

***Análisis espacial de la distribución en el territorio de variables sanitarias y socioeconómicas mediante técnicas de inteligencia artificial y sus aplicaciones en la planificación sociosanitaria de la salud mental***

DOCTORANDO: José Alberto Salinas Pérez

DIRECTORES: Dr. Carlos Ramón García Alonso y Dr. Luis Salvador Carulla

PROGRAMA DE DOCTORADO: Ciencias jurídicas y empresariales

LÍNEA DE INVESTIGACIÓN: Aprendizaje de modelos neuro y bioinspirados aplicados a problemas Reales

DEPARTAMENTO: Departamento de Gestión Empresarial y Métodos Cuantitativos

FACULTAD: Facultad de Ciencias Económicas y Empresariales (ETEA) – Universidad de Córdoba

**TITULO:** *Análisis espacial de la distribución en el territorio de variables sanitarias y socioeconómicas mediante técnicas de inteligencia artificial y sus aplicaciones en la planificación sociosanitaria de la salud mental.*

**AUTOR:** *JOSÉ ALBERTO SALINAS PÉREZ*

---

© Edita: Servicio de Publicaciones de la Universidad de Córdoba.  
Campus de Rabanales  
Ctra. Nacional IV, Km. 396 A  
14071 Córdoba

[www.uco.es/publicaciones](http://www.uco.es/publicaciones)  
[publicaciones@uco.es](mailto:publicaciones@uco.es)

---





## **TÍTULO DE LA TESIS:**

ANÁLISIS ESPACIAL DE LA DISTRIBUCIÓN EN EL TERRITORIO DE VARIABLES SANITARIAS Y SOCIOECONÓMICAS MEDIANTE TÉCNICAS DE INTELIGENCIA ARTIFICIAL Y SUS APLICACIONES EN LA PLANIFICACIÓN SOCIOSANITARIA DE LA SALUD MENTAL

**DOCTORANDO/A:** JOSÉ ALBERTO SALINAS PÉREZ

### **INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS**

(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

La Tesis Doctoral titulada:

ANÁLISIS ESPACIAL DE LA DISTRIBUCIÓN EN EL TERRITORIO DE VARIABLES SANITARIAS Y SOCIOECONÓMICAS MEDIANTE TÉCNICAS DE INTELIGENCIA ARTIFICIAL Y SUS APLICACIONES EN LA PLANIFICACIÓN SOCIOSANITARIA DE LA SALUD MENTAL

es la culminación de un proceso intensivo de investigación científica sobre la distribución espacial de fenómenos relacionados con la salud de la población. Durante su desarrollo, se han realizado numerosos proyectos de I+D+i, asistido a relevantes eventos científicos y publicado monografías y artículos científicos. El proceso ha incluido: i) la revisión sistemática e intensiva de la bibliografía científica relacionada con la materia objeto del estudio, ii) el diseño y desarrollo de herramientas informáticas para la resolución de los distintos problemas matemático-estadísticos, iii) la validación de las distintas metodologías y procesos informáticos en situaciones reales y, finalmente, iv) la elicitación de conocimiento experto de diferentes agentes y decisores sobre el que se fundamentaron los procesos, herramientas y conclusiones del trabajo presentado.

La presente Tesis Doctoral es, por tanto, resultado de un esfuerzo conjunto entre la universidad y diferentes administraciones públicas, siendo en nuestra opinión un ejemplo de investigación científica que trasciende el mero conocimiento académico y se proyecta sobre la planificación y gestión de situaciones reales. Teniendo en cuenta esta característica, las principales dificultades que se afrontaron fueron: i) la estructura y diseño de las bases de datos disponibles –sanitarias, ii) la discrepancia metodológica extrema que existe en este ámbito científico, iii) la inexistencia de software especializado en este tipo de análisis y, finalmente, iv) la dispersión del conocimiento experto disponible. Todas estas dificultades se han abordado con rigurosidad científica y, como ya se ha mencionado, validado en situaciones reales.

La metodología diseñada y desarrollada en este trabajo de investigación ha demostrado ser de especial utilidad en el ámbito del estudio y de la planificación de

servicios de salud mental. Basándose en la evidencia, los entes decisores tienen a su disposición un conjunto flexible de herramientas para apoyar sus decisiones. Esta flexibilidad permite incorporar nuevos métodos al análisis de una forma sencilla puesto que los procesos desarrollados no están condicionados por determinadas técnicas específicas.

En la actualidad, el equipo de investigación del cual forma parte esencial el doctorando continúa trabajando en estos temas, siendo requerido por determinadas administraciones públicas responsables en la gestión sanitaria para ampliar el ámbito de estudio en temas relacionados con: i) la relación de la distribución geográfica de enfermedades con variables socioeconómicas, ii) la accesibilidad, adecuación y eficiencia de los servicios y iii) el análisis específico de áreas macrouurbanas.

Por todo ello consideramos que la Tesis Doctoral presentada tiene todos los avales necesarios para su presentación y defensa.

Agradecemos profundamente la colaboración de la Junta de Andalucía, la Generalitat de Catalunya, el Gobierno Autónomo de Madrid, el Gobierno Autónomo de Cantabria, la Diputación Foral de Guipúzcoa, la Diputación Foral de Vizcaya, así como al Instituto de Salud Carlos III y al Séptimo Programa Marco de la UE.

Producción científica:

#### *Proyectos de investigación*

2005-2006. Atlas de la Salud Mental de Andalucía. Convenio entre la Asociación Científica Psicost y el Servicio Andaluz de Salud.

2005-2008. Atlas de la Salud Mental de Catalunya. Convenio entre la Asociación Científica Psicost y el Servei Català de la Salut (CatSalut).

2009. Atlas de la Salud Mental de Cantabria. Consejería de Sanidad del Gobierno de Cantabria.

2009. Atlas de la Salud Mental de España. Ministerio de Sanidad y Política Social.

2009-2010. Identificación, localización y análisis espacial de zonas de alta prevalencia en enfermedades con alto impacto social mediante técnicas de inteligencia artificial. Proyecto de investigación de evaluación de tecnologías sanitarias. Instituto de Salud Carlos III. Ministerio de Sanidad y Consumo.

2009-2011. Atlas de la Salud Mental de Madrid. Consejería de Sanidad de la Comunidad de Madrid.

2010. Elaboración de un mapa sanitario de servicios y recursos destinados a la atención integral de personas con enfermedad mental en Cataluña y aplicación de sistemas de información geográfica para la ayuda a la decisión en la planificación de servicios en Cataluña. Departamento de Salud de Cataluña.

2010-2012. Research on Financing systems' Effect on the Quality of Mental health care (REFINEMENT). Seventh Framework Programme. Coordinado por la Universidad de Verona.

2011-2012. Atlas de Salud Mental de Gipuzkoa. Instituto Vasco de Innovación Sanitaria / o+berri (Fundación Vasca de Innovación e Investigación Sanitarias).

2012-2014. Análisis espacial y de clasificación ordinal de la distribución geográfica de enfermedades mentales en Andalucía. Proyecto de investigación de evaluación de tecnologías sanitarias. Instituto de Salud Carlos III. Ministerio de Sanidad y Consumo.

2012. Atlas de Salud Mental de Bizkaia. Red de Salud Mental de Bizkaia.

### *Congresos*

García-Alonso C, Salinas-Pérez JA, Salvador-Carulla L. Spatial Identification of Hot and Cold Spots for the Prevalence of Schizophrenia and Depression. International Conference on Operations Research, OR2010. Munich, 1 – 3 de septiembre de 2010.

García-Alonso C, Salinas-Pérez JA, Salvador-Carulla L. Spatial Analysis of Schizophrenia Prevalence Using a Multiobjective Evolutionary Algorithm. 5th Workshop on Data Mining and Health Informatics. Institute for Operations Research and the Management Sciences INFORMS Annual Meeting 2010. Austin, 5 – 11 de noviembre de 2010.

Pérez-Ortiz M, García-Alonso C, Gutiérrez PA, Salvador-Carulla L, Salinas-Pérez JA, Hervás-Martínez C. Ordinal classification of depression spatial hotspots of prevalence. ISDA 2011: Eleventh International Conference on Intelligent Systems Design and Applications. Córdoba, 22 – 24 de noviembre de 2011.

García Alonso CR, Salinas Pérez JA, Salvador Carulla L, Almenara Barrios J, Peralta Sáez JL, Poole Quintana, ML, Rodero Cosano ML, De los Reyes Vázquez SM. Identificación y caracterización socioeconómica de hot-spots de depresión en Andalucía. Póster presentado a la XXX Reunión científica de la Sociedad Española de Epidemiología: Epidemiología en tiempos de crisis: Haciendo sostenible el sistema de salud. Santander, 17-19 octubre de 2012.

### *Publicaciones*

Garrido M, Salinas JA, Almenara J, Salvador L. Atlas de Salud Mental de Andalucía 2005. Sevilla: Servicio Andaluz de Salud; 2006. ISBN: 978-84-690-6666-9.

Vázquez-Barquero JL, Gaité L, Salvador L, Salinas JA. Atlas de Salud Mental de Cantabria. Santander: Consejería de Sanidad; 2010. ISBN: 978-84-693-6768-1.

Salinas-Perez JA, Garcia-Alonso CR, Molina-Parrilla C, Jorda-Sampietro E, Salvador-Carulla L. Identification and location of hot and cold spots of treated prevalence of depression in Catalonia (Spain). International Journal of Health Geographics 2012 ago;11(1):36.

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, \_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_

Firma del/de los director/es

Fdo.: \_\_\_\_\_ Fdo.: \_\_\_\_\_



## Índice

I.	INTRODUCCIÓN .....	17
1.1.	El estudio de la distribución espacial de los datos .....	17
1.1.1.	El Análisis Espacial de Datos.....	17
1.1.2.	Las agrupaciones espaciales de datos.....	18
1.2.	Métodos de análisis de las agrupaciones espaciales de datos.....	20
1.2.1.	Origen de los principales métodos de análisis.....	20
1.2.2.	Principales métodos de AED.....	21
1.3.	Análisis de los patrones espaciales de las enfermedades .....	23
1.3.1.	Epidemiología Espacial.....	23
1.3.2.	Análisis espacial aplicado a enfermedades .....	25
1.3.3.	Análisis espacial aplicado a enfermedades mentales .....	27
1.4.	Justificación y planteamiento de la investigación .....	28
1.4.1.	Relevancia de la metodología de análisis espacial .....	28
1.4.2.	Relevancia del estudio de las enfermedades mentales .....	30
1.4.3.	Estructura y organización del trabajo de investigación.....	32
II.	HIPÓTESIS Y OBJETIVOS.....	33
2.1.	Hipótesis.....	33
2.2.	Objetivo general .....	33
2.3.	Objetivos específicos.....	33
III.	METODOLOGÍA .....	35
3.1.	Diseño .....	35
3.1.1.	Estudio de Demostración A.....	37
3.1.2.	Estudio de Demostración B.....	37
3.2.	Creación de una nueva herramienta de Análisis Espacial de Datos .....	38
3.2.1.	Los Algoritmos Evolutivos y la resolución de problemas multi-objetivos .....	38
3.2.2.	Elementos de un AEMO .....	39
3.2.3.	Funcionamiento de un AEMO .....	43
3.3.	Estudios de Demostración .....	47
3.3.1.	Fuentes de información .....	47
3.3.1.1.	El Sistema de Información de Salud Mental de Andalucía .....	47
3.3.1.1.1.	Descripción del SISMA .....	47
3.3.1.1.2.	Explotación del SISMA .....	49
3.3.1.2.	El Conjunto Mínimo Básico de Datos de Salud Mental Ambulatoria.....	50
3.3.1.2.1.	Descripción del CMBD-SMA.....	50
3.3.1.2.2.	Explotación del CMBD-SMA.....	52



3.3.2.	Variables espaciales de análisis.....	53
3.3.2.1.	Las relaciones entre variables espaciales.....	53
3.3.2.2.	Prevalencia administrativa.....	56
3.3.2.3.	Análisis espacial exploratorio de la prevalencia administrativa.....	57
3.3.2.3.1.	<i>I</i> de Moran global.....	58
3.3.2.3.2.	<i>G</i> de Getis y Ord global.....	60
3.3.2.4.	Índices locales para el análisis espacial de datos.....	61
3.3.2.4.1.	Definición de las relaciones espaciales empleadas en el estudio.....	61
3.3.2.4.2.	Indicadores Locales de Asociación Espacial (LISA).....	63
3.3.2.4.3.	Modelo Autorregresivo Condicional Bayesiano (Modelo CAR).....	67
3.4.	Sistemas de Información Geográfica (SIG).....	69
3.5.	Limitaciones de los Estudios de Demostración.....	71
3.6.	Valoración del Modelo AEMO/AE.....	71
IV.	RESULTADOS.....	73
4.1.	El Algoritmo Evolutivo Multi-Objetivo de Análisis Espacial (AEMO/AE).....	73
4.1.1.	Fases y procedimientos del AEMO/AE.....	73
4.1.2.	El Algoritmo Evolutivo de Fuerza de Pareto v.2.....	74
4.1.3.	Diseño del AEMO/AE.....	75
4.1.3.1.	El problema multi-objetivo.....	75
4.1.3.2.	Objetivos y funciones objetivo.....	76
4.1.3.2.1.	Estudio de Demostración A.....	76
4.1.3.2.2.	Estudio de Demostración B.....	79
4.1.3.3.	Parámetros del AEMO/AE.....	81
4.1.3.4.	Funciones de aptitud.....	84
4.1.3.4.1.	<i>Fine-grained Strength Pareto</i> .....	84
4.1.3.4.2.	<i>Weighted Sum of Objectives</i> .....	85
4.1.3.4.3.	<i>Standard Rankings Selection</i> .....	85
4.1.3.4.4.	<i>Fuzzy Evaluation of Weighted Objectives</i> .....	86
4.1.4.	Selección de las soluciones finales.....	86
4.1.4.1.	Soluciones del AEMO/AE.....	86
4.1.4.2.	Selección mediante estadísticas de los extremos: QQ-Plots.....	87
4.1.4.3.	Método de selección ad hoc.....	89
4.1.4.4.	Selección de hot-spots.....	89
4.2.	Estudio de Demostración A: aplicación del AEMO/AE sobre datos de Enfermedades Mentales en Andalucía.....	91
4.2.1.	Caso 1: <i>Hot-spots</i> de prevalencia asistida de la esquizofrenia en Andalucía durante una serie temporal de cuatro años.....	91

4.2.1.1.	Distribución geográfica de la prevalencia administrativa de la esquizofrenia ..	91
4.2.1.2.	Dependencia espacial de la prevalencia administrativa de la esquizofrenia .....	96
4.2.1.3.	Hot-spots de prevalencia administrativa de la esquizofrenia .....	104
4.2.2.	Caso 2: <i>Hot-spots</i> de prevalencia asistida de la depresión en Andalucía durante una serie temporal de cuatro años .....	120
4.2.2.1.	Distribución geográfica de la prevalencia administrativa de la depresión .....	120
4.2.2.2.	Dependencia espacial de la prevalencia administrativa de la depresión .....	125
4.2.2.3.	Hot-spot de prevalencia administrativa de la depresión .....	133
4.3.	Estudio de Demostración B: aplicación del AEMO/AE sobre datos de Enfermedades Mentales en Cataluña .....	148
4.3.1.	Caso 3: <i>Hot-spots</i> y <i>cold-spots</i> de prevalencia asistida de la esquizofrenia en Cataluña en 2009 .....	148
4.3.1.1.	Distribución geográfica de la prevalencia administrativa de la esquizofrenia .....	148
4.3.1.2.	Hot-spot y cold-spot de prevalencia administrativa de la esquizofrenia .....	152
4.3.2.	Caso 4: <i>Hot-spots</i> y <i>cold-spots</i> de prevalencia asistida de la depresión en Cataluña en 2009 .....	158
4.3.2.1.	Distribución geográfica de la prevalencia administrativa de la depresión .....	158
4.3.2.2.	Hot-spot y cold-spot de prevalencia administrativa de la depresión .....	162
V.	DISCUSIÓN .....	169
5.1.	El modelo AEMO/AE .....	169
5.2.	Estudios de Demostración .....	172
5.3.	Limitaciones del estudio .....	175
5.4.	<i>Hot-spots/cold-spots</i> y posibles factores explicativos .....	179
5.5.	Difusión del modelo AEMO/AE .....	182
5.6.	AED como apoyo a la decisión en planificación de la salud mental .....	186
5.7.	Líneas futuras de desarrollo de la investigación .....	187
VI.	CONCLUSIONES .....	189
VII.	BIBLIOGRAFÍA .....	191
	Anexo 1 .....	205
	Anexo 2 .....	211

## Índices de tablas, figuras y mapas

### Índice de tablas

Tabla 1: Actividades y pacientes recogidos en el Sistema de Información de Salud Mental de Andalucía .....	49
Tabla 2: Pacientes con esquizofrenia y depresión atendidos en servicios de salud mental ambulatorio en Andalucía .....	50
Tabla 3: Actividades y pacientes recogidos en el Conjunto Mínimo Básico de Datos de Salud Mental Ambulatoria de Cataluña .....	52
Tabla 4: Pacientes con esquizofrenia y depresión atendidos en servicios de salud mental ambulatorios de Cataluña .....	53
Tabla 5: Estructura del AEMO/AE. Descripción de fases y procedimientos.....	73
Tabla 6: Procedimiento del AEMO/AE .....	75
Tabla 7: Estadísticas descriptivas de la prevalencia administrativa de la esquizofrenia en Andalucía .....	91
Tabla 8: Distribución de frecuencias de la prevalencia administrativa de la esquizofrenia en Andalucía .....	93
Tabla 9: Análisis de la autocorrelación espacial global de la prevalencia asistida de la esquizofrenia en la serie temporal (Índices globales de la <i>I</i> de Moran y la <i>G</i> de Getis y Ord)...	94
Tabla 10: Estadísticas descriptivas de la <i>I</i> de Moran .....	96
Tabla 11: Estadísticas descriptivas de la <i>C</i> de Geary .....	97
Tabla 12: Estadísticas descriptivas de la <i>G*</i> de Getis y Ord.....	97
Tabla 13: Estadísticas descriptivas del CAR Bayesiano .....	97
Tabla 14: Estadísticas descriptivas de los <i>hot-spots</i> de esquizofrenia en 2004.....	104
Tabla 15: Estadísticas descriptivas de los <i>hot-spots</i> de esquizofrenia en 2006.....	104
Tabla 16: Estadísticas descriptivas de los <i>hot-spots</i> de esquizofrenia en 2007.....	105
Tabla 17: Estadísticas descriptivas de los <i>hot-spots</i> de esquizofrenia en 2008.....	105
Tabla 18: Estadísticas descriptivas de los <i>hot-spots</i> de esquizofrenia .....	105
Tabla 19: Municipios que forman parte de los hot-spots de esquizofrenia durante la serie temporal.....	106
Tabla 20: Comparación de las distribuciones y medianas de las unidades espaciales que forman los diferentes <i>hot-spots</i> de esquizofrenia y aquellas que no forman parte de ninguno .....	108
Tabla 21: Estadísticas descriptivas de la prevalencia administrativa de la depresión en Andalucía .....	120
Tabla 22: Distribución de frecuencias de la prevalencia administrativa de la depresión en Andalucía .....	122

Tabla 23: Análisis de la autocorrelación espacial global de la prevalencia asistida de la depresión en la serie temporal (Índices globales de la <i>I</i> de Moran y la <i>G</i> de Getis y Ord) .....	123
Tabla 24: Estadísticas descriptivas de la <i>I</i> de Moran .....	125
Tabla 25: Estadísticas descriptivas de la <i>C</i> de Geary .....	126
Tabla 26: Estadísticas descriptivas de la <i>G*</i> de Getis & Ord.....	126
Tabla 27: Estadísticas descriptivas del CAR Bayesiano .....	126
Tabla 28: Estadísticas descriptivas de los <i>hot-spots</i> de depresión .....	133
Tabla 29: Estadísticas descriptivas de los <i>hot-spots</i> de esquizofrenia en 2004.....	133
Tabla 30: Estadísticas descriptivas de los <i>hot-spots</i> de esquizofrenia en 2006.....	134
Tabla 31: Estadísticas descriptivas de los <i>hot-spots</i> de esquizofrenia en 2007.....	134
Tabla 32: Estadísticas descriptivas de los <i>hot-spots</i> de esquizofrenia en 2008.....	134
Tabla 33: Municipios que forman parte de los <i>hot-spots</i> de depresión durante la serie temporal .....	135
Tabla 34: Comparación de las distribuciones y medianas de las unidades espaciales que forman los diferentes <i>hot-spots</i> de depresión y aquellas que no forman parte de ninguno .....	137
Tabla 35: Estadísticas descriptivas de la prevalencia administrativa de la esquizofrenia en Cataluña.....	148
Tabla 36: Distribución de frecuencias de la prevalencia administrativa de la esquizofrenia en Cataluña.....	150
Tabla 37: Análisis de la autocorrelación espacial global de la prevalencia asistida de la depresión en la serie temporal (Índices globales de la <i>I</i> de Moran y la <i>G</i> de Getis y Ord) .....	150
Tabla 38: Estadísticas descriptivas de los <i>hot-spots</i> y <i>cold-spots</i> de esquizofrenia .....	152
Tabla 39: Municipios que forman parte de los <i>hot-spots</i> de esquizofrenia en 2009 .....	153
Tabla 40: Municipios que forman parte de los <i>cold-spots</i> de esquizofrenia en 2009 .....	153
Tabla 41: Comparación de las distribuciones y medianas de las unidades espaciales que forman los diferentes <i>hot-spots</i> y <i>cold-spots</i> de esquizofrenia y aquellas que no forman parte de ninguno .....	155
Tabla 42: Estadísticas descriptivas de la prevalencia administrativa de la depresión en Cataluña .....	158
Tabla 43: Distribución de frecuencias de la prevalencia administrativa de la depresión en Cataluña.....	160
Tabla 44: Análisis de la autocorrelación espacial global de la prevalencia asistida de la depresión en la serie temporal (Índices globales de la <i>I</i> de Moran y la <i>G</i> de Getis y Ord) .....	160
Tabla 45: Estadísticas descriptivas de los <i>hot-spots</i> de depresión .....	162
Tabla 46: Municipios que forman parte de los <i>hot-spots</i> de depresión en 2009 .....	163
Tabla 47: Municipios que forman parte de los <i>cold-spots</i> de depresión en 2009 .....	164

Tabla 48: Comparación de las distribuciones y medianas de las unidades espaciales que forman los diferentes <i>hot-spots</i> de depresión y aquellas que no forman parte de ninguno .....	165
---	-----

## Índice de figuras

Figura 1: Fases del trabajo de investigación. ....	36
Figura 2: Dominancia de soluciones en la Teoría del Óptimo de Pareto. ....	41
Figura 3: Esquema de un Algoritmo Evolutivo.....	44
Figura 4: Fases del algoritmo Evolutivo. ....	44
Figura 5: Funcionamiento del operador genético de recombinación o cruzamiento.....	45
Figura 6: Funcionamiento del operador genético de mutación. ....	46
Figura 7: Ejemplo de matrices de pesos espaciales.....	54
Figura 8: Medición de la distancias entre unidades espaciales .....	79
Figura 9: Procedimiento de la selección por torneo .....	83
Figura 10: Esquema de la selección de <i>hot-spot</i> y <i>cold-spot</i> finales.....	90
Figura 11: Histogramas de frecuencia de la prevalencia administrativa de la esquizofrenia en la serie temporal.....	92
Figura 12: <i>Box-plot</i> de las poblaciones de prevalencia administrativa de la esquizofrenia en la serie temporal.....	93
Figura 13: <i>Box-Plot</i> de los valores de la <i>I</i> de Moran para la prevalencia de la esquizofrenia.....	98
Figura 14: <i>Box-Plot</i> de los valores de la <i>C</i> de Geary para la prevalencia de la esquizofrenia ....	98
Figura 15: <i>Box-Plot</i> de los valores de la <i>G*</i> de Getis & Ord para la prevalencia de la esquizofrenia .....	99
Figura 16: <i>Box-Plot</i> de los valores del modelo <i>CAR</i> Bayesiano para la prevalencia de la esquizofrenia .....	99
Figura 17: Histogramas de frecuencia de la prevalencia administrativa de la esquizofrenia de los <i>hot-spots</i> en la serie temporal.....	106
Figura 18: Histogramas de frecuencia de la prevalencia administrativa de la depresión en la serie temporal.....	121
Figura 19: <i>Box-plot</i> de las poblaciones de prevalencia administrativa de la depresión en la serie temporal.....	122
Figura 20: <i>Box-Plot</i> de los valores de la <i>I</i> de Moran para la prevalencia de la depresión.....	127
Figura 21: <i>Box-Plot</i> de los valores de la <i>C</i> de Geary para la prevalencia de la depresión .....	127
Figura 22: <i>Box-Plot</i> de los valores de la <i>G</i> de Getis & Ord para la prevalencia de la depresión .....	128
Figura 23: <i>Box-Plot</i> de los valores del modelo <i>CAR</i> Bayesiano para la prevalencia de la depresión .....	128

Figura 24: Histogramas de frecuencia de la prevalencia administrativa de la depresión de los <i>hot-spots</i> en la serie temporal.....	135
Figura 25: Histogramas de frecuencia de la prevalencia administrativa de la esquizofrenia en 2009.....	149
Figura 26: <i>Box-plot</i> de las poblaciones de prevalencia administrativa de la esquizofrenia en 2009.....	149
Figura 27: Histogramas de frecuencia de la prevalencia administrativa de la esquizofrenia de los <i>hot-spots</i> en 2009 .....	152
Figura 28: Histogramas de frecuencia de la prevalencia administrativa de la depresión en 2009 .....	158
Figura 29: <i>Box-plot</i> de las poblaciones de prevalencia administrativa de la depresión en 2009 .....	159
Figura 30: Histogramas de frecuencia de la prevalencia administrativa de la esquizofrenia de los <i>hot-spots</i> y <i>cold-spots</i> en 2009 .....	162

## Índice de mapas

Mapa 1: Áreas de cobertura de las Unidades de Salud Mental Comunitaria de Andalucía (770 municipios).....	48
Mapa 2: Áreas de cobertura de los Centros de Salud Mental de Adultos de Cataluña (946 municipios).....	51
Mapa 3: Prevalencia administrativa de la esquizofrenia en los municipios de Andalucía.....	95
Mapa 4: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la esquizofrenia en los municipios de Andalucía (2004) .....	100
Mapa 5: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la esquizofrenia en los municipios de Andalucía (2006) .....	101
Mapa 6: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la esquizofrenia en los municipios de Andalucía (2007) .....	102
Mapa 7: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la esquizofrenia en los municipios de Andalucía (2008) .....	103
Mapa 8: Resultados del AEMO/AE por <i>función de aptitud</i> para los <i>hot-spot</i> de esquizofrenia en Andalucía (2004).....	112
Mapa 9: <i>Hot-spots</i> de dependencia espacial de la esquizofrenia en Andalucía (2004).....	113
Mapa 10: Resultados del AEMO/AE por <i>función de aptitud</i> para los <i>hot-spot</i> de esquizofrenia en Andalucía (2006).....	114
Mapa 11: <i>Hot-spots</i> de dependencia espacial de la esquizofrenia en Andalucía (2006).....	115

Mapa 12: Resultados del AEMO/AE por <i>función de aptitud</i> para los <i>hot-spot</i> de esquizofrenia en Andalucía (2007).....	116
Mapa 13: <i>Hot-spots</i> de dependencia espacial de la esquizofrenia en Andalucía (2007).....	117
Mapa 14: Resultados del AEMO/AE por <i>función de aptitud</i> para los <i>hot-spot</i> de esquizofrenia en Andalucía (2008).....	118
Mapa 15: <i>Hot-spots</i> de dependencia espacial de la esquizofrenia en Andalucía (2008).....	119
Mapa 16: Prevalencia administrativa de la depresión en los municipios de Andalucía.....	124
Mapa 17: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la depresión en los municipios de Andalucía (2004) .....	129
Mapa 18: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la depresión en los municipios de Andalucía (2006) .....	130
Mapa 19: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la depresión en los municipios de Andalucía (2007) .....	131
Mapa 20: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la depresión en los municipios de Andalucía (2008) .....	132
Mapa 21: Resultados del AEMO/AE por <i>función de aptitud</i> para los <i>hot-spot</i> de depresión en Andalucía (2004).....	140
Mapa 22: <i>Hot-spots</i> de dependencia espacial de la depresión en Andalucía (2004).....	141
Mapa 23: Resultados del AEMO/AE por <i>función de aptitud</i> para los <i>hot-spot</i> de depresión en Andalucía (2006).....	142
Mapa 24: <i>Hot-spots</i> de dependencia espacial de la depresión en Andalucía (2006).....	143
Mapa 25: Resultados del AEMO/AE por <i>función de aptitud</i> para los <i>hot-spot</i> de depresión en Andalucía (2007).....	144
Mapa 26: <i>Hot-spots</i> de dependencia espacial de la depresión en Andalucía (2007).....	145
Mapa 27: Resultados del AEMO/AE por <i>función de aptitud</i> para los <i>hot-spot</i> de depresión en Andalucía (2008).....	146
Mapa 28: <i>Hot-spots</i> de dependencia espacial de la depresión en Andalucía (2008).....	147
Mapa 29: Prevalencia administrativa de la esquizofrenia en Centros de Salud Mental de Cataluña.....	151
Mapa 30: <i>Hot-spots</i> y <i>cold-spots</i> de prevalencia administrativa de la esquizofrenia en Cataluña (2009) .....	157
Mapa 31: Prevalencia administrativa de la depresión en Centros de Salud Mental de Cataluña .....	161
Mapa 32: <i>Hot-spots</i> y <i>cold-spots</i> de prevalencia administrativa de la depresión en Cataluña (2009) .....	167

## ACRÓNIMOS

*AED: Análisis Espacial de Datos*

*AEMO/AE: Algoritmo Evolutivo Multi-objetivo para análisis espacial*

*AEMO: Algoritmos Evolutivos Multi-Objetivos*

*AG: Algoritmo Genético*

*BYM: Besag, York y Mollie*

*CAR: Conditional Autorregresive Model*

*CIE-10: Clasificación Internacional de Enfermedades en su 10ª revisión*

*CIE-9: Clasificación Internacional de Enfermedades en su 9ª revisión*

*CMBD: Conjunto Mínimo de Datos al Alta Hospitalaria*

*CMBD-SMA: Conjunto Mínimo Básico de Datos de Centros de Salud Mental Ambulatorios*

*CSd: Cold-spot depresión*

*CSe: Cold-spot esquizofrenia*

*CSMA: Centros de salud Mental de Adultos*

*CSMIJ: Centros de Salud Mental Infanto-juveniles*

*EBC: European Brain Council*

*EbCA: Expert-based Cooperative Analysis*

*ECM: Error Cuadrático Medio*

*ECNP: European College of Neuropsychopharmacology*

*ENDSF: External non-dominated solution file*

*ENSE: Encuesta Nacional de Salud*

*ESEMeD: The European Study of the Epidemiology of Mental Disorders*

*ESRI: Enviromental Systems Research Institute*

*FI: Fuzzy Evaluation of Weighted Objectives*

*FSS: Flexible Scan Statistics*

*HSd: Hot-spot depresión*

*HSe: Hot-spot esquizofrenia*

*INE: Instituto Nacional de Estadística*

*ISCHIII: Instituto de Salud Carlos III*

*KDD: Knowledge Discovery from Data*

*LISA: Local Indicators of Spatial Association*

*MCMC: Cadenas de Markov Monte-Carlo*

*MOGA: Multi-Objective Genetic Algorithm*

*NSGA: Non-Dominated Sorting Genetic Algorithm*

*NSGA-II: Non-Dominated Sorting Genetic Algorithm II*

*Organisation for Economic Co-operation and Development (OECD)*



*PAES: Pareto Archived Evolution Strategy*

*QQ-Plots: Quantile-Quantile Plots*

*REFINEMENT: Research on Financing systems' Effect on the Quality of Mental health care*

*SAS: Servicio Andaluz de Salud*

*SIG: Sistemas de Información Geográfica*

*SISMA: Sistema de Información de Salud Mental de Andalucía*

*SPEA: Strength Pareto Evolutionary Algorithm*

*SPEA2: Strength Pareto Evolutionary Algorithm*

*SRS: Standard Rankings Selection*

*ULS: Upper Level Set Scan Statistics*

*USMC: Unidades de Salud Mental Comunitaria*

*WMH-CIDI: Composite International Diagnostic Interview*

*WO: Weighted Sum of Objectives*

# I. INTRODUCCIÓN

---

## 1.1. El estudio de la distribución espacial de los datos

### 1.1.1. El Análisis Espacial de Datos

El Análisis Espacial de Datos (AED) puede definirse como el estudio cuantitativo de fenómenos ubicados en el espacio (Bailey y Gatrell, 1995). El AED se emplea cuando la ubicación geográfica de las observaciones y las relaciones espaciales entre las mismas se consideran importantes para el análisis y la interpretación de un determinado fenómeno. Con este fin el AED reúne un conjunto de técnicas para el estudio de las propiedades topológicas, geométricas y geográficas de los datos espaciales. La relevancia del AED radica en que muchos de los datos que se emplean habitualmente en trabajos de investigación poseen una base geográfica o componente espacial, es decir, describen fenómenos (población, actividades, entorno, morbilidad, etc.) que se localizan o desarrollan en unidades espaciales definidas (países, regiones, ciudades, etc.).

Los datos espaciales son aquellos que se recopilan en el territorio y, por tanto, tienen una localización geográfica conocida (Haining, 1993). Lo que distingue a los datos espaciales de otros tipos de datos es que poseen una ubicación en el espacio y relaciones espaciales entre ellos. La localización de los datos puede aportarse de forma exacta mediante coordenadas geográficas (*georreferenciación*) o de forma aproximada si se localizan en una unidad espacial definida (ciudad, comarca, provincia, región, etc.).

Este tipo de datos se administran en bases de datos espaciales donde los elementos geográficos del mundo real se abstraen en objetos espaciales. Las bases de datos espaciales pueden organizarse en dos modelos de datos: el modelo ráster y el modelo vectorial. Por una parte, el modelo ráster conceptualiza los datos espaciales mediante píxeles, que es la unidad espacial mínima de información, y cuyas agrupaciones constituyen los diferentes objetos espaciales del territorio. Y por otra, el modelo vectorial representa los datos espaciales de forma geométrica mediante el punto, la línea y el polígono como objetos espaciales básicos (Smith, Goodchild, et al., 2007).

En las últimas décadas, estos estudios se han visto facilitados e impulsados con la utilización y desarrollo de estadísticas espaciales o métodos geoestadísticos y de los Sistemas de Información Geográfica (SIG). Estos sistemas computerizados facilitan la proyección espacial y el análisis

de datos (Anselin y Hudak, 1992). Los SIG permiten gestionar las bases de datos espaciales permitiendo almacenar, recuperar, manipular y analizar los datos geográficos (Cromley y McLafferty, 2002). Su gran difusión en las últimas décadas responde a que su entorno de usuario es cada vez más accesible a personas no expertas (*user-friendly*) y a la generalización de la disponibilidad de datos espaciales y cartografía digital (Griffith, 2011).

### **1.1.2. Las agrupaciones espaciales de datos**

Uno de los múltiples objetos de estudio del AED es el análisis de patrones espaciales (*spatial patterns*) en la distribución de los datos en el espacio geográfico (Shekhar, Evans, et al., 2011). El AED facilita las técnicas necesarias para conocer si existen agrupaciones espaciales (*spatial clusterings*) de valores similares o disímiles de una (autocorrelación espacial) o varias variables (correlación espacial multivariante). También permite analizar sus características con la finalidad de poder establecer hipótesis para su explicación. Las agrupaciones espaciales de datos se explican por la existencia de dependencia espacial y/o heterogeneidad espacial. Ambas constituyen los denominados *efectos espaciales* que se producen sobre los datos geográficos (Anselin, 1988; Moreno y Vayá, 2000).

La dependencia espacial señala la existencia de relaciones entre variables provocadas por su localización y distribución en el espacio geográfico, es decir, los valores de una variable no se distribuyen en el territorio de forma aleatoria sino que vienen dados por los valores ubicados en su entorno. La dependencia se definiría por la similitud o disimilitud entre los valores de una variable y entre las ubicaciones espaciales de la misma. Supone la materialización de la Primera Ley de la Geografía de W. R. Tobler que señala que todos los elementos en el espacio están relacionados entre sí, pero los elementos cercanos están más relacionados que los lejanos (Tobler, 1970).

Las concentraciones espaciales de valores semejantes, normalmente aquellos que sean significativamente altos o bajos, indicarían la existencia de dependencia espacial positiva (concentración), en cambio, si se agrupan valores significativamente diferentes la dependencia espacial sería negativa (dispersión). Las variables son espacialmente independientes cuando sus valores no están influidos por su localización y su distribución geográfica es aleatoria. La importancia de considerar la dependencia espacial en un conjunto de datos también radica en que su existencia puede rebatir la asunción de independencia y de distribución aleatoria de los mismos, que en principio son asumidas como tales por ser recogidas de forma independiente y bajo las mismas condiciones (Cressie, 1991).

Por su parte, la heterogeneidad o heterocedasticidad espacial es debida a la existencia de estructuras espaciales subyacentes que provocan la ausencia de estabilidad espacial en el comportamiento de los datos espaciales. En este caso, los valores de las variables espaciales están influidos por las características geográficas inherentes a su ubicación que no son homogéneas en todo el espacio (Moreno y Vayá, 2000). La heterogeneidad espacial aparece cuando se quiere explicar un fenómeno a partir de datos de unidades espaciales muy diferentes (espacio no uniforme) debido a su localización, tamaño o a causas sociológicas y/o políticas. Por ejemplo, las diferencias en el nivel económico o las políticas de los países o regiones influye de forma decisiva en determinadas variables (salarios, precios, inversión, etc.) apareciendo estructuras centro-periferia, norte-sur, etc. Esta circunstancia se manifiesta en una varianza no constante de los residuos o coeficientes en los modelos de regresión que utilicen estos valores.

Los *efectos espaciales* pueden aparecer debido a varias causas (Haining, 2001):

- a) La continuidad espacial de algunos fenómenos geográficos provoca que valores próximos en el espacio sean semejantes. Por ejemplo, la temperatura o el relieve cambian de forma progresiva, por lo que si se toman sus valores en puntos cercanos serán muy parecidos y presentarán dependencia espacial.
- b) La existencia de errores en la recogida de datos también puede provocar dependencia espacial.
- c) La distribución geográfica de los valores de una variable que mide un determinado fenómeno puede ser influida por procesos subyacentes:
  - a. Procesos de difusión (*diffusion*): un fenómeno puede transmitirse en el espacio por proximidad.
  - b. Procesos de derrame (*spillover*): las condiciones del área geográfica y de sus vecinas pueden influir sobre un fenómeno.
  - c. Procesos de interacción (*interaction*): cuando las acciones que se realizan sobre una localización influyen sobre las acciones que se realizan en otras y afectan a un fenómeno en ambas localizaciones.

- d. Procesos de dispersión (*dispersal*): la dispersión de los individuos en el territorio hace que compartan los valores de un determinado fenómeno.

El estudio de la existencia y localización de concentraciones espaciales de variables socioeconómicas, sanitarias o de otro tipo, resulta especialmente interesante de cara al diseño de políticas territoriales desde diferentes ámbitos como la economía, la sanidad, la asistencia social, las obras públicas, etc. Los fenómenos de asociación espacial de valores significativamente altos o bajos de una determinada variable socioeconómica (*hot-spots* y *cold-spots*) pueden señalar ciertas situaciones espaciales no deseables, como estancamiento demográfico, desequilibrios económicos, riesgos para la salud o desestructuración social, que son objeto prioritario de intervención desde la Administración.

## **1.2. Métodos de análisis de las agrupaciones espaciales de datos**

### **1.2.1. Origen de los principales métodos de análisis**

El desarrollo de los diversos métodos cuantitativos que conforman el AED ha sido abordado tradicionalmente por estadísticos y econométricos, dentro de los campos de conocimiento de la Geoestadística o Estadística Espacial (estadística y geografía) y de la Ciencia Regional (economía y geografía). Las diferencias entre la estadística espacial y la econometría son mínimas y ambas han aportado importantes avances en las técnicas de estudio. El matiz diferenciador viene de la especialización de la segunda en los modelos económicos (Anselin, 1988).

En el ámbito de la Ciencia Regional, A. D. Cliff y J. K. Ord, a finales de los 60 y principios de los 70, advierten de la importancia de considerar la localización espacial de los datos en los modelos espaciales para evitar errores que pusieran en cuestión su validez por no evaluar la existencia de autocorrelación espacial (Cliff y Ord, 1969, 1970, 1972). En la década de los 70, J. Paelinck sería el precursor de un nuevo campo de la Econometría, que consideraba los *efectos espaciales* en los análisis econométricos urbanos y regionales, con la publicación de la primera monografía sobre Econometría Espacial (Paelinck y Klaassen, 1979). Desde entonces se han publicado otros manuales específicos entre los que destacan el de L. Anselin (Anselin, 1988), el de A. Getis, J. Mur y H. G. Zoller (Getis, Mur, et al., 2004), el de J. LeSage y R. K. Pace (LeSage y Pace, 2009) y, en español, el de R. Moreno y E. Vayá (Moreno y Vayá, 2000). A pesar de su reconocida importancia científica, hoy en día la autocorrelación o dependencia espacial no es mencionada en muchos de los principales manuales de Econometría quedando fuera de muchos planes de estudio universitarios (Getis, 2007).

La aportación desde la estadística espacial ha sido fundamentalmente metodológica. En este sentido, la utilización del AED en un amplio espectro de disciplinas con interés en el territorio viene de la mano de la estadística espacial más preocupada en el método que en las aplicaciones concretas. El mismo Student puso de manifiesto los efectos producidos por la dependencia espacial sobre la validez de los métodos estadísticos (Student, 1914). Entre los estadísticos que han desarrollado su trabajo en el campo del AED destacan Patrick A. P. Moran, Roy C. Geary, John Keith Ord, Julian Besag o Martin Kulldorff. Estos estadísticos han desarrollado nuevos índices de autocorrelación, han perfeccionado los existentes o han introducido nuevas metodologías.

De esta manera, el AED se ha constituido como una herramienta multidisciplinar ampliamente utilizada en numerosos campos del conocimiento como la economía, epidemiología, ecología, geología, criminología, arqueología, etc. Continuamente salen a luz nuevos trabajos científicos que analizan los patrones espaciales de información de diverso origen, así como constantes avances en el desarrollo y validación de nuevos métodos de AED.

### **1.2.2. Principales métodos de AED**

Los métodos utilizados para analizar clústeres o concentraciones significativas de fenómenos en el espacio son muy diversos (Auchincloss, Gebreab, et al., 2012) y se pueden clasificar de diferentes modos. Una primera distinción sería diferenciar entre métodos univariantes y multivariantes aunque hay técnicas que se adaptan a ambas premisas (Moreno y Vayá, 2000). Los métodos también pueden ser clasificados en generales o focalizados según estén destinados a detectar dependencia espacial en un ámbito territorial o en el entorno de un punto geográfico o foco (Tango, 1995). Otras formas de clasificación de los métodos son la distinción entre métodos paramétricos y no paramétricos, o en base al tipo de unidad espacial de análisis, ya sean puntos o polígonos (Bithell, 2000).

En la práctica se observa que cada aproximación metodológica aplicada sobre un mismo conjunto de datos puede señalar patrones generales de agrupación comunes, pero también presentar diferencias notorias que dificultan la interpretación de los resultados y que llevan a recomendar la utilización simultánea de diferentes aproximaciones. La comparación de métodos de análisis de clústeres es un tema periódicamente tratado en publicaciones científicas (Bithell, 2000; Lawson, Biggeri, et al., 2000; Song y Kulldorff, 2003; Leyland y Davies, 2005; Best, Richardson, et al., 2005; Kulldorff, Song, et al., 2006; Aamodt, Samuelson, et al., 2006; Huang, Pickle, et al., 2008; Jackson, Huang, et al., 2009).

### *Autocorrelación espacial.*

Los métodos de autocorrelación espacial detectan la existencia de dependencia espacial a nivel univariante y se clasifican en globales o locales. Los métodos globales analizan toda el área de estudio permitiendo conocer si existe dependencia espacial en el conjunto de observaciones. Los indicadores globales tienen dificultades en detectar situaciones en las que existen clústeres en áreas concretas del ámbito de estudio mientras que en el resto del territorio la distribución es aleatoria. Como complemento, los métodos locales aportan el emplazamiento concreto de las agrupaciones o clústeres espaciales. En los estudios de análisis espacial se utilizan normalmente ambos tipos porque el análisis general de los primeros puede enmascarar situaciones en áreas geográficas pequeñas que pueden detectar los segundos. Por esta razón en ocasiones los métodos globales y locales pueden dar resultados diferentes sobre una misma distribución. En primer lugar se aplican los métodos globales para conocer la posible existencia de dependencia espacial en la muestra y, posteriormente, se usan los índices locales para localizar los clústeres geográficamente.

Los métodos de autocorrelación espacial global más conocidos son los índices de la *I* de Moran (Moran, 1950), la *C* de Geary (Geary, 1954) y la *G* de Getis & Ord (Getis y Ord, 1992). Los métodos locales, denominados Índices Locales de Asociación Espacial (*Local Indicators of Spatial Association-LISA*), surgieron posteriormente y los más conocidos proceden de la adaptación de los índices globales de Moran y Geary (Anselin, 1995) y Getis & Ord (Ord y Getis, 1995).

Las relaciones espaciales son consideradas en los diferentes índices mediante una matriz de pesos espaciales o distancias (*W*) que recoge las relaciones por pares de observaciones o unidades espaciales a partir de la distancia geográfica o la vecindad.

### *Inferencia Bayesiana.*

Los métodos de inferencia Bayesiana han sido aplicados exitosamente para el estudio de patrones en la disposición espacial de datos (Besag y Green, 1993; Bithell, 2000; Wakefield, Best, et al., 2000; Best, Richardson, et al., 2005). Una de sus principales ventajas es que permite minimizar el efecto producido por el tamaño de la población sobre la aparición de un determinado fenómeno. Cuando se trabaja con tasas de población éstas pueden ser muy elevadas cuando aparece un fenómeno en unidades espaciales con un número reducido de habitantes produciendo una amplia variabilidad de los resultados en diferentes momentos. El enfoque

Bayesiano suaviza el riesgo local (*smoothing*) y además introduce dos interesantes factores en el análisis. Por un lado, el conocimiento experto sobre la variable analizada a nivel global del área de estudio y, por otro, la distancia geográfica o relaciones espaciales entre las observaciones para considerar los *efectos espaciales* mediante una matriz de pesos ( $W$ ). El modelo Bayesiano más difundido en análisis espacial es el desarrollado por Besag, York y Mollie conocido como modelo BYM o *Conditional Autorregresive Model* (CAR) (Besag, York, et al., 1991; Earnest, Morgan, et al., 2007).

#### *Spatial scan statistics.*

Uno de los métodos más empleados actualmente son las diferentes versiones de las *spatial scan statistics*. El método se basa en superponer ventanas de diferente forma y tamaño sobre las diferentes localizaciones de una variable. Estas áreas son consideradas candidatas a clúster y se calcula su probabilidad de serlo en base a los casos esperados y observados dentro y fuera del área. La significación estadística se determina generando un gran número de conjuntos de datos aleatorios bajo la hipótesis nula de no estar agrupados, entonces se calcula la máxima probabilidad para cada conjunto de la misma manera que con los datos reales. Si la máxima probabilidad de los datos reales se encuentra entre el 5% más alto, el grupo es significativo al nivel de 5%. Es preciso tener en cuenta la forma y tamaño de la ventana debido a las diferentes distribuciones que pueden tener las variables de estudio, la existencia de covariables o el enorme número de cálculos que supone trabajar con un gran número de clústeres candidatos. La *spatial scan statistic* más popular es la diseñada por Kulldorf (Kulldorff y Nagarwalla, 1995), que posteriormente adaptó a ventanas circulares (Kulldorff, 1997) y a elípticas (Kulldorff, Huang, et al., 2006). Otras variantes de *spatial scan statistics* son la *Upper Level Set Scan Statistics* (ULS) donde la ventana se adapta a la distribución de los datos (Patil y Taillie, 2004), y la *Flexible Scan Statistics* (FSS) donde la forma de la ventana es irregular en base a la forma de las unidades espaciales (Tango y Takahashi, 2005).

### **1.3. Análisis de los patrones espaciales de las enfermedades**

#### **1.3.1. Epidemiología Espacial**

Uno de los campos en los que más se ha aplicado y desarrollado el análisis de los clústeres espaciales es la epidemiología espacial en su estudio de las variaciones geográficas de las enfermedades. Esta disciplina describe y analiza información sanitaria geográficamente referenciada y su relación con factores de riesgo demográficos, ambientales, del comportamiento, socioeconómicos, genéticos e infecciosos (Elliott y Wartenberg, 2004;



Rezaeian, Dunn, et al., 2007). Su origen se remonta al siglo XIX con el desarrollo de los primeros mapas de enfermedades, entre los que destaca el trabajo pionero del Dr. John Snow que relacionó la mortalidad por cólera y los pozos de agua en la epidemia que asoló la ciudad de Londres en 1854 (Snow, 1855).

La epidemiología espacial estudia múltiples variables de salud pero principalmente analiza la prevalencia (casos por habitante), la incidencia (nuevos casos por habitante) y la mortalidad (defunciones por habitante) (Bithell, 2000), generalmente en forma de tasas estandarizadas por edad y sexo debido a los efectos de la estructura de la población sobre la morbilidad (Staines y Järup, 2000; Cromley y McLafferty, 2002). Los datos epidemiológicos se pueden localizar en el espacio considerando los lugares de residencia, trabajo, atención o cualquier otro lugar que sea considerado relevante (Bithell, 2000).

Las fuentes de información proceden fundamentalmente de los registros sanitarios sobre la utilización de servicios de salud por parte de la población, las encuestas de salud a la población o los registros de defunciones. El cálculo de tasas y razones requiere conocer la población susceptible o en riesgo de experimentar el fenómeno de estudio. Estos datos pueden ser extraídos de fuentes demográficas clásicas como los censos y padrones de población (Walter, 2000).

Resulta de gran interés aplicar sobre estas variables epidemiológicas las herramientas de AED con el objetivo de detectar posibles concentraciones en el espacio (*disease clustering*), estadísticamente significativas, de valores elevados (*hot-spot*) o bajos (*cold-spot*), que indiquen la posible existencia de factores de riesgo comunes de una determinada enfermedad o de procesos de transmisión por proximidad (Wakefield, Kelsall, et al., 2000). En el primer caso, estaríamos ante procesos de reacción y en el segundo ante procesos de interacción, aunque también se pueden dar ambos conjuntamente. La reacción ocurre cuando el valor es semejante en áreas vecinas porque comparten un factor de riesgo común como los factores ambientales o sociales (Weiss y McMichael, 2004; Beale, Hodgson, et al., 2010). Y la interacción ocurre cuando el valor es similar a sus vecinos porque la proximidad facilita la transmisión, como ocurre en la transmisión de enfermedades infecciosas a partir de un foco u origen (Ostfeld, Glass, et al., 2005; Nelson y Williams, 2007).

Los resultados del análisis espacial de datos sanitarios se plasman en la realización de mapas de enfermedades (*disease mapping*) (Walter, 2000). La finalidad de estos mapas es mostrar la distribución de los casos de enfermedad o mortalidad en el territorio, descubrir patrones o concentraciones de valores altos y/o bajos, y comparar patrones de enfermedad o series

temporales (Holowaty, Norwood, et al., 2010). Estas prácticas no están exentas de errores y es necesario considerar las limitaciones inherentes a este tipo de estudios como la falacia ecológica, el problema de la unidad mínima modificable, el ajuste de tasas, la interpretación de correlaciones ecológicas o la fiabilidad de las bases de datos (Ocaña-Riola, 2010).

La capacidad de los SIG de interrelacionar diferente información espacial resulta extremadamente útil para cruzar información sanitaria con información ambiental y socioeconómica, permitiendo relacionar la enfermedad o la mortalidad con posibles factores de riesgo de un territorio (Jarup, 2004; Kistemann, Dangendorf, et al., 2002). La explicación de las causas y el comportamiento de una enfermedad se suele realizar mediante técnicas de modelización como la regresión y la clasificación espacial (Anselin y Bera, 1998; Wakefield, 2007). Como es conocido, la modelización presenta resultados que no son generalizables y que sólo son válidos para determinadas enfermedades en ciertos territorios y en un momento dado.

El conocimiento de cuando y donde se desarrolla una determinada enfermedad permite establecer hipótesis sobre sus causas ayudando a conocer mejor su etiología y a identificar sus factores de riesgo (Ward, 2008). La identificación de patrones de agrupamiento espacial en enfermedades es de gran utilidad para las políticas y la planificación en Salud Pública. Este conocimiento puede servir de apoyo a la decisión en la localización y asignación de nuevos recursos, la gestión de los existentes, la definición de acciones destinadas a enfermedades prioritarias, y el diseño de programas de prevención, vigilancia y control (Lawson, Biggeri, et al., 2000).

### **1.3.2. Análisis espacial aplicado a enfermedades**

La revisión de la literatura científica permite encontrar una gran cantidad de investigaciones de epidemiología espacial que analizan los patrones espaciales de enfermedades endógenas y exógenas mediante el AED. En otros casos, los estudios analizan la mortalidad en general, sus principales causas o determinados tipos de especial interés como el suicidio.

La enfermedad endógena más recurrente en este tipo de estudios es el cáncer y algunas de sus principales tipologías. La distribución espacial de la mortalidad por cáncer de pulmón en la provincia de Lecce (Italia) entre 1992 y 2001 ha sido analizada utilizando una variante del modelo BYM Bayesiano (Bilancia y Fedespina, 2009). También se han estudiado las relaciones entre la elevada contaminación atmosférica y la distribución de las tasas de incidencia del cáncer en Haifa (Israel) entre 1995 y 1999 (Eitan, Yuval, et al., 2010), y la distribución del

riesgo de cáncer a nivel de barrio en Ontario en el periodo 1999-2003 (Canadá) (Holowaty, Norwood, et al., 2010) a través de modelos jerárquicos Bayesianos.

En España, el Centro Nacional de Epidemiología del Instituto de Salud Carlos III ha publicado varios trabajos analizando la distribución geográfica de la mortalidad por cáncer de riñón (López-Abente, Aragonés, et al., 2008), cáncer de ovario (Lope, Pollán, et al., 2008), cáncer gástrico (Aragonés, Pérez-Gómez, et al., 2009) y cáncer de mama (Pollán, Ramis, et al., 2007) en la década de los 90 y principios del presente siglo mediante el *Modelo Condicional Autorregresivo Bayesiano* (CAR).

Otras investigaciones sobre enfermedades endógenas que han sido objeto de publicaciones científicas han sido la detección de *hot-spots* de mortalidad por la enfermedad de Creutzfeldt-Jakob en Japón entre 1995 y 2004 mediante *spatial scan statistics* (Doi, Yokoyama, et al., 2008); el análisis de agrupaciones espaciales en la incidencia de síndromes mielodisplásicos (preleucemia) en el Estado de Washington (EE.UU.) entre 2002 y 2006 utilizando también las *scan statistics* y el test de Besag & Newell (Ross et al. 2010); el estudio de la distribución geográfica de la mortalidad por Parkinson en la década de los 90 en España utilizando el modelo CAR (Pedro-Cuesta, Rodríguez-Farré, et al., 2009); y el análisis geográfico de las prescripciones de medicamentos cardiovasculares en núcleos de población de Taiwan en 2004 mediante los Índices Locales de Asociación Espacial (Cheng, Chen, et al., 2011).

En cuanto a las enfermedades exógenas, se pueden destacar varios estudios realizados mediante las *spatial scan statistics*, como la distribución espacial de las enfermedades de transmisión sexual en Kisumu (Kenia) en 2005/2006 (Westercamp, Moses, et al., 2010); la incidencia de la fiebre del Nilo Occidental en EE.UU. entre 2002 y 2008, complementándolo también con un LISA (Sugumaran, Larson, et al., 2009); o el análisis de *hot-spots* de incidencia de la tuberculosis en la isla de Montreal (Canadá) desde 1996 a 2000 (Haase, Olson, et al., 2007). Otro análisis de *hot-spot* pero mediante índices LISA se ha realizado sobre la incidencia de la malaria en el área de East Shoa (Etiopía) entre 2002 y 2006 (Yeshiwondim, Gopal, et al., 2009). En España, destaca el estudio de la incidencia de la legionelosis en el periodo 2003-2007 empleando el suavizado de tasas Bayesiano y el índice local de Moran (Gómez-Barroso, Nogareda, et al., 2011).

Por último, también es posible encontrar numerosos ejemplos de estudios que no se refieren a enfermedades concretas sino a la mortalidad por diferentes causas. Los LISA se han utilizado para relacionar la existencia de clústeres de mortalidad general con indicadores de privación en Escocia en torno al año 2001 (Sridharan, Tunstall, et al., 2007); explorar la existencia de *hot-*

*spots* de las 20 principales causas de mortalidad en Taiwan (China) durante el 2006 (Tsai, Lin, et al., 2009); y señalar los patrones espaciales de la mortalidad en Jordania durante el año 2008 (Khamis, 2010).

En esta línea, se han empleado modelos jerárquicos Bayesianos para el estudio de los patrones espaciales de la mortalidad en Andalucía (España) entre 1981 y 2006 (Ocaña-Riola y Mayoral-Cortés, 2010), y de la mortalidad por suicidio en Taiwan entre 1999 y 2007 (Chang, Sterne, et al., 2011). Las *spatial scan statistic* han sido utilizadas para analizar la distribución espacial de la mortalidad por suicidio en Queensland (Australia) entre 1999 y 2003 (Qi, Tong, et al., 2010). Por último, también se ha estudiado la variación espacial de la mortalidad infantil en Manhiça (Mozambique) durante el periodo 1997-2005 empleando un modelo de regresión espacial (Escaramis, Carrasco, et al., 2011).

### **1.3.3. Análisis espacial aplicado a enfermedades mentales**

En el ámbito de la salud mental, aunque en menor medida que en otro tipo de enfermedades, se pueden encontrar estudios que emplean diferentes métodos de análisis espacial. Por ejemplo, se han empleado modelos Bayesianos para el análisis de la relación de la pobreza y el aislamiento social con los ingresos psiquiátricos en distritos postales de Londres y Nueva York en 1998 y 2000 (Curtis, Copeland, et al., 2006); la variación de la incidencia de la esquizofrenia en áreas urbanas del sureste de Londres entre 1997 y 1999 (Kirkbride, Fearon, et al., 2007); la relación entre la hospitalización por esquizofrenia o depresión y las características socioeconómica en los condados de 14 Estados de los EE.UU. en el año 2000 (Fortney, Rushton, et al., 2007; Fortney, Xu, et al., 2009); o el estudio de la correlación entre clústeres de retraso mental y trastornos del desarrollo entre los años 1996 y 2001 (Zhen, McDermott, et al., 2009).

Las *spatial scan statistics* se han aplicado al análisis de clústeres de trastornos mentales ocasionados por consumo de sustancias psicoactivas y trastornos neuróticos, del estrés y somatomorfos en relación con las características de pobreza y desestructuración social de los barrios de Malmö (Chaix, Leyland, et al., 2006).

Los indicadores LISA han servido para analizar la distribución espacial del gasto en salud mental en Inglaterra junto a la modelización del gasto a través de operadores de regresión que tienen en consideración variables socioeconómicas junto al espacio (Moscone, Knapp, et al., 2007). También han facilitado junto a la inferencia Bayesiana la detección de *hot-spots* de prevalencia de la esquizofrenia en una pequeña área sanitaria de Granada sobre datos de 1999 (Moreno, García-Alonso, et al., 2008).

Finalmente, cabe señalar dos estudios resultado del presente trabajo. Por una parte, la detección de *hot-spots* de casos de esquizofrenia en Andalucía en el año 2004 (García-Alonso, Salvador-Carulla, et al., 2010) y, por otra, la identificación de *hot-spots* y *cold-spots* de casos de depresión en Cataluña en el 2009 (Salinas-Pérez, García-Alonso, et al., 2012).

#### **1.4. Justificación y planteamiento de la investigación**

##### **1.4.1. Relevancia de la metodología de análisis espacial**

Esta investigación pretende diseñar y probar una herramienta de AED para la identificación y localización de patrones espaciales basada en los Algoritmos Evolutivos Multi-Objetivo (AEMO). Los AEMO son herramientas destinadas a la resolución de problemas complejos, compuestos por varios objetivos que hay que optimizar para obtener soluciones eficientes desde el punto de vista del Óptimo de Pareto (Coello-Coello, Lamont, et al., 2007). Son empleados frecuentemente como herramientas de apoyo a la decisión en la que los decisores plantean y describen el problema, definen las variables que intervienen y su importancia, y establecen las posibles restricciones a su resolución.

En esta investigación se ha diseñado e implementado un Algoritmo Evolutivo Multi-Objetivo para análisis espacial (AEMO/AE) en el que el problema multi-objetivo busca la identificación de concentraciones significativas de valores de una o más variables en el espacio, ya sean de valores altos (*hot-spots*) o bajos (*cold-spots*). La herramienta ha sido probada en dos Estudios de Demostración que han empleado variables de salud mental, concretamente la prevalencia administrativa de la esquizofrenia y la depresión.

El modelo AEMO/AE fue diseñado y probado sobre datos de salud mental en Andalucía (Estudio de Demostración A) en el marco del proyecto “*Identificación, localización y análisis espacial de zonas de alta prevalencia en enfermedades con alto impacto social mediante técnicas de inteligencia artificial*” (expediente PI08/90752). Dicho proyecto fue desarrollado por INSA-ETEA bajo la dirección del Dr. Carlos Ramón García Alonso y financiado por el Instituto de Salud Carlos III y los Fondos de Desarrollo Regional de la Unión Europea durante los años 2009 y 2010. La relevancia de los resultados obtenidos ha facilitado la concesión de su continuación denominada “*Análisis espacial y de clasificación ordinal de la distribución geográfica de enfermedades mentales en Andalucía*” (expediente PI11/02008), que se está desarrollando actualmente (2012-2014).

El Estudio de Demostración B es una parte del proyecto “*Elaboración de un mapa sanitario de servicios y recursos destinados a la atención integral de personas con enfermedad mental en Cataluña, y aplicación de sistemas de información geográfica para la ayuda a la decisión en la planificación de servicios en Cataluña*” desarrollado por las Asociación Científica PSICOST, la Universidad Loyola Andalucía y la Fundació Sant Joan de Dèu para el Departament de Salut de la Generalitat de Catalunya durante los años 2010-2012.

La atención especializada de la salud mental es una de las prestaciones del sistema sanitario público de ambas Comunidades. La principal base de datos sobre la atención a la salud mental en Andalucía era el Sistema de Información de Salud Mental de Andalucía (SISMA)<sup>1</sup>. En Cataluña, la base de datos es el Conjunto Mínimo Básico de Datos de Centros de Salud Mental Ambulatorios (CMBD-SMA). La información facilitada por el SISMA y el CMBD-SMA ha supuesto una gran oportunidad para obtener datos de prevalencia *georreferenciados*, imprescindibles para analizar sus patrones espaciales y la posible existencia de *hot-spot* y/o *cold-spot*. El Programa de Salud Mental de Andalucía del Servicio Andaluz de Salud (SAS) y la Dirección General de Regulación, Planificación y Recursos Sanitarios del Departamento de Salud de Cataluña apoyaron la realización de este estudio que se realizó ajustándose escrupulosamente a la legislación sobre protección de datos.

En base a dicha información, se han planteado dos problemas de AED, diferentes aunque relacionados, que se pretenden resolver mediante el AEMO/AE como Estudios de Demostración. En ambos casos la variable de estudio es la prevalencia administrativa o número de casos o pacientes de esquizofrenia o depresión atendidos en centros de salud mental ambulatorios por 1.000 habitantes. La unidad espacial de estudio son los municipios andaluces y catalanes. La existencia de agrupaciones espaciales se plantea como un problema multi-objetivo, compuesto por una serie de objetivos y restricciones, abordable por el Algoritmo Evolutivo en la búsqueda de soluciones eficientes.

En primer lugar, se ha empleado el AEMO/AE para encontrar una solución de consenso entre los resultados de diferentes técnicas de AED para el estudio de concentraciones espaciales de la esquizofrenia y la depresión en Andalucía. Como se ha indicado anteriormente sus resultados presentan en muchas ocasiones claras diferencias por lo que sería interesante encontrar una solución de compromiso. En segundo lugar, el AEMO/AE ha analizado la existencia de puntos calientes (*hot-spots*) y puntos fríos (*cold-spots*) en la distribución geográfica de la prevalencia de la esquizofrenia y la depresión en Cataluña. Una vez identificados los *hot-spot* y *cold-spot* se

---

<sup>1</sup> Sustituida por el sistema Diraya a partir del año 2008.

<sup>2</sup> En el año 2009 la Entidad Local Autónoma Villanueva de la Concepción se segregó del Municipio de Antequera constituyendo el

ha procedido a su descripción y representación cartográfica mediante un SIG con el fin de facilitar el conocimiento de la distribución geográfica de la esquizofrenia y la depresión en Andalucía y Cataluña.

#### **1.4.2. Relevancia del estudio de las enfermedades mentales**

El análisis espacial de las enfermedades mentales, y más concretamente la detección y localización de *hot-spot* y *cold-spot* en el territorio, posee un especial interés por su fuerte impacto social y económico. En muchas ocasiones los trastornos mentales se convierten en enfermedades crónicas con un largo tratamiento que produce importantes costes directos e indirectos (Gustavsson, Svensson, et al., 2011). En España, los costes directos incluyen la asistencia sanitaria y social junto a las prestaciones por minusvalía o dependencia que son soportados por el Sistema Público, mientras que los costes de la asistencia informal recaen sobre las familias. Los costes indirectos se refieren a la pérdida de productividad por bajas laborales (permanentes o temporales) o muerte prematura (suicidios). Las enfermedades mentales supusieron la segunda causa de bajas laborales y el tercer grupo de enfermedades que conllevaron mayor gasto público en España durante el año 2002. El gasto público en este año ha sido cifrado en el 1% del Producto Interior Bruto nacional, con un montante de 7.019 millones de euros. Los costes relacionados con los días de pérdida de productividad por bajas laborales permanentes, temporales y muertes prematuras alcanzaron los 2.997 millones de euros (Oliva-Moreno, López-Bastida, et al., 2009).

Los Estudios de Demostración se han restringido a los diagnósticos psiquiátricos de esquizofrenia y depresión, que son dos de las principales enfermedades mentales por prevalencia y carga. La repercusión económica de ambos diagnósticos es substancial aunque resulta complicado cifrarla de forma real por las restricciones de información. Es posible hacerse una idea de su importancia recurriendo a los resultados de diversos trabajos de investigación. Por ejemplo en España los costes directos de asistencia médica de la esquizofrenia han sido evaluados como el 37,6% de los costes totales en salud mental en 2002 (Oliva-Moreno, López-Bastida, et al., 2009). En el caso de la depresión, se han cifrado unos costes directos e indirectos de 118 billones de euros en Europa durante el 2004 (Sobocki, Jönsson, et al., 2006). Desde el punto de vista social ambas enfermedades constituyen una de las principales causa de suicidio, como señala una investigación sobre costes de salud mental en el Reino Unido que cifró que eran responsables del 90% de este tipo de defunciones (Patel y Knapp, 1998)

La prevalencia de las enfermedades mentales en la población es objeto de numerosos estudios epidemiológicos. Un reciente estudio del *European Brain Council* (EBC) y el *European College of Neuropsychopharmacology* (ECNP), en el que se han revisado diversos estudios epidemiológicos europeos, ha estimado que el 38,2% de la población europea (164,8 millones) sufre algún trastorno mental cada año. La esquizofrenia afecta al 1,2% de la población mayor de 17 años (5 millones) y la depresión mayor al 6,9% de la población mayor de 14 años (30,3 millones) (Wittchen HU, Jacobi F, Rehm J, et al. 2011).

Anteriormente, el proyecto ESEMeD (*The European Study of the Epidemiology of Mental Disorders*) a través de la encuesta *Composite International Diagnostic Interview* (WMH-CIDI) estimó la prevalencia-año de los trastornos mentales al 9,6% de las personas mayores de 17 años y la depresión mayor al 3,9% de este grupo de población en el año 2003 (Alonso, Angermeyer, et al., 2004). Las diferencias en las cifras con el estudio anterior se debe, entre otras cosas, a que el ESEMeD no recoge algunos diagnósticos por estar poco representados en su muestra como los trastornos psicóticos (esquizofrenia), trastornos obsesivos compulsivos, trastornos de la alimentación, trastornos relacionados con el abuso de drogas y trastornos infantiles. El proyecto ESEMeD-España cifró la prevalencia al año de los trastornos mentales en el 8,5% de la población de 18 o más años. El diagnóstico más frecuente fue la depresión mayor con el 4% de la población de este grupo de edad (Haro, Palacín, et al., 2006).

En España, la Encuesta Nacional de Salud (ENSE) de 2006 arroja que el 13,8% de la población española mayor de 16 años declara padecer depresión, ansiedad u otro trastorno mental diagnosticado por un médico en los últimos 12 meses. La cifra para Andalucía es del 14,8% y para Cataluña del 13,4%. Esta encuesta no ofrece datos por diagnósticos específicos.

La Evaluación de la Estrategia en Salud Mental del Sistema Nacional de Salud (Ministerio de Sanidad y Política Social, 2010) recopila datos de morbilidad atendida en los servicios ambulatorios de salud mental por Comunidades Autónomas. Según el registro de casos andaluz en el año 2007, la morbilidad atendida por esquizofrenia fue del 0,232% y por depresión mayor el 0,031% de la población. En el caso de Cataluña, la prevalencia asistida de la esquizofrenia fue de 0,231% y la depresión mayor de 0,694%. El total nacional para la esquizofrenia fue del 0,198% y del 0,278% para la depresión mayor. Como se puede comprobar, la morbilidad asistida de la depresión en Andalucía presenta un valor muy bajo respecto a la prevalencia nacional y catalana sin que el citado documento explique la razón.



### **1.4.3. Estructura y organización del trabajo de investigación**

Este trabajo de investigación se estructura en siete grandes epígrafes. Tras esta Introducción (Capítulo I) se exponen las hipótesis y se definen los objetivos generales y específicos del estudio (Capítulo II). A continuación, en la Metodología (Capítulo III) se presenta el diseño del trabajo y de los Estudios de Demostración, las bases de datos empleadas, las herramientas de estadística espacial, una introducción a los Algoritmos Evolutivos, las limitaciones del estudio y la revisión de la producción científica derivada de este trabajo. El diseño del Algoritmo Evolutivo Multi-Objetivo para Análisis Espacial (AEMO-AE) y los resultados de los dos Estudios de Demostración se presentan en el epígrafe de Resultados (Capítulo IV). En el epígrafe Discusión se discute sobre los resultados obtenidos (Capítulo V) y, finalmente, se exponen las Conclusiones de este trabajo (Capítulo VI). Todas las referencias bibliográficas del texto quedan recogidas de forma exhaustiva en la Bibliografía (Capítulo VII).

## II. HIPÓTESIS Y OBJETIVOS

---

### 2.1. Hipótesis

1. Los Algoritmos Evolutivos Multi-objetivo (AEMO), como herramientas para la resolución de problemas con múltiples objetivos, pueden ser aplicados eficazmente al análisis espacial de datos de variables sanitarias.
2. El modelo de análisis espacial basado en un AEMO (AEMO/AE) permite obtener una solución de consenso de los resultados de varios índices estadísticos de análisis de la dependencia espacial aplicados sobre una variable sanitaria (prevalencia administrativa).
3. El modelo de análisis espacial AEMO/AE permite identificar y localizar concentraciones o clústeres espaciales de valores significativamente elevados (*hot-spot*) y bajos (*cold-spot*) de una variable sanitaria (prevalencia administrativa).

### 2.2. Objetivo general

Diseño de un Algoritmo Evolutivo Multi-Objetivo de Análisis Espacial (AEMO/AE) de variables sanitarias; y la prueba del modelo mediante dos Estudios de Demostración sobre la distribución municipal de la prevalencia administrativa de la esquizofrenia y la depresión en las Comunidades Autónomas de Andalucía y Cataluña.

### 2.3. Objetivos específicos

1. Diseñar un Algoritmo Evolutivo Multi-Objetivo para el análisis espacial de variables sanitarias (AEMO/AE).
  - a. Planteamiento y programación del AEMO/AE.
  - b. Selección de parámetros para el funcionamiento del AEMO/AE.
2. Estudio de Demostración A. Aplicar el modelo AEMO/AE para buscar una solución de consenso entre cuatro métodos de análisis espacial aplicados sobre la prevalencia administrativa de la esquizofrenia (Caso 1) y la depresión (Caso 2) en la Comunidad Autónoma de Andalucía en los años 2004, 2006, 2007 y 2008.

- a. Planteamiento del problema multi-objetivo para su resolución por el AEMO/AE.
  - b. Análisis de la dependencia espacial de la prevalencia administrativa de la esquizofrenia y la depresión mediante 3 índices locales de asociación espacial (LISA).
    - i. Análisis de la dependencia espacial mediante el índice local de Moran ( $I$ ).
    - ii. Análisis de la dependencia espacial mediante el índice local de Geary ( $C$ ).
    - iii. Análisis de la dependencia espacial mediante el índice local de Getis & Ord ( $G^*$ ).
  - c. Análisis de la dependencia espacial de la prevalencia administrativa de la esquizofrenia y la depresión mediante un modelo de inferencia Bayesiana.
    - i. Análisis de la dependencia espacial mediante un Modelo Condicional Autorregresivo ( $CAR$ ).
  - d. Identificación de *hot-spots* de dependencia espacial de la prevalencia administrativa de la esquizofrenia y la depresión mediante el modelo AEMO/AE.
3. Estudio de Demostración B. Aplicar el modelo AEMO/AE para buscar *hot-spots* y *cold-spots* de prevalencia administrativa de la esquizofrenia (Caso 3) y la depresión (Caso 4) en la Comunidad Autónoma de Cataluña en el año 2009.
- a. Planteamiento del problema multi-objetivo para su resolución por el AEMO/AE.
  - b. Identificación de *hot-spots* y *cold-spots* de prevalencia administrativa de la esquizofrenia y la depresión mediante el modelo AEMO/AE.

## III. METODOLOGÍA

---

### 3.1. Diseño

La finalidad de esta investigación es el desarrollo de una herramienta de análisis espacial de datos basada en técnicas de inteligencia artificial y destinada a la identificación y localización de agrupaciones estadísticamente significativas de datos espaciales. La metodología ha sido probada y evaluada mediante dos Estudios de Demostración sobre datos epidemiológicos referidos a los pacientes con esquizofrenia y depresión atendidos en los servicios públicos especializados de salud mental de tipo ambulatorios de Andalucía y Cataluña (Figura 1).

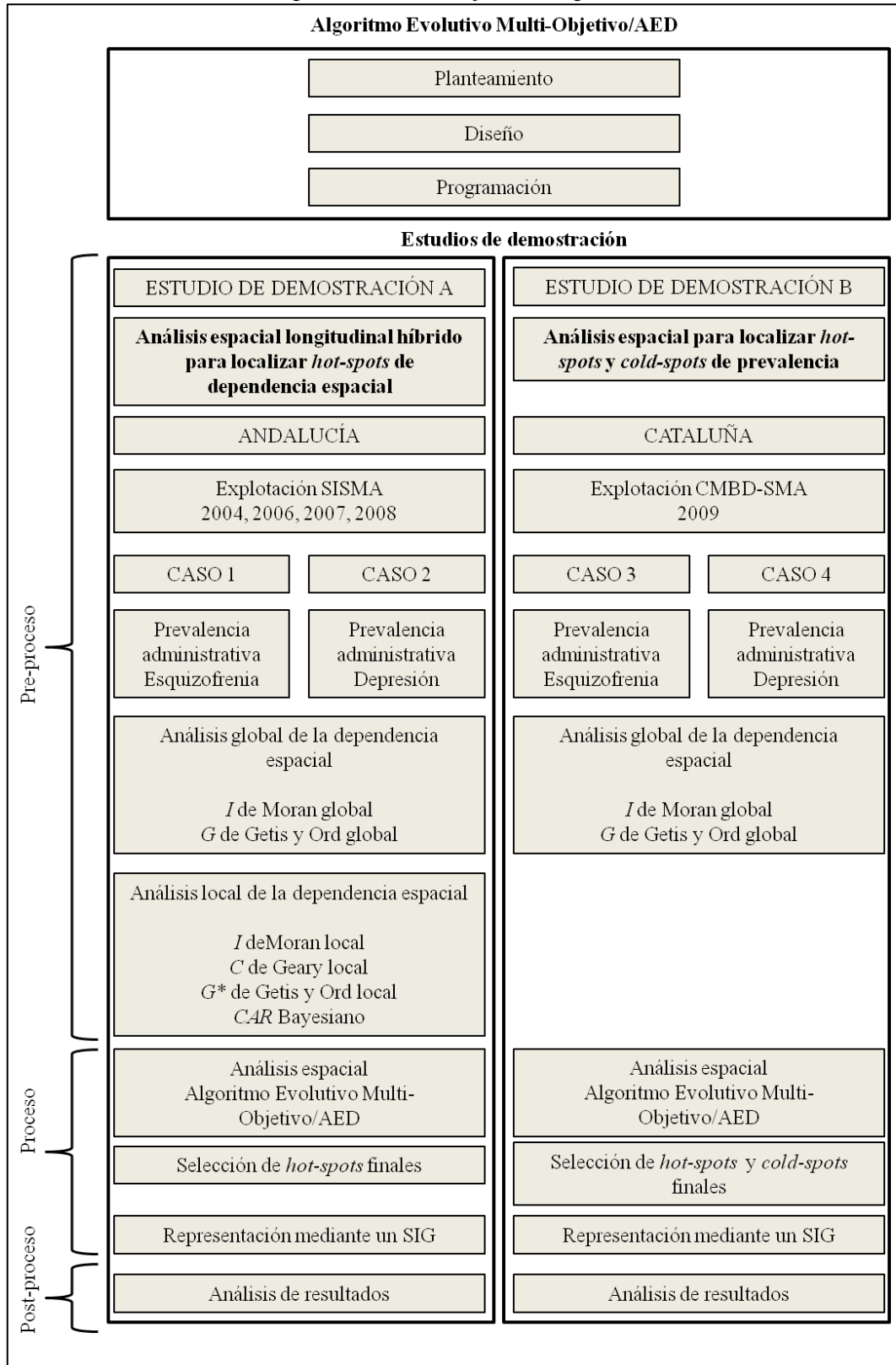
Los Estudios de Demostración en el campo de la salud se dirigen a evaluar la viabilidad técnica de un nuevo o mejorado procedimiento, método, tecnología o sistema destinado a abordar diferentes problemas de salud. La implementación de la nueva técnica se realiza en un lugar real donde se define una medida del problema a solucionar (en este caso la prevalencia) y una medida de seguimiento del problema (el modelo AEMO/AE) y se evalúan los beneficios de llevarla a cabo (National Institute for Occupational Safety and Health, 1993).

Los dos Estudios de Demostración pretenden evaluar la viabilidad técnica del Algoritmo Evolutivo Multi-Objetivo de Análisis Espacial (AEMO/AE) para detectar y localizar concentraciones espaciales de valores significativamente elevados y/o bajos de una o más variables. La herramienta se ha aplicado sobre variables sanitarias, en este caso, los datos municipales de prevalencia administrativa de la esquizofrenia y la depresión en Andalucía y Cataluña. El municipio es la unidad territorial mínima más precisa sobre la que es posible *georreferenciar* la información disponible. Los resultados han sido evaluados por grupos de expertos en salud mental, planificación y gestión de servicios de ambas Comunidades Autónomas.

Los Estudios de Demostración se han planteado como estudios epidemiológicos de tipo ecológico, en el que la *población de estudio* son los pacientes con esquizofrenia y depresión atendidos en servicios de salud mental públicos de tipo ambulatorio que residen en cualquiera de los municipios andaluces y catalanes. Los estudios ecológicos se caracterizan por ser exploratorios y generadores de hipótesis y, por tanto, diferentes a los estudios etiológicos explicadores de hipótesis (Borja-Aburto, 2000).

Los Estudios de Demostración han probado el modelo de análisis espacial en dos tipos de estudios diferentes que han permitido probar su viabilidad.

Figura 1: Fases del trabajo de investigación.



### **3.1.1. Estudio de Demostración A**

*Variables de análisis:* El AEMO/AE se ha aplicado sobre cuatro índices que evalúan la existencia de dependencia o autocorrelación espacial en la distribución territorial de la prevalencia administrativa de la esquizofrenia (Caso 1) y la depresión (Caso 2). Se pretende de esta forma encontrar una solución de consenso entre los cuatro métodos que señale áreas geográficas donde exista una elevada dependencia espacial de la prevalencia (*hot-spots*).

*Ámbito espacial de análisis:* La Comunidad Autónoma de Andalucía que se organiza administrativamente en 8 provincias.

*Unidades espaciales de análisis:* Los 770 municipios<sup>2</sup> andaluces que se organizan administrativamente en 71 pequeñas áreas de atención a la salud mental.

*Marco temporal:* Estudio longitudinal en el que la serie temporal ha sido seleccionada considerando la disponibilidad de las bases de datos sanitarias y que eran referidas a los años 2004, 2006, 2007 y 2008. Su finalidad es detectar la persistencia o variabilidad de las concentraciones espaciales en el tiempo.

### **3.1.2. Estudio de Demostración B**

*Variable de análisis:* El AEMO/AE ha buscado concentraciones espaciales de valores significativamente elevados (*hot-spot*) o bajos (*cold-spot*) de prevalencia administrativa de la esquizofrenia (Caso 3) y la depresión (Caso 4). En este Estudio B se ha analizado la distribución geográfica de la prevalencia administrativa por sí misma sin el empleo de índices de AED.

*Ámbito espacial de análisis:* La Comunidad Autónoma de Cataluña que se divide administrativamente en 4 provincias.

*Unidad espacial de análisis:* Los 946 municipios catalanes que se organizan administrativamente en 74 pequeñas áreas de atención a la salud mental.

*Marco temporal:* Estudio de un único año, concretamente el año 2009.

---

<sup>2</sup> En el año 2009 la Entidad Local Autónoma Villanueva de la Concepción se segregó del Municipio de Antequera constituyendo el municipio número 771 de Andalucía. Este nuevo municipio no ha sido considerado en los análisis debido a que no existía durante el marco temporal definido.

## 3.2. Creación de una nueva herramienta de Análisis Espacial de Datos

### 3.2.1. Los Algoritmos Evolutivos y la resolución de problemas multi-objetivos

Los Algoritmos Evolutivos Multi-Objetivo (AEMO), dentro del ámbito de la inteligencia artificial, se han confirmado en los últimos años como una potente y robusta metodología de optimización metaheurística sin restricciones relativas a su estructura algebraica. Son especialmente útiles cuando los métodos clásicos son inoperantes porque demandan demasiada computación o amplios periodos de tiempo para resolver el problema (Bäck, 1996; Coello-Coello, Lamont, et al., 2007). Los AEMO pueden ser empleados para obtener la máxima aproximación al óptimo de eficiencia de Pareto cuando se tiene que optimizar funciones objetivo no lineales para resolver un problema multi-objetivo de decisión.

Un AEMO es un tipo de Algoritmo Genético (AG) aplicado a la resolución de problemas multi-objetivo. El funcionamiento de los AG se basa en los mecanismos de selección de la Naturaleza en el que sobreviven los individuos más aptos o aquellos que se adaptan más fácilmente a los cambios de su entorno. Este funcionamiento natural fue planteado por R. S. Rosenberg (Rosenberg, 1970) y llevado al mundo de las computadoras por J. Holland a finales de la década de los 60 con la finalidad de que las computadoras aprendieran por sí mismas (Holland, 1975).

La principal aplicación que se le ha dado a los AG es la búsqueda de soluciones a problemas de optimización multi-objetivo, también denominada optimización vectorial u optimización con criterios múltiples. Una completa definición de un problema de optimización multi-objetivo fue dada por A. Osyczka:

*“El problema de encontrar un vector de variables de decisión que satisfaga las restricciones y optimice una función vectorial cuyos elementos representen las funciones objetivo. Estas funciones forman una descripción matemática de los criterios de desempeño que usualmente están en conflicto entre sí. Por lo tanto, el término “optimizar” significa encontrar una solución tal que proporcione valores para todos los objetivos que resulten aceptables para el diseñador” (Osyczka, 1984).*

La optimización trata de encontrar la mejor solución entre todas las soluciones posibles, también conocidas como *óptimo global*, que constituyen el *espacio de búsqueda* o *espacio de soluciones*. El concepto de óptimo no es universal por lo que no es posible establecer comparaciones directas entre las diferentes metodologías de resolución de problemas multi-

objetivo. Además, la solución final siempre debe ser resultado de la decisión subjetiva de un experto.

### 3.2.2. Elementos de un AEMO

#### *Variables de decisión y restricciones*

Al abordar un problema multi-objetivo se busca optimizar un conjunto de *variables de decisión* referidas a un problema determinado que se quiere resolver satisfaciendo ciertas *restricciones*. Las *variables de decisión* y las *restricciones* de un problema multi-objetivo son seleccionadas siguiendo los criterios de los decisores en base a su conocimiento experto del problema. Las *variables de decisión* constituyen el vector  $x$  que marcará una solución:

$$\text{Variables de decisión} \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (1)$$

Por su parte, las *restricciones* a la optimización proceden de las características del medio o la disponibilidad de recursos (límites al consumo de recursos, tiempo, espacio...). Las soluciones aceptables deben considerar y satisfacer estas *restricciones*. Las *restricciones* pueden ser de desigualdad o igualdad:

$$\text{Restricciones de desigualdad} \quad g_i(\vec{x}) \leq 0 \quad i = 1, 2, \dots, m \quad (2)$$

$$\text{Restricciones de igualdad} \quad h_j(\vec{x}) = 0 \quad j = 1, 2, \dots, p \quad (3)$$

El número de *restricciones* de igualdad ( $p$ ) debe ser inferior al número de *variables de decisión* ( $n$ ) para evitar *sobre-restricciones* ( $p < n$ ). Las *restricciones* pueden ser explícitas o implícitas dependiendo de si se conocen o no. Las *restricciones* definen la *región factible* donde encontrar las *soluciones factibles* al problema multi-criterio.



### ***Funciones objetivo***

Las *variables de decisión* que se pretenden optimizar se definen mediante las *funciones objetivo*. Las funciones pueden ser *comparables* o *no comparables* dependiendo de si se expresan en la misma unidad de medida. Las diferentes *funciones objetivo* que se proponen optimizar para la resolución de un problema componen la *función vector*  $\vec{f}(x)$ :

$$\text{Funciones objetivo} \quad \vec{f}(x) = \begin{bmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_k(\vec{x}) \end{bmatrix} \quad (4)$$

La relación entre las *funciones objetivo* puede ser *independiente* si al mejorar o empeorar cualquiera de ellas no se produce ningún efecto sobre las demás; *armónica* si al mejorar o empeorar una de ellas provoca el mismo efecto en las otras; o *conflictiva* si al mejorar unas provoca el empeoramiento de las otras y viceversa. En la mayor parte de las ocasiones las *funciones objetivo* estarán en conflicto entre sí, por ejemplo unas deberán ser maximizadas mientras que por el contrario otras requerirán ser minimizadas. Por esta razón, la optimización busca encontrar soluciones aceptables o de consenso para todas las *funciones objetivo* del problema.

La optimización multi-objetivo busca un conjunto de soluciones próximas a la eficiencia ya que no existe una única solución a un problema multi-objetivo compuesto de varias *funciones objetivo*. La optimización de las distintas funciones se realizará de forma simultánea requiriendo maximizar y/o minimizar las mismas. Las *funciones objetivos* frecuentemente necesitan re-escalarsse o cambiar de signo en aras de evitar valores negativos y diferencias en las escalas de medida. En resumen, un problema multi-objetivo consiste en optimizar (maximizar y/o minimizar) un grupo de *variables de decisión* ( $n$ ) reflejadas en varias *funciones objetivo* ( $k$ ) considerando ciertas *restricciones* ( $m+p$ ).

### ***Óptimo de Pareto***

Las soluciones son encontradas siguiendo la Teoría del Óptimo de Pareto planteada por V. Pareto en el siglo XIX (Pareto, 1896; Ehrgott, 2005) por la cual las *funciones objetivo* de las

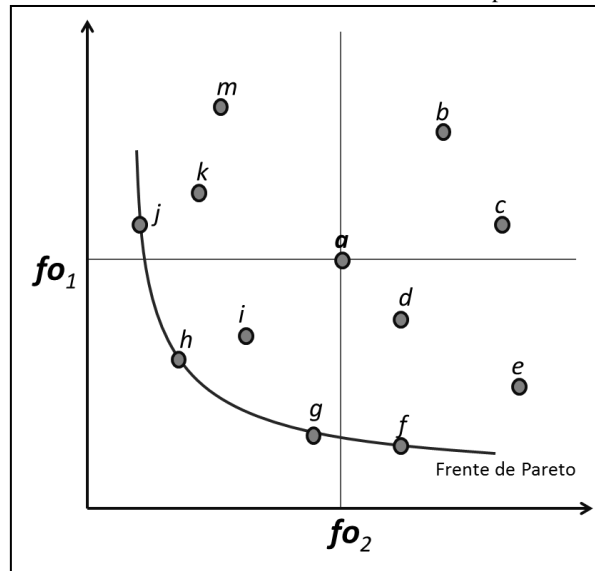
soluciones no dominarán unas sobre otras. Un vector  $\vec{u}$  dominará a otro  $\vec{v}$  ( $u \leq v$ ) si y solo si  $u$  es parcialmente menor que  $v$ :

$$\forall i \in \{1, \dots, k\}, u_i \leq v_i \wedge \exists i \in \{1, \dots, k\}: u_i < v_i \quad (5)$$

En la Figura 2 se representa el concepto de dominancia de Pareto. Un vector domina a otro si al menos uno de sus objetivos es mejor y en el resto de objetivos no es peor; y es dominado si al menos alguno de sus objetivos es peor que el del otro vector aunque el resto sea igual. En cambio, dos vectores no son comparables en términos de dominancia si algunos objetivos son mejores y otros peores. Los vectores que no son dominados por ningún otro son aquellos que se encuentran en el Óptimo o Frente de Pareto en el que no existen soluciones mejores que otras. En la siguiente expresión el Frente de Pareto es  $FP^*$ ,  $P^*$  es un conjunto de Óptimos de Pareto y  $\vec{f}(x)$  es la *función objetivo* a optimizar:

$$FP^* := \{ \vec{f} = [f_1(x), \dots, f_k(x)] \mid x \in P^* \} \quad (6)$$

Figura 2: Dominancia de soluciones en la Teoría del Óptimo de Pareto.



Fuente: Elaboración propia.

En un problema de minimización  $a$  domina a  $b$  y  $c$ ; y es dominada por  $i$ ,  $h$  y  $g$ . La dominancia no es comparable con  $a$  para los vectores  $d$ ,  $e$ ,  $f$ ,  $j$ ,  $k$  y  $m$ . Los vectores  $f$ ,  $g$ ,  $h$  y  $j$  no se dominan entre sí al situarse en el Frente de Pareto.

En la búsqueda del *óptimo global* (óptimo de todas las variables en juego), el incremento de una *función objetivo* no supondrá el decrecimiento de otra y viceversa. Así, una solución ( $x$ ) será Óptimo de Pareto si y solo si no existe otro vector ( $x'$ ) cuya función ( $F(x')$ ) domina a la función de  $x$  ( $F(x)$ ). O la solución  $x$  no será Óptimo de Pareto si existe un vector  $x'$  que al mejorar uno de los objetivos provoque el empeoramiento de otros objetivos.

### ***La intervención del decisor en los AEMO***

Como las soluciones del Frente de Pareto son igualmente óptimas, debe ser un decisor el que seleccione las soluciones finales entre éstas en base a su conocimiento experto del problema. Las técnicas multi-objetivo, y los algoritmos evolutivos como parte de ellas, se pueden clasificar en tres categorías atendiendo a cuando se realiza la decisión, si antes, durante o después de la optimización (Cohon y Marks, 1975).

- “A priori”: el decisor tiene que definir la relevancia relativa de cada objetivo antes de resolver el modelo. Las preferencias de decisor son modeladas para evaluar y comparar soluciones. Aproximaciones evolutivas “a priori” son por ejemplo la ordenación lexicográfica (*lexicographic ordering*), funciones de agregación lineal (*linear aggregating functions*) o funciones de agregación no lineal (*nonlinear aggregating functions*).
- Progresivas: los decisores incorporan sus preferencias durante el proceso de forma interactiva. No existen técnicas evolutivas de este tipo.
- “A posteriori”: exploran el espacio de variables tratando de obtener el número máximo de elementos diferentes del conjunto de Pareto que sea posible. El decisor en este caso actuará una vez completada la búsqueda de soluciones. Técnicas evolutivas de este tipo son el muestreo independiente (*independent sampling techniques*), la selección por criterios (*criterion selection techniques*), la selección agregativa (*aggregation techniques*), las restricciones ( *$\epsilon$ -constraint techniques*) o las técnicas basadas en Pareto (*Pareto sampling techniques*).

Las técnicas “a priori” presentan el inconveniente de que la elección de los pesos de los objetivos es subjetiva y pueden perderse soluciones relevantes al priorizar arbitrariamente los objetivos restringiendo el *espacio de búsqueda*. Las técnicas progresivas precisan que el decisor establezca metas y preferencias para influir en la búsqueda de soluciones y además evalúe la

validez de cada conjunto de soluciones, lo que puede ralentizar y dificultar el proceso cuando se conoce poco acerca del problema. También incorpora el elemento subjetivo de la preferencia del decisor. Por último, en las técnicas “a posteriori” el decisor interpreta el espacio de soluciones una vez obtenido por lo que no introduce la posible subjetividad en el proceso. En principio, ninguno de estos tres enfoques es mejor que los demás para la resolución de un problema multi-objetivo (Wolpert y Macready, 1997).

### ***Funciones de aptitud***

Las soluciones se evalúan a través de una *función de aptitud* (*fitness function*) que valora la eficiencia o el óptimo de las *funciones objetivo* del problema que se plantea resolviendo representando el compromiso entre ellas. Las *funciones de aptitud* permiten identificar o discriminar las mejores soluciones para aplicar sobre ellas el operador genético de selección y ser mejoradas, generación a generación, mediante los operadores de recombinación y mutación hasta alcanzar el óptimo. En un problema mono-objetivo la *función de aptitud* se identifica con la *función objetivo*, en cambio, en los problemas multi-objetivos la *función de aptitud* debe considerar las diferentes dimensiones de cada *función objetivo*.

Normalmente se emplea la dominancia de Pareto para valorar la aptitud de las soluciones pero, como se ha señalado anteriormente, existen dificultades cuando el número de objetivos a optimizar es alto. Ante estos problemas normalmente se recurre a reducir dimensiones eliminando *funciones objetivo* con un comportamiento armónico que resulte redundante; o se plantea un método alternativo de búsqueda de soluciones que sea más estricto (*grano fino*).

Las *funciones de aptitud* son complejas y demandan un gran esfuerzo de computación. Algunos ejemplos de *funciones de aptitud* son las funciones de creación de "rankings" de objetivos, de ponderación de objetivos, los modelos especiales tipo SPEA2 o la inferencia difusa – lógica difusa (Coello-Coello, Lamont, et al., 2007).

### **3.2.3. Funcionamiento de un AEMO**

Un AEMO funciona mediante un proceso iterativo que permite la mejora sucesiva de las soluciones analizadas, mientras evita *óptimos locales*, hasta que éstas no se pueden mejorar más o se alcanza un criterio de parada. Para la determinación del conjunto de soluciones factibles y no dominadas o eficientes desde el punto de vista de Pareto, los AEMO utilizan los operadores genéticos clásicos: selección, recombinación y mutación (Osyczka, 2002; Deb, 2004) (Figura 3).

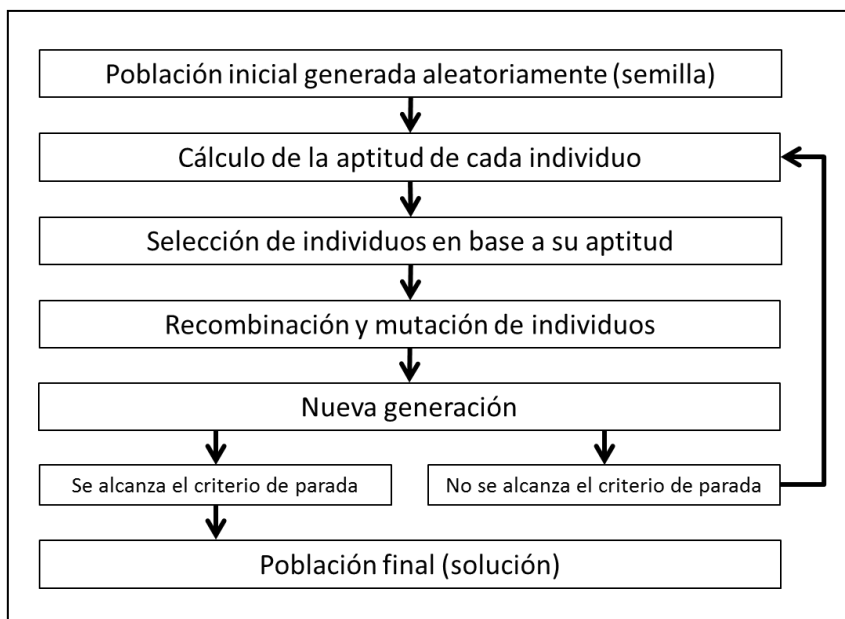
Figura 3: Esquema de un Algoritmo Evolutivo.

```

t := 0
inicializar P(0) := {a1(0), ..., aμ(0)} ∈ Iμ(0);
while (i({P(0), ..., P(t)}) ≠ verdadero do
    recombinar : P'(t) := rθr(t)(P(t));
    mutar : P''(t) := mθm(t)(P'(t));
    seleccionar :
if χ
    then P(t+1) := s(θs(t), φ)(P''(t));
    else P(t+1) := s(θs(t), φ)(P''(t) ∪ P(t));
fi
    t : t + 1
od
    
```

Fuente: (Coello-Coello, Lamont, et al., 2007)

Figura 4: Fases del algoritmo Evolutivo.

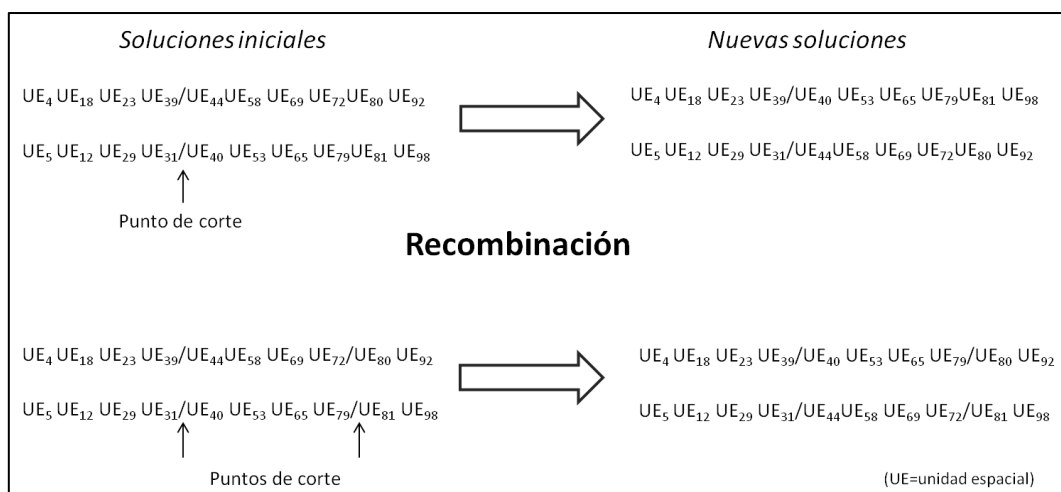


Fuente: elaboración propia.

El procedimiento comienza con la fijación de un número determinado de soluciones o *semillas* (*población inicial*) compuestas por grupos de soluciones aleatorias o seleccionadas mediante métodos determinísticos (*individuos*) (Figura 4). A través de un proceso iterativo las *semillas* son mejoradas generación a generación mediante los operadores genéticos o evolutivos:

- Selección: mediante este operador se eliminan aquellas soluciones con valores bajos de la *función de aptitud* y el resto son seleccionadas para ser mejoradas mediante los siguientes operadores genéticos. La selección se puede realizar mediante múltiples métodos, siendo los más comunes la ruleta (las soluciones con mayor aptitud tienen mayor probabilidad de ser seleccionadas), el torneo (se comparan los valores de aptitud por pares de soluciones y se selecciona el mayor) o la categoría (las soluciones se ordenan por su valor de aptitud y se seleccionan aquellas con mayores puntuaciones). Las soluciones seleccionadas para ser eliminadas son sustituidas por las ganadoras manteniéndose constante el número de individuos de la *población inicial* para la siguiente generación.
- Recombinación o cruzamiento: se mezclan los componentes de dos soluciones diferentes, mediante uno o dos puntos de corte, para crear una nueva solución en la que se mejore el valor de la *función de aptitud* (Figura 5). La recombinación explora las regiones más prometedoras del *espacio de búsqueda* al combinar dos soluciones.

Figura 5: Funcionamiento del operador genético de recombinación o cruzamiento.



Fuente: elaboración propia.

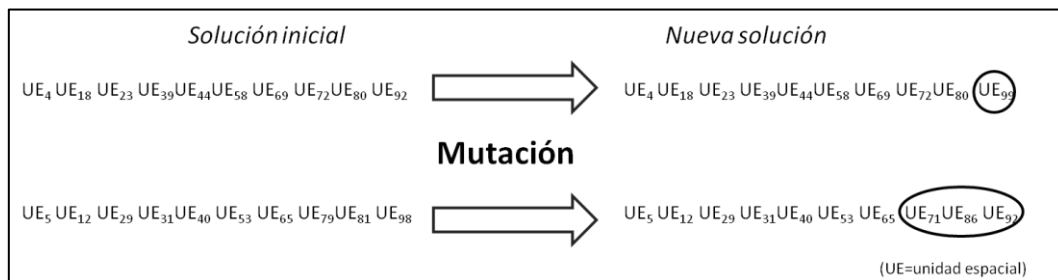
La selección de las soluciones para su recombinación es aleatoria y se ajusta mediante un porcentaje o probabilidad sobre el total. Al tratarse de una probabilidad pueden existir generaciones en la que no se efectúe la recombinación.

- Mutación: se sustituyen uno o más componentes de una solución buscando mejorar la puntuación de la *función de aptitud* (Figura 6). Con la mutación se consigue evitar una

convergencia rápida entre soluciones en un óptimo local manteniendo su diversidad al explorar nuevas regiones del *espacio de búsqueda*.

La selección de las soluciones de una generación que van a mutar se realiza aleatoriamente mediante un porcentaje o probabilidad sobre el total. La mutación ocurre más esporádicamente que la recombinación por lo que este porcentaje será muy inferior al establecido para ésta. La mutación puede no existir en todas las generaciones en función de esta probabilidad de ocurrencia. La selección de soluciones a mutar también puede ser dirigida atendiendo a ciertos criterios que mejoren su valor de *aptitud*. Hay diversos métodos de mutación entre los que destacan las permutaciones, la programación genética o la representación real.

Figura 6: Funcionamiento del operador genético de mutación.



Fuente: elaboración propia.

Es además muy frecuente, aunque opcional, considerar la existencia de cierto elitismo mediante la supervivencia de las mejores soluciones (valor alto de la *función de aptitud*) que son separadas para que ni muten ni se recombinen. De esta forma las mejores soluciones prevalecen inalteradas hasta que aparezca alguna mejor. Con esto se consigue que el valor de *aptitud* máximo de las soluciones no disminuya de una generación a la siguiente.

El proceso iterativo se detiene mediante el establecimiento de criterios de parada obteniéndose la *población* o soluciones finales para cada *semilla*. De esta forma, se puede fijar un número máximo de generaciones que cuando se alcanza detiene al algoritmo, esperar al momento en que las soluciones no mejoran significativamente sus valores de *aptitud* de una generación a otra, o se pueden combinar ambos criterios.

Los parámetros de probabilidad de los operadores genéticos, criterios de parada, número de generaciones, tamaño de la *población*, número de objetivos, existencia de elitismo, tipo de

selección, etc., son fijados inicialmente y condicionarán de forma crucial el comportamiento del algoritmo.

### **3.3. Estudios de Demostración**

#### **3.3.1. Fuentes de información**

##### **3.3.1.1. El Sistema de Información de Salud Mental de Andalucía**

###### **3.3.1.1.1. Descripción del SISMA**

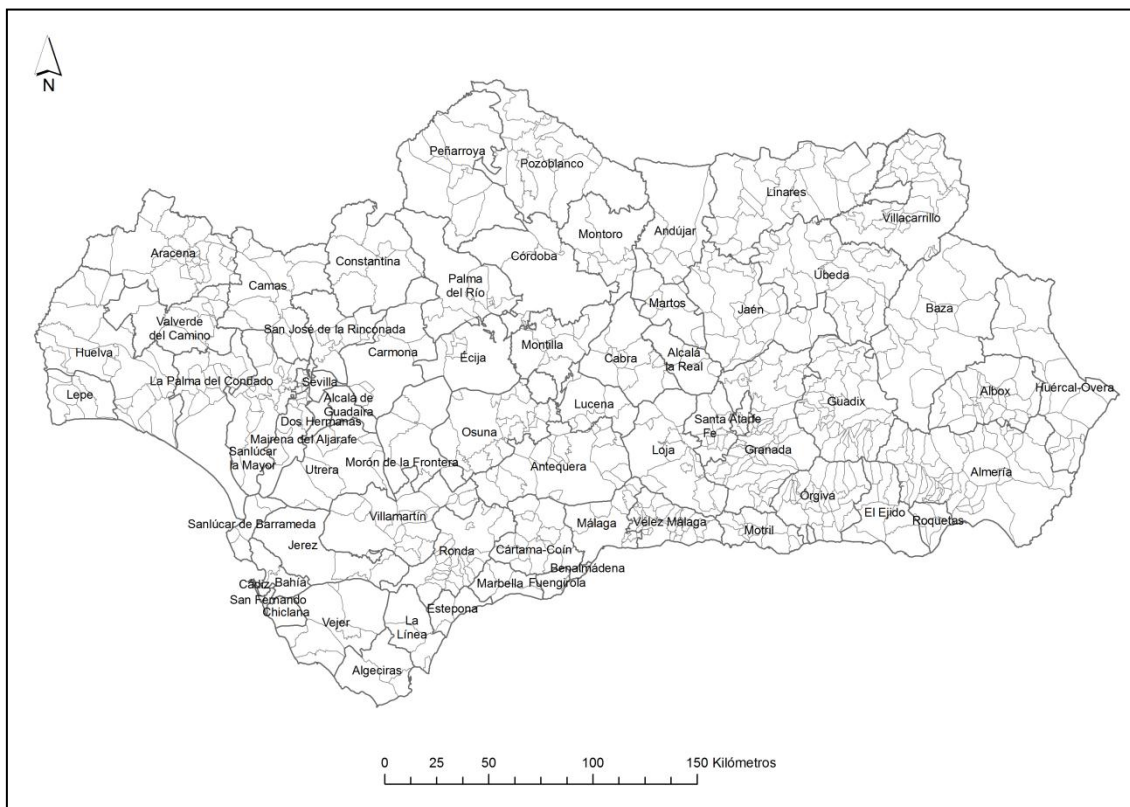
El Sistema de Información de Salud Mental de Andalucía (SISMA) (Consejería de Salud, 2003) ha sido la base de datos que recopilaba la actividad asistencial realizada en las Unidades de Salud Mental Comunitaria (USMC) del Servicio Andaluz de Salud (SAS) hasta el año 2008 en el que fue sustituido por el sistema Diraya (González Cocina y Pérez Torres, 2007; Consejería de Salud, 2008a). El Programa de Salud Mental de Andalucía facilitó información del SISMA para los años 2004, 2006, 2007 y 2008. No se ha incluido el año 2005 debido a problemas con la calidad de los datos. Esta investigación ha salvaguardado el anonimato de los pacientes mediante su identificación con un código numérico.

Las USMC constituyen el dispositivo básico de la red de atención especializada de salud mental del SAS. A estas Unidades son remitidos aquellos casos que no pueden ser atendidos desde la atención primaria según el *Decreto 77/2008 de ordenación administrativa y funcional de los servicios de Salud Mental en el ámbito del Servicio Andaluz de Salud* (Consejería de Salud, 2008b). Las USMC prestan atención comunitaria de tipo ambulatorio y domiciliario a las personas con enfermedad mental de cualquier edad, aunque determinados casos en población infanto-juvenil se derivan a las Unidades de Salud Mental Infanto-juveniles. Cada una de las USMC tiene asignada un área de cobertura compuesta por varios municipios o, en el caso de las grandes capitales (Sevilla, Málaga, Córdoba y Granada), por distritos urbanos.

En el Mapa 1 se representa la asignación de las unidades espaciales de estudio (municipios) a las USMC. Las capitales señaladas anteriormente son atendidas por más de una Unidad cuya delimitación inframunicipal no se representa en el mapa por no constituir la escala de este trabajo.



Mapa 1: Áreas de cobertura de las Unidades de Salud Mental Comunitaria de Andalucía (770 municipios).



Fuente: Programa de Salud Mental de Andalucía, 2008

El SISMA recoge información sobre los diferentes tipos de actividades realizadas en las USMC, ya sean asistenciales o no, y sobre los pacientes atendidos. Los registros de la base de datos que no presentan pacientes asociados corresponden a actividades propias de los profesionales sanitarios como reuniones de coordinación, acciones de formación, asesoramiento a otros profesionales, etc. Las actividades con pacientes (visitas) pueden ser de diverso tipo como rehabilitación, administración de fármacos, entrevistas de acogida, etc. En estas últimas se identifica al paciente y se recogen datos sociodemográficos obtenidos en la entrevista de acogida como la edad, el sexo, el municipio de residencia, la ocupación, el nivel educativo, etc., y también datos sanitarios como el motivo de alta, el tipo de profesional que lo ha atendido, el diagnóstico principal, secundario, etc. En la Tabla 1 se presentan las cifras globales de la serie temporal analizada.

Tabla 1: Actividades y pacientes recogidos en el Sistema de Información de Salud Mental de Andalucía

	2004	2006	2007	2008
Actividades totales	1.030.858	1.131.020	1.184.242	1.085.525
Actividades con pacientes	892.844	983.952	1.027.996	950.823
Pacientes	192.904	198.955	201.977	195.522

Fuente: SISMA, 2004, 2006, 2007 y 2008

Durante este periodo se encontraban en funcionamiento 72 USMC, a las que en el último año se añadieron 3 nuevas hasta llegar a las 75 USMC. Como limitaciones de la base de datos cabe señalar la ausencia de datos de la USMC de Antequera durante los cuatro años de la serie temporal debido a problemas en la comunicación de datos al Programa de Salud Mental. Las USMC de Santa Fe, Sanlúcar, Villamartín, Jerez y Cádiz tampoco presentan registros para el año 2008 debido al proceso de sustitución progresiva del SISMA por el Diraya.

### 3.3.1.1.2. Explotación del SISMA

Las variables que se han analizado en este estudio han sido la edad y sexo del paciente, su diagnóstico y su municipio de residencia de cara a la *georreferenciación* de la información en las unidades espaciales de estudio. Se han analizado los pacientes con diagnóstico principal de esquizofrenia y depresión durante la serie temporal de cuatro años.

La explotación inicial de los datos se ha realizado mediante el programa estadístico *SPSS Statistics versión 17.0* desarrollado por IBM Company. Se han definido diversos filtros sobre las bases de datos para obtener el número de pacientes, con los diagnósticos seleccionados, para cada municipio o unidad espacial. Cada registro del SISMA corresponde a una actividad por lo que se descartaron los registros que describían actividades sin usuarios. Posteriormente se retiraron aquellos registros que describían distintas visitas de un mismo paciente para conseguir que cada registro correspondiese a un paciente diferente. A continuación, se filtró la variable diagnóstico principal extrayendo los códigos de la Clasificación Internacional de Enfermedades 10ª revisión – CIE-10 (WHO, 1993) correspondientes a la esquizofrenia (códigos F20 a F29) y a la depresión (códigos F32 y F33).

Una vez depurada la base de datos se realizó el recuento de pacientes por municipio y se descartaron aquellos que no tenían completa esta variable (Tabla 2). A pesar de esto, se ha podido analizar aproximadamente el 90% de los casos con diagnósticos de la serie temporal.

Tabla 2: Pacientes con esquizofrenia y depresión atendidos en servicios de salud mental ambulatorio en Andalucía

<b>ESQUIZOFRENIA (CIE-10: F20-F29)</b>			
	<b>Pacientes diagnosticados totales</b>	<b>Pacientes diagnosticados sin municipio</b>	<b>Pacientes diagnosticados con municipio</b>
2004	18.198	1.581 (8,7%)	16.617 (91,3%)
2006	20.424	1.521 (7,4%)	18.903 (92,6%)
2007	21.131	1.526 (7,2%)	19.505 (92,3%)
2008	19.725	1.431 (7,3%)	18.294 (92,7%)
<b>DEPRESIÓN (CIE-10: F32-F33)</b>			
	<b>Pacientes diagnosticados totales</b>	<b>Pacientes diagnosticados sin municipio</b>	<b>Pacientes diagnosticados con municipio</b>
2004	19.804	2.034 (10,3%)	17.770 (89,7%)
2006	21.608	2.060 (9,5%)	19.548 (90,5%)
2007	22.100	2.157 (9,8%)	19.943 (90,2%)
2008	21.292	1.825 (8,6%)	19.467 (91,4%)

Fuente: SISMA, 2004, 2006, 2007 y 2008

### **3.3.1.2. El Conjunto Mínimo Básico de Datos de Salud Mental Ambulatoria**

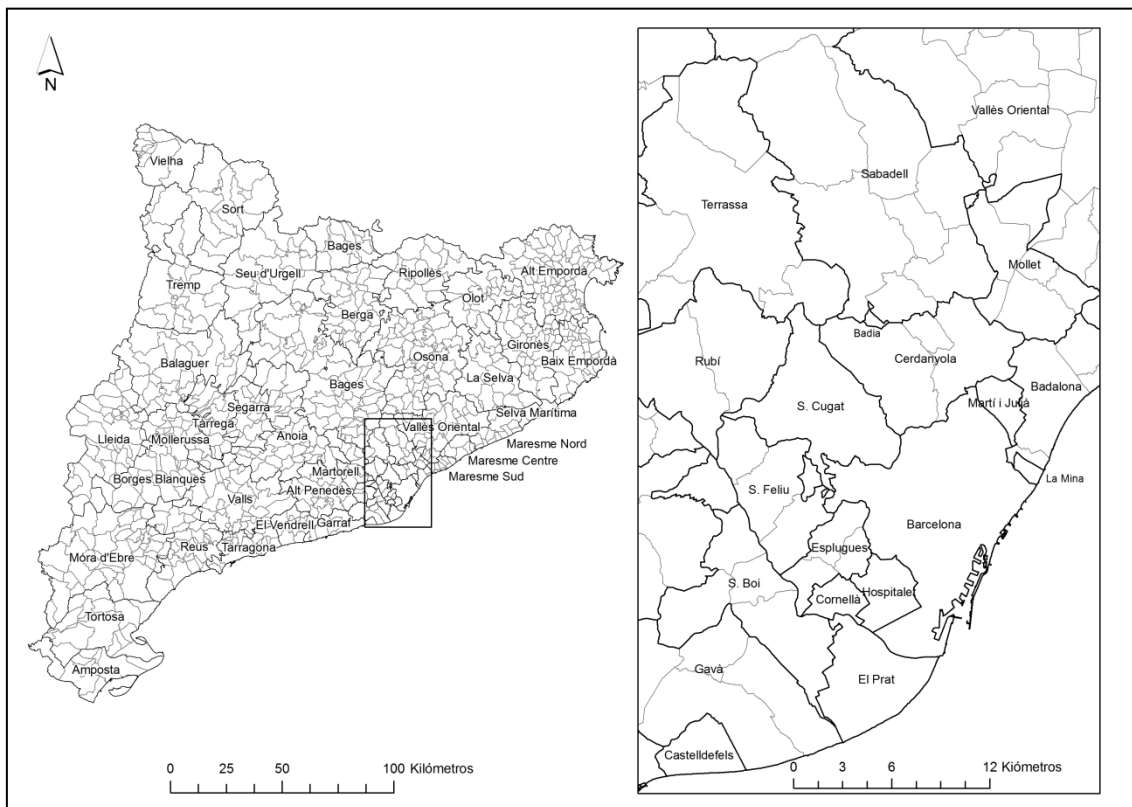
#### **3.3.1.2.1. Descripción del CMBD-SMA**

La gestión y planificación de la atención psiquiátrica en Cataluña se lleva a cabo desde la Dirección General de Regulación, Planificación y Recursos Sanitarios del Departamento de Salud. La información de casos psiquiátricos atendidos en los Centros de salud Mental de Adultos (CSMA) y Centros de Salud Mental Infanto-juveniles (CSMIJ) de Cataluña es recopilada en el Conjunto Mínimo Básico de Datos de Centros de Salud Mental Ambulatorios (CMBD-SMA) (Departament de Sanitat i Seguretat Social, 2003). El Departamento de Salud facilitó los datos del CMBD-SMA sustituyendo los nombres de los pacientes por códigos numéricos.

Los CSMA y CSMIJ son los servicios de salud mental de tipo ambulatorio donde son atendidos los casos que no pueden ser tratados en atención primaria. Poseen equipos multidisciplinares de profesionales para ofrecer una atención integral y comunitaria. Cada CSMA y CSMIJ tiene asignado un grupo de municipios a los que prestan cobertura formando pequeñas áreas de atención de salud mental. Estas áreas son inframunicipales en el caso de los siete municipios más poblados (Barcelona, Badalona, l'Hospitalet de Llobregat, Sabadell, Sant Adrià del Vallès, Tarragona y Terrassa).

En el Mapa 2 se presentan los municipios de Cataluña y su distribución por áreas de cobertura o pequeñas áreas de salud mental. Al igual que en el caso de Andalucía se ignoran las áreas inframunicipales por no ser objeto de este estudio.

Mapa 2: Áreas de cobertura de los Centros de Salud Mental de Adultos de Cataluña (946 municipios)



Fuente: Departamento de Salud de Cataluña, 2009

El CMBD-SMA ofrece datos del paciente y de las actividades desarrolladas. La información sobre el paciente recoge edad, sexo, municipio de residencia, área básica, distrito y región sanitaria, el centro donde se le atendió, diagnóstico principal y secundario o si el paciente está en un Programa de Trastorno Mental Severo. También se recoge de cada paciente las fechas y los motivos de ingreso y alta junto a la financiación de la atención. El CMBD-SMA únicamente recoge actividades realizadas con pacientes, a diferencia del SISMA andaluz, ya sean en el centro o a domicilio. Las actividades descritas son las siguientes: primera visita, visita de seguimiento, pruebas complementarias, psicoterapia individual, psicoterapia grupal, tratamiento familiar, atención de enfermería, trabajo social, visita domiciliaria y visita no programada. En la Tabla 3 se muestra información sobre actividades totales y número de pacientes registrados en el CMBD-SMA en el año 2009.

Tabla 3: Actividades y pacientes recogidos en el Conjunto Mínimo Básico de Datos de Salud Mental Ambulatoria de Cataluña

	2009
Actividades totales	1.431.898
Actividades con pacientes	1.431.898
Pacientes	214.000

Fuente: SISMA, 2004, 2006, 2007 y 2008

La Red de Salud Mental de Cataluña se compone de 74 CSMA y 53 CSMIJ. Las áreas de cobertura de los CSMA incluyen una o varias áreas de CSMIJ aunque en ocasiones el ajuste entre áreas no es exacto. En 2009 hubo un total de seis CSMA que no informaron al CMBD, concretamente: Ciutat Badia, l'Hospitalet 2, la Mina, Maragall, Sant Martí Nord y Vielha. Los CSMIJ sin datos fueron otros seis: Cerdanyola-Ripollet, Ciutat Badia, Ciutat Vella, la Mina y Sant Martí Nord y Sant Martí Sud.

### 3.3.1.2.2. Explotación del CMBD-SMA

Las variables utilizadas fueron el sexo, edad, municipio de residencia y diagnóstico principal. Esta información se ha extraído para todos los casos atendidos de esquizofrenia y depresión. En esta base de datos la codificación utiliza la Clasificación Internacional de Enfermedades en su 9 revisión – CIE-9 (Arbor, 1978). Los códigos de la esquizofrenia son el 295, 297, 298 y 301.22, y para la depresión el 296.2, 296.3 y 311.

La explotación se ha realizado de nuevo con el paquete estadístico *SPSS Statistics* filtrando los diagnósticos seleccionados por municipios. El total de pacientes considerados en los análisis han sido 25.270 con esquizofrenia y 24.580 con depresión como diagnóstico principal (Tabla 4). Los datos por pacientes se han agregado por municipio como unidad espacial de análisis. También por sexo y grupo de edad para poder calcular las tasas estandarizadas de prevalencia administrativa (por 1.000 habitantes) mediante el método directo considerando la Comunidad como población estándar. Los casos con esquizofrenia o depresión perdidos por tener vacío o erróneo el campo referido al municipio de residencia no han superado el 1% para cada diagnóstico.

Tabla 4: Pacientes con esquizofrenia y depresión atendidos en servicios de salud mental ambulatorios de Cataluña

<b>ESQUIZOFRENIA (CIE-9: 295, 297, 298 y 301.22)</b>			
	<b>Pacientes diagnosticados totales</b>	<b>Pacientes diagnosticados sin municipio de residencia</b>	<b>Pacientes diagnosticados con municipio de residencia</b>
2009	25.488	218 (0,9%)	25.270 (99,1%)
<b>DEPRESIÓN (CIE-9: 296.2, 296.3 y 311)</b>			
	<b>Pacientes diagnosticados totales</b>	<b>Pacientes diagnosticados sin municipio de residencia</b>	<b>Pacientes diagnosticados con municipio de residencia</b>
2009	24.677	97 (0,4%)	24.580 (99,6%)

Fuente: CMBD-SMA, 2009

### 3.3.2. Variables espaciales de análisis

#### 3.3.2.1. Las relaciones entre variables espaciales

La dependencia espacial no puede ser abordada con las técnicas estadísticas y econométricas clásicas utilizadas para tratar la dependencia en otros tipos de datos como las series temporales (Getis, 2007). Mientras que la dependencia temporal es unidireccional (pasado-presente) la dependencia espacial se desarrolla en múltiples direcciones (espacio geográfico). Por esta razón, se han desarrollado técnicas específicas que consideran las relaciones espaciales entre observaciones basadas en su posición geográfica ya sea por distancia (proximidad) o contigüidad.

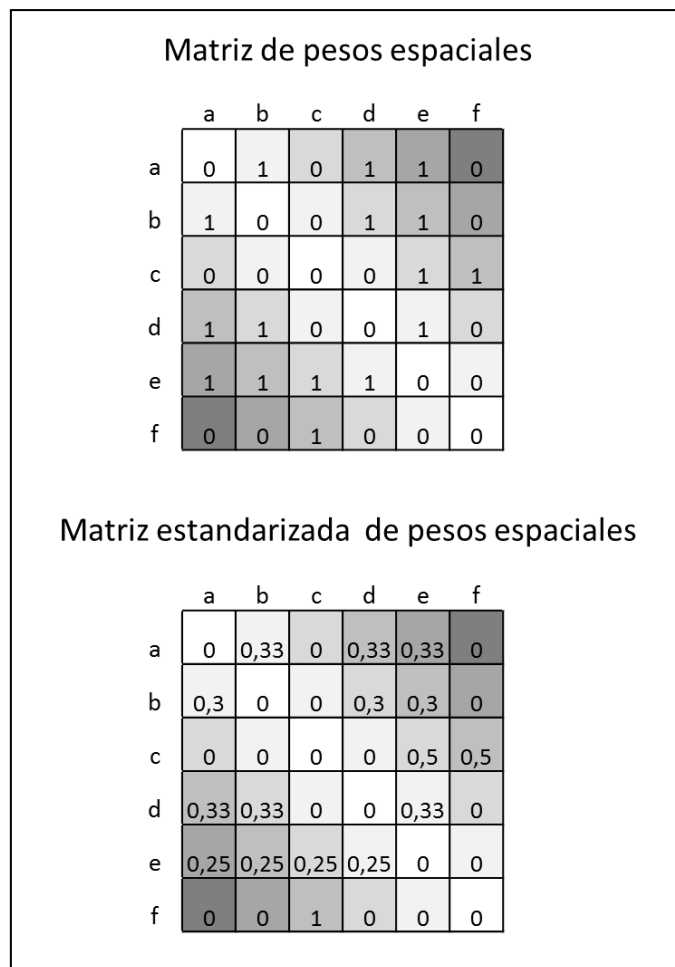
El estudio de las relaciones entre unidades espaciales es complejo ya que dependen del comportamiento espacial de la variable que se esté analizando, de las características geográficas del área de estudio, de los medios de transporte y vías de comunicación existentes, etc. Además, la heterogeneidad en tamaños, formas y poblaciones entre unidades espaciales pueden distorsionar los resultados.

Las unidades espaciales pueden ser *modelizadas* como puntos (p. e. poblaciones, actividades, etc.), polígonos (p. e. municipios, parcelas, etc.) o líneas (p. e. infraestructuras lineales, cursos fluviales, etc.). La geometría de las unidades espaciales estará influida también por el nivel de detalle territorial del estudio, por ejemplo una ciudad puede ser representada como un punto o un polígono según la escala. A pesar de esto, según las necesidades es posible transformar el

modelo de datos espacial mediante diversas técnicas (centroide del polígono, ubicación del núcleo principal, polígonos de Thiessen a partir de puntos, etc.).

La valoración de las relaciones espaciales entre las unidades espaciales se realiza mediante la matriz de pesos espaciales ( $W$ ). La matriz valora por pares la intensidad de las relaciones entre las unidades espaciales ( $W_{ij}$ ), de esta forma se recoge la multi-direccionalidad que caracteriza a las relaciones entre los datos espaciales.

Figura 7: Ejemplo de matrices de pesos espaciales



Fuente: elaboración propia

La matriz de pesos espaciales, independientemente de la forma en que refleja las relaciones espaciales, es cuadrada porque relaciona a todas las unidades entre sí; no estocástica porque las relaciones son estables y medibles; y, normalmente, simétrica porque las relaciones entre pares de unidades espaciales suelen ser recíprocas (Figura 7). Los pesos entre elementos deben ser positivos y finitos. En este sentido, las matrices de pesos espaciales habitualmente se

estandarizan mediante la división de los valores de cada fila por la suma de la misma (la suma de la fila será 1). La estandarización permite ponderar por igual la influencia de las unidades vecinas independientemente de su número y facilita la interpretación y la comparabilidad de los pesos espaciales con otros ámbitos geográficos, por contra se pierde la simetría de la matriz (Anselin, 2002).

Existen dos formas principales de definir las relaciones espaciales en la matriz de pesos<sup>3</sup>:

- a) La *distancia o proximidad* entre puntos o centroides de polígonos. Se puede medir mediante la distancia euclídea, la distancia euclídea inversa, la distancia ajustada por densidad de población, el camino más corto, una distancia fijada, etc. Normalmente se establece un radio de distancia a partir del cual no se consideran las relaciones espaciales entre observaciones evitando analizar pares de unidades lejanas.
  
- b) La *vecindad o contigüidad* se mide de forma binaria en base a si las *unidades espaciales* comparten límites (si:  $w_{ij}=1$  ó no:  $w_{ij}=0$ ), o de forma continua a través de la longitud de la frontera compartida, la longitud ponderada del perímetro compartido, la relación del perímetro compartido con las áreas, etc. También se pueden considerar distintos órdenes de contigüidad, de primer orden son las unidades que comparten límites, el segundo orden las vecinas a las contiguas y así sucesivamente. El número, la irregularidad y el tamaño de las unidades espaciales dificultan fijar un criterio de vecindad entre las mismas. Así, dos unidades pueden considerarse vecinas únicamente por compartir un punto en común o, en cambio, se puede exigir que compartan una determinada longitud de frontera.

Existen varios problemas comunes a ambas aproximaciones como el efecto frontera. Las unidades espaciales localizadas en los límites del área geográfica de estudio poseen relaciones con unidades espaciales externas de las que no se conoce el valor de la variable analizada, lo que distorsionará la naturaleza de la dependencia espacial. Otro problema es que algunos métodos exigen que las relaciones espaciales sean recíprocas cuando en la realidad puede suceder que no sean iguales en ambos sentidos (p. e. un municipio con alto desarrollado mantiene relaciones comerciales desiguales con otro municipio con menor desarrollo). Por

---

<sup>3</sup> La mayor parte de los estudios utilizan alguna de estas dos formas de valorar las relaciones espaciales. Algunos autores han introducido acercamientos alternativos a las relaciones espaciales como la valoración de la accesibilidad (Bodson y Peeters, 1975), aunque este concepto está basado en estimaciones más o menos objetivas; o la distancia económica (Case, Rosen, et al., 1993) considerando la renta per cápita, grado de intercambio comercial, etc; o aquella que considera la estructura de las redes sociales entre unidades espaciales (Doreian, 1980).



tanto, existen diversas formas de definir las relaciones geográficas entre unidades espaciales y diversas problemáticas aparejadas, lo que convierte a esta decisión en un punto crítico a la hora de realizar un análisis espacial de datos (Anselin, 2002).

### 3.3.2.2. Prevalencia administrativa

La prevalencia administrativa de una enfermedad no es equivalente a la prevalencia total ya que recoge únicamente los casos atendidos y, dentro de estos, solo aquellos que lo han sido en servicios que registran su actividad en un sistema de información sanitario. La inexistencia o incluso el mal funcionamiento de los registros sanitarios pueden suponer la subvaloración de los casos que realmente existen en un determinado territorio. La tasa de prevalencia administrativa por municipio es el resultado del cociente de los pacientes atendidos en servicios de salud mental ambulatorios de cada municipio entre la población del mismo.

$$p^A = \frac{x_j^A}{Pob^A} \times 1.000 \quad (7)$$

Donde,  $p^A$  es la prevalencia administrativa en la unidad espacial  $A$ ,  $x_j^A$  son los casos atendidos de la enfermedad  $j$  en la ubicación  $A$ , y  $Pob^A$  es la población de  $A$ .

Las cifras de población utilizadas en el denominador han sido obtenidas del Padrón Municipal de Población. El Padrón es el registro administrativo de los habitantes que residen habitualmente en cada municipio de España. Sus datos son oficiales y se revisan anualmente por los Ayuntamientos bajo la coordinación de los Institutos regionales de estadística y el Instituto Nacional de Estadística (INE) (Instituto Nacional de Estadística, 2011). El Padrón recoge diversos datos como el sexo, edad, lugar de nacimiento y nacionalidad de los vecinos de cada municipio a 1 de enero de cada año. En este estudio se han obtenido cifras de población por grupo de edad y sexo de cada uno de los 770 municipios andaluces para los años 2004, 2006, 2007 y 2008, y los 946 municipios catalanes para el 2009.

Las diferencias en la distribución geográfica de los indicadores sanitarios, en este caso la prevalencia de una enfermedad, a través de las unidades espaciales no se debe exclusivamente a las variaciones en la intensidad del fenómeno estudiado, sino que también influye otras variables, llamadas variables de confusión, como la estructura de la población ya sea por edad,

sexo, etc. (Abellana y Ascaso, 2004; Rezaeian, Dunn, et al., 2007). Por este motivo, la prevalencia ha sido corregida por edad y sexo para evitar los efectos producidos por la existencia de estructuras de población muy diferentes en los municipios andaluces. La tasa de prevalencia sobre la que se ha realizado el análisis espacial se ha estandarizado o ajustado mediante el método directo. Se ha seleccionado este método porque se conocía la distribución de los pacientes por grupo de edad y sexo. En los casos en los que no está disponible la información por estrato de edad se suele recurrir al método de estandarización indirecto (Ocaña-Riola, 2010).

El método de estandarización directa calcula la prevalencia administrativa que tendría cada unidad espacial si su estructura de población por sexo y edad fuera la misma. El procedimiento consiste en seleccionar una población estándar, en este caso la población regional andaluza o catalana, a la que se le aplica las tasas de prevalencia por grupos de sexo y edad de las diferentes unidades espaciales. De este modo se obtiene el número de casos que cada unidad tendría si su población fuera la estándar. La prevalencia ajustada final para cada unidad espacial es el resultado del cociente entre la suma de los casos esperados en cada unidad espacial y la población estándar total.

$$p_e = \frac{\sum_{x=0}^w p_x^A \times P_{ob}^{estandar}}{\sum_{x=0}^w P_{ob}^{estandar}} \quad (8)$$

Donde,  $p_e$  es la prevalencia estandarizada,  $p_x^A$  es la prevalencia en la población  $A$  para los distintos grupos de sexo y edad  $x$ ,  $P_{ob}^{estandar}$  es la población estándar para el grupo de población  $x$  y  $P_{ob}^{estandar}$  es la población estándar total.

### 3.3.2.3. Análisis espacial exploratorio de la prevalencia administrativa

Los índices de autocorrelación espacial global son métodos univariantes que miden la existencia de dependencia espacial en la distribución geográfica de una variable. Estos índices nos permiten detectar un esquema general de dependencia espacial pero no localizarla en el espacio (Moreno y Vayá, 2000). Existen diversos índices aunque los más conocidos son la  $I$  de Moran (Moran, 1948), la  $C$  de Geary (Geary, 1954) y la  $G$  de Getis y Ord (Getis y Ord, 1992).

Se ha realizado un análisis espacial exploratorio mediante un SIG con la finalidad de comprobar si existe dependencia espacial en las poblaciones de prevalencia administrativa objeto de

estudio. En caso positivo avalaría la realización del análisis a escala local para localizar concretamente en el territorio donde se encuentran las concentraciones de prevalencia administrativa. El análisis exploratorio emplea una matriz de pesos espaciales que pondera las relaciones entre cada par de unidades espaciales mediante la distancia euclídea inversa que existe entre las mismas (centroides de los polígonos municipales). Además, los pesos de la matriz han sido estandarizados por filas. Se ha usado una matriz de pesos espaciales convencional calculada por defecto por el SIG por la naturaleza exploratoria de este análisis. En cambio, como se verá más adelante, se ha diseñado expresamente una matriz para los índices espaciales que se integran en el AEMO/AE. Este análisis se ha efectuado mediante dos índices de autocorrelación espacial global que se detallan a continuación.

### 3.3.2.3.1. *I* de Moran global

El índice de la *I* de Moran (Moran, 1948) permite conocer si existe agrupación espacial de valores similares o disímiles, o si la distribución es aleatoria.

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n W_{i,j}} \frac{\sum_{i=1}^n \sum_{j=1}^n W_{i,j} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (9)$$

Donde,  $x_i$  y  $x_j$  son los valores de la variable  $x$  en las localizaciones  $i$  y  $j$ ;  $\bar{x}$  es la media de los valores de la variable  $x$ ; y  $W_{ij}$  representa los pesos espaciales entre pares de localizaciones ( $i$  y  $j$ ).

Un valor positivo y significativo de *I* indica dependencia espacial positiva (valores similares), un valor negativo y significativo señala dependencia espacial negativa (valores diferentes) y un valor en torno a 0 o no significativo expresa que la distribución es aleatoria.

La interpretación de la *I* de Moran precisa normalizar sus valores por la desviación estándar calculando sus *valores z* ( $z_I$ ). Los valores  $z_I$  permitirán calcular la probabilidad (*valor p*), en base a la curva de Gauss o curva normal, de que se cumpla o no la hipótesis nula ( $H_0$ ) en la que no existen patrones espaciales y se está ante una distribución espacial aleatoria de los datos. Los valores *I* que tengan una probabilidad por debajo de 0,05 ( $z_I \geq 1,960$  y  $z_I \leq -1,960$ ) serán significativos porque solo existe un 5% de posibilidades de que se cumpla la  $H_0$  de que la distribución se deba al azar<sup>4</sup>.

---

<sup>4</sup> De forma más exigente un valor *p* de 0,01 se dará cuando  $z_I \geq 2,580$  y  $z_I \leq -2,580$

$$z_I = \frac{I - E[I]}{\sqrt{V[I]}} \quad (9.1)$$

$$E[I] = -1/(n - 1) \quad (9.1.2)$$

$$V[I] = E[I^2] - E[I]^2 \quad (9.1.3)$$

$$E[I^2] = \frac{A - B}{C} \quad (9.1.4)$$

$$A = n[(n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2] \quad (9.1.5)$$

$$B = D[(n^2 - n)S_1 - 2nS_2 + 6S_0^2] \quad (9.1.6)$$

$$C = (n - 1)(n - 2)(n - 3)S_0^2 \quad (9.1.7)$$

$$D = \frac{\sum_{i=1}^n z_i^4}{(\sum_{i=1}^n z_i^2)^2} \quad (9.1.8)$$

$$S_1 = \left(\frac{1}{2}\right) \sum_{i=1}^n \sum_{j=1}^n (W_{i,j} + W_{j,i})^2 \quad (9.1.9)$$

$$S_2 = \sum_{i=1}^n (\sum_{j=1}^n W_{i,j} + \sum_{j=1}^n W_{j,i})^2 \quad (9.1.10)$$

### 3.3.2.3.2. $G$ de Getis y Ord global

El índice  $G$  de Getis y Ord (Getis y Ord, 1992) informa si existe agrupación de valores similares y si ésta es de valores altos o bajos. También señala si la distribución es aleatoria. Su expresión es la siguiente:

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} x_i x_j}{\sum_{i=1}^n \sum_{j=1}^n x_i x_j}, \forall j \neq i \quad (10)$$

Donde,  $x_i$  y  $x_j$  son los valores de la variable  $x$  en la localizaciones  $i$  y  $j$ ; y  $W_{ij}$  representa los pesos espaciales entre pares de unidades espaciales  $i$  y  $j$ .

Un valor positivo y significativo de  $G$  señala agrupación geográfica de datos con valores elevados, el valor significativo negativo indica agrupación de valores bajos. Los valores no significativos señalan que la distribución es aleatoria. El valor  $z$  ayudará a la interpretación del índice y a conocer su significación.

$$z_G = \frac{G - E[G]}{\sqrt{V[G]}} \quad (10.1)$$

$$E[G] = \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij}}{n(n-1)}, \forall j \neq i \quad (10.1.1)$$

$$V[G] = E[G^2] - E[G]^2 \quad (10.1.2)$$

$$E[G^2] = \frac{A+B}{c} \quad (10.1.3)$$

$$A = D_0 \left( \sum_{i=1}^n x_i^2 \right)^2 + D_1 \sum_{i=1}^n x_i^4 + \left( \sum_{i=1}^n x_i \right)^2 \sum_{i=1}^n x_i^2 \quad (10.1.4)$$

$$B = D_3 \sum_{i=1}^n x_i \sum_{i=1}^n x_i^3 + D_4 \left( \sum_{i=1}^n x_i \right)^4 \quad (10.1.5)$$

$$C = [(\sum_{i=1}^n x_i)^2 - \sum_{i=1}^n x_i^2]^2 \times n(n-1)(n-2)(n-3) \quad (10.1.6)$$

$$D_0 = (n^2 - 3n + 3)S_1 - nS_2 + 3W^2 \quad (10.1.7)$$

$$D_1 = -[(n^2 - n)S_1 - 2nS_2 + 6W^2] \quad (10.1.8)$$

$$D_2 = -[2nS_1 - (n + 3)S_2 + 6W^2] \quad (10.1.9)$$

$$D_3 = 4(n - 1)S_1 - 2(n + 1)S_2 + 8W^2 \quad (10.1.10)$$

$$D_4 = S_1 - S_2 + W^2 \quad (10.1.11)$$

$$W = (\sum_{i=1}^n \sum_{j=1, i \neq j}^n W_{i,j}) \quad (10.1.12)$$

$$S_1 = \left(\frac{1}{2}\right) \sum_{i=1}^n \sum_{j=1, i \neq j}^n (W_{i,j} + W_{j,i})^2 \quad (10.1.13)$$

$$S_2 = \sum_{i=1}^n (\sum_{j=1, i \neq j}^n W_{i,j} + \sum_{j=1}^n W_{i,j})^2 \quad (10.1.15)$$

### 3.3.2.4. Índices locales para el análisis espacial de datos

#### 3.3.2.4.1. Definición de las relaciones espaciales empleadas en el estudio

Las técnicas locales de análisis espacial de datos empleadas en el Estudio de Demostración A precisaron del cálculo de la correspondiente matriz de pesos espaciales ( $W$ ) entre pares de municipios andaluces (unidades espaciales). La elección de la forma de definir las relaciones espaciales ha buscado evitar los problemas que provocan tanto las diferencias de tamaño entre

unidades espaciales como las irregularidades en sus formas (Anselin, 2002). De hecho, en Andalucía existen casos en los que el centroide de un municipio de gran extensión está a mayor distancia del centroide de una unidad espacial contigua que de otra que no es contigua.

Las relaciones espaciales se han definido considerando tanto la distancia entre centroides de las unidades espaciales como la vecindad basada en las fronteras. De este modo, en el cálculo de los Índices Locales de Asociación Espacial (Moran, Geary y Getis & Ord) y del Riesgo Bayesiano (CAR), se ha considerado que una unidad espacial será vecina de otra si comparte frontera, pero también lo será de aquellas que se encuentran dentro de un radio de distancia estadísticamente definido en base a la distribución geográfica de las unidades espaciales.

Las unidades espaciales ( $UE$ ) que comparten frontera son siempre identificadas como vecinas  $UE_j \in UEV_j$ , siendo  $UE_j$  la unidad espacial que tomamos de referencia y  $UEV_j$  el conjunto de unidades vecinas a  $UE_j$ . Cuando las unidades no son contiguas se identifican como tales secuencialmente siguiendo el siguiente procedimiento:

1. Se identifica la unidad espacial ( $UE_i$ ) más próxima a la unidad espacial de referencia ( $UE_j$ ) que no sea contigua ( $UE_i \notin UEV_j$ ).
2. Se calcula el Error Cuadrático Medio (ECM)<sup>5</sup> de las distancias de la unidad espacial  $UE_j$  a todas las unidades definidas como vecinas ( $UEV_j$ ) incluida la unidad no contigua ( $UE_i$ ).
3. Si el ECM es inferior a un valor predefinido, seleccionado en función del estudio realizado mediante un análisis de sensibilidad a este parámetro, se incluye en el conjunto de unidades vecinas ( $UE_i \in UEV_j$ ) y se vuelve al paso 1) seleccionando una nueva unidad espacial próxima; en cambio, si el error es superior se descarta como unidad espacial vecina ( $UE_i \notin UEV_j$ ).

La matriz de pesos espaciales ( $W$ ) empleada en el cálculo de los índices de dependencia espacial recoge las relaciones entre pares de unidades de forma binaria, con valor 1 cuando éstas son vecinas y con valor 0 cuando no son vecinas.

---

<sup>5</sup> El error cuadrático medio (ECM) es la esperanza ( $E$ ) de la diferencia al cuadrado entre el valor medio ( $\bar{X}$ ) y el estimado ( $\mu$ ), siendo  $N$  el tamaño de la muestra.

$$ECM = E(\bar{X} - \mu)^2 = \left(\frac{\sigma}{\sqrt{N}}\right)^2 = \frac{\sigma^2}{N} \quad (11)$$

### 3.3.2.4.2. Indicadores Locales de Asociación Espacial (LISA)

Los Indicadores Locales de Asociación Espacial (LISA) son índices de autocorrelación espacial univariante que miden el grado de asociación espacial de cada observación de una variable en el espacio con sus vecinas. Los LISA expresan la contribución de cada observación espacial al indicador global de autocorrelación espacial. De esta manera, la suma de los valores de los indicadores locales de cada observación será proporcional al indicador global (Anselin, 1995).

Los indicadores locales se desarrollaron a partir de los índices de autocorrelación espacial globales que miden la dependencia espacial del conjunto de las observaciones de la variable. Frente a los índices de autocorrelación global, los LISA permiten detectar la posible existencia de agrupaciones significativas de valores similares en ciertas localizaciones del espacio (dependencia espacial), la existencia de áreas con procesos espaciales no estacionarios (heterogeneidad espacial) o valores fuera de rango (*outliers*). Los resultados de los indicadores globales y locales pueden ser coincidentes, pero es posible que una variable espacial con una distribución espacial aleatoria según los indicadores globales presente áreas concretas en las que exista dependencia espacial.

Los LISA que se presentan a continuación han sido calculados mediante un programa informático diseñado a medida sobre el lenguaje de programación XBase.

#### A) Índice de la *I* de Moran local

El test local de la *I* de Moran es resultado de la adaptación realizada por Luc Anselin (Anselin, 1995) del método de autocorrelación espacial global creado por P.A.P. Moran (Moran, 1948). Su expresión es la siguiente:

$$I_i = \frac{(x_i - \bar{x})}{\sigma^2} \sum_{j=1; j \neq i}^n W_{ij} (x_j - \bar{x}) \quad (12)$$

Donde,  $x_i$  es el valor de la variable  $x$  en la localización  $i$ ;  $\bar{x}$  es la media de los valores de la variable  $x$ ;  $x_j$  es el valor de la variable  $x$  en cualquier localización excepto  $i$  ( $j \neq i$ );  $W_{ij}$  representa a la matriz de pesos espaciales;  $\sigma^2$  es la varianza de la variable  $x$ :



$$\sigma^2 = \sum_{j=1}^n (x_j - \bar{x})^2 / n \quad (12.1)$$

Un valor positivo y significativo de  $I_i$  indica la existencia de dependencia espacial positiva ( $I_i > 0$ ). En cambio, un valor negativo y significativo señala la presencia de dependencia espacial negativa ( $I_i < 0$ ). Un valor similar a 0 o no significativo indicaría una distribución espacial aleatoria ( $I_i \approx 0$ ). La  $I$  de Moran no expresa por sí mismo el tipo de dependencia espacial positiva de las unidades espaciales, es decir, no señala si se está ante asociaciones de valores altos o bajos de la variable de estudio.

El valor  $z$  ( $zI_i$ ) indica si la significación de la prueba es suficiente para descartar la hipótesis nula de aleatoriedad. Para rechazar la  $H_0$  de distribución aleatoria, los valores  $z$  deben ser  $\geq 1,960$  ó  $\leq -1,960$ , siendo el nivel significación del 0,05. El valor  $z$  se extrae de la siguiente operación:

$$zI_i = \frac{I_i - E(I_i)}{\sqrt{V(I_i)}} \quad (12.2)$$

Donde  $I_i$  es el valor de la  $I$  de Moran en la unidad espacial  $x_i$ ;  $E(I_i)$  es el valor esperado de  $I_i$ ; y  $V(I_i)$  es la varianza de  $I_i$ . Así, asumiendo que todas las permutaciones de valores en cualquier localización son igualmente probables:

$$E(I_i) = -\frac{W_i}{n-1} \quad (12.2.1)$$

$$V(I_i) = E[I_i^2] - E[I_i]^2 \quad (12.2.2)$$

$$E[I_i^2] = A - B \quad (12.2.3)$$

$$A = \frac{(n - b_{2i}) \sum_{j=1; j \neq i}^n W_{ij}^2}{n-1} \quad (12.2.4)$$

$$B = \frac{(2b_{2i} - n) \sum_{k=1, k \neq i}^n \sum_{h=1, h \neq i}^n (W_{ik} W_{ih})}{(n-1)(n-2)} \quad (12.2.5)$$

$$b_{2i} = \frac{\sum_{i=1, i \neq j}^n (x_i - \bar{x})^4}{(\sum_{i=1, i \neq j}^n (x_i - \bar{x})^2)^2} \quad (12.2.6)$$

## B) Índice de la C de Geary local

Anselin también transformó el índice de autocorrelación espacial global de R. C. Geary (Geary, 1954) en índice local. La expresión del índice local de la C de Geary (Anselin, 1995) es:

$$C_i = \frac{\sum_{j=1}^n W_{ij} (x_i - x_j)^2}{\sigma^2} \quad (13)$$

Donde,  $x_i$  es el valor de la variable  $x$  en la unidad espacial  $i$ ;  $x_j$  es el valor de la variable  $x$  en cualquier localización excepto  $i$  ( $j \neq i$ );  $W_{ij}$  representa a la matriz de pesos espaciales;  $\sigma^2$  es la varianza de la variable  $x$  (ver expresión 1.1).

Las unidades espaciales que presenten valores significativamente altos de la C de Geary estarán asociadas a las unidades vecinas con valores disímiles o dependencia espacial negativa. En cambio, cuando los valores sean significativamente bajos existirá asociación de valores similares o dependencia espacial positiva. Una distribución aleatoria presentará valores similares a la media o valores no significativos. Este índice al igual que el de Moran no indica si la agrupación es de valores altos o bajos.

La significación de la C de Geary se extrae de la normalización de sus valores obteniendo  $zC_i$ .

$$zC_i = \frac{C_i - E(C_i)}{\sqrt{V(C_i)}} \quad (13.1)$$

Asumiendo que todas las permutaciones de datos son posibles, la esperanza y la varianza de  $C_i$  serán:

$$E(C_i) = -\frac{2nW_i}{n-1} \quad (13.1.1)$$

$$V(C_i) = \left(\frac{n}{n-1}\right)(W_i^2 - W_{i(2)})(3 - b_2) - \left(\frac{2nW_i}{n-1}\right)^2 \quad (13.1.2)$$

$$b_{2i} = \frac{\sum_{i=1; i \neq j}^n (x_i - \bar{x})^4}{\left(\sum_{i=1; i \neq j}^n (x_i - \bar{x})^2\right)^2} \quad (13.1.3)$$

### C) Índice de la $G^*$ de Getis y Ord local

A. Getis y J. K. Ord también adaptaron su índice global  $G$  (Getis y Ord, 1992) para conocer la aportación de cada observación a la dependencia espacial de toda la muestra. La  $G$  de Getis y Ord (Ord y Getis, 1995) no cumple la característica de los LISA de que la suma de sus valores es proporcional al valor del índice global. Por esta razón no puede considerarse estrictamente como un LISA aunque se suele incluir entre ellos (Moreno y Vayá, 2000). Existen dos versiones del índice local  $G$  según si considera o no la observación  $x_i$  que se está calculando ( $G: j \neq i$  y  $G^*: j=i$ ). En este estudio se ha utilizado la  $G^*$  que posee la siguiente expresión:

$$G_i^* = \frac{\sum_{j=1}^n W_{ij}x_j}{\sum_{j=1}^n x_j} \quad (14)$$

Donde  $x_j$  es el valor de la variable  $x$  en cualquier localización incluida la unidad espacial  $i$ ;  $W_{ij}$  representa a la matriz de pesos espaciales.

Si el valor de  $G_i^*$  es positivo y estadísticamente significativo indica que existen agrupaciones de unidades espaciales de valores altos en torno a la unidad  $i$ . En cambio, si  $G_i^*$  es bajo y significativo los valores de las unidades en torno a  $i$  serán similarmente bajos. En ambos casos nos señala la existencia de dependencia espacial positiva con agrupaciones de unidades espaciales de valores similares mientras que no informa de la existencia de dependencia espacial negativa.

De nuevo, asumiendo el total de permutaciones, la significación vendrá dada por  $zG_i^*$ :

$$zG_i^* = \frac{G_i^* - E(G_i^*)}{\sqrt{V(G_i^*)}} \quad (14.1)$$

$$E(G_i^*) = \frac{W_i}{n-1} \quad (14.1.1)$$

$$V(G_i^*) = \frac{W_i(n-W_i)\sigma^2}{n^2(n-1)\bar{x}^2} \quad (14.1.2)$$

Donde  $\sigma^2$  es la varianza y  $\bar{x}$  la media de los valores observados. Las unidades espaciales con valores de la  $G_i^*$  significativos (0,05) serán aquellas con un  $z \geq 1,960$  o  $z \leq -1,960$ .

### 3.3.2.4.3. Modelo Autorregresivo Condicional Bayesiano (Modelo CAR)

La inferencia Bayesiana ha sido aplicada sobre datos espaciales con éxito y es una de las metodologías de AED más utilizadas en la actualidad. Los modelos Bayesianos tienen en cuenta las fluctuaciones aleatorias que se dan sobre los fenómenos espaciales, en este caso la prevalencia, debido al tamaño y las características de la población en riesgo. También hay que considerar los efectos aleatorios específicos que se dan en una distribución espacial. Por una parte, los efectos espaciales que varían de forma estructurada en el espacio, como es la dependencia espacial y/o la heterogeneidad espacial producida por variables externas no analizadas. Y por otra, los efectos que varían de forma no estructurada en el espacio debidos a la existencia de variables externas no contempladas que provocan heterogeneidad no espacial.

Estos problemas se abordan mediante el *suavizado* (*smoothing*) de las tasas de mortalidad o prevalencia estandarizadas, calculando el riesgo relativo de ocurrencia del fenómeno en relación al grupo de población de referencia, con el fin de reducir su extra-variación. El *suavizado* parte del supuesto de que los casos observados de una enfermedad se ajustan a una distribución de acontecimientos aleatorios de Poisson.

$$O_i \approx \text{Poisson}(\mu_i E_i) \quad (15)$$

Donde  $O_i$  es el número de casos observados de enfermedad en la unidad espacial  $i$ ,  $\mu_i$  es el riesgo relativo del área  $i$  y  $E_i$  son los casos esperados en  $i$ . El riesgo relativo de la enfermedad también se ajustaría a una distribución aleatoria convencional a priori, en este caso una distribución de probabilidad *gamma*.

En este trabajo de investigación se ha empleado el modelo Bayesiano propuesto por Besag, York y Mollie (Besag, York, et al., 1991) denominado modelo autorregresivo condicional (CAR) o modelo BYM. Este modelo es considerado el más robusto de aquellos con estructura espacial para analizar la distribución de enfermedades en áreas pequeñas (Lawson, Biggeri, et al., 2000; Best, Richardson, et al., 2005). El modelo CAR recoge la estructura espacial de la variable de estudio a través de dos términos espaciales aleatorios que recogen los dos problemas anteriormente señalados, el primero no estructurado para considerar la heterogeneidad entre unidades espaciales y el segundo estructurado para recoger las agrupaciones de casos en el espacio (Ocaña-Riola y Mayoral-Cortés, 2010). La variabilidad estructural espacial se especifica mediante una distribución autorregresiva condicional (CAR).

$$\log(\mu_i) = \delta_i : \delta_i = \beta_0 + \beta_1 x_{(1)} + \beta_2 x_{(2)} + \dots + \beta_k x_{(k)} + v_i + S_i \quad (16)$$

$$[S_i | S_j, i \neq j] \approx N(\bar{S}_i, \sigma_i^2) \quad (16.1)$$

$$S_i = \frac{1}{m_i} \sum_{k \in \delta_i} u_k \quad (16.2)$$

$$\sigma_i^2 = \frac{\sigma_u^2}{m_i} \quad (16.3)$$

$$v_i | v_{-i} \approx N \left[ \frac{\sum_{j \neq i}^n W_{ij} u_j}{\sum_{j \neq i}^n W_{ij}}, \frac{\sigma_u^2}{\sum_{j \neq i}^n W_{ij}} \right] \quad (16.4)$$

Donde  $\mu_i$  es el riesgo relativo en la unidad espacial  $i$ ,  $\alpha$  es el riesgo relativo total de la región de estudio,  $E_i$  son los casos esperados en  $i$ ,  $v_i$  es el efecto aleatorio no espacial que se asume normalmente distribuido con media 0 y varianza constante ( $\sigma^2$ ),  $S_i$  es el efecto aleatorio espacial que según el modelo CAR es el promedio de la dependencia espacial de las áreas vecinas a  $i$ ,  $n_i$  es la población en riesgo en la unidad  $i$ ,  $\beta$  son los coeficientes de regresión,  $x$  es la variable de estudio,  $\delta_i$  son las unidades espaciales vecinas de  $i$  y  $m_i$  es el número de unidades espaciales vecinas de  $i$ . La relación espacial entre unidades geográficas se recoge de nuevo mediante la matriz de pesos espaciales ( $W_{ij}$ ) descrita anteriormente.

La determinación del riesgo relativo Bayesiano se ha realizado mediante simulación a través de las cadenas de Markov Monte-Carlo (MCMC) (Gilks, Richardson, et al., 1996) que han fijado los límites superior e inferior del intervalo de confianza (0,05). Se descartaron las primeras 40.000 iteraciones para anular el efecto de la carga de arranque (*burn-in*) y la cadena se ejecutó por un periodo de 4.000 iteraciones más. El carácter iterativo del MCMC demanda comprobar la convergencia de las simulaciones para validar los resultados del modelo, para ello se realizó un diagnóstico Gelman y Rubin (Brooks y Gelman, 1998).

Los valores que adopta el riesgo relativo estimado por el modelo Bayesiano indican el valor de la variable en cada unidad espacial considerando también los valores de sus unidades vecinas, por lo que sus valores tenderán a estar positivamente autocorrelados. Las unidades espaciales con valores significativamente elevados del CAR Bayesiano indican la existencia de valores elevados de la variable analizada en relación a los valores de la misma en las unidades vecinas permitiendo identificar clústeres espaciales de valores altos. En cambio, los valores significativamente bajos indican que el valor pequeño de la unidad espacial es similar a los de su entorno y el clúster será de valores bajos.

La implementación del enfoque CAR Bayesiano se ha llevado a cabo mediante la extensión GeoBUGS versión 1.2 del programa de libre difusión WinBUGS versión 1.4.1 (MRC Biostatistics Unit, 2008). Estas herramientas han sido desarrolladas por el Departamento de Epidemiología y Salud Pública del Imperial College del Hospital de St. Mary de Londres.

### **3.4. Sistemas de Información Geográfica (SIG)**

El *National Center for Geographic Information and Analysis* (NCGIA) (Goodchild y Kemp, 1990) define a los Sistemas de Información Geográfica como:

*“Un Sistema de Información Geográfica (SIG) es un sistema de hardware, software y procedimientos diseñado para apoyar la captura, gestión, manipulación, análisis, modelización y visualización de datos referenciados geográficamente para la resolución de problemas complejos de planificación y gestión”.*

Por tanto, los Sistemas de Información Geográfica (SIG) son herramientas informáticas que permiten recopilar, almacenar, gestionar, consultar, analizar y cartografiar datos *georreferenciados* (Ostfeld, Glass, et al., 2005). Los SIG suponen una herramienta de apoyo a la toma de decisiones territoriales tanto públicas como privadas con respecto a la ordenación del territorio y la planificación urbana, la planificación sectorial con incidencia territorial, la prevención y la protección ambiental, la planificación de redes de transporte, abastecimiento y saneamiento, la localización de servicios públicos, industriales y comerciales, la protección frente a riesgos naturales y antrópicos, etc.

Los SIG no son un mero *software* de representación cartográfica sino que incluyen potentes herramientas para el análisis espacial de información modelizadas como puntos, redes y áreas (puntos, líneas y polígonos) o como teselas. Las técnicas de análisis espacial que puede incluir un SIG son diversas como la estadística espacial, el análisis de redes y accesibilidad, el análisis de superficies, las áreas de influencia, etc. (Smith, Goodchild, et al., 2007).

En este estudio se ha empleado ArcGIS v.9 creado por la empresa ESRI (*Environmental Systems Research Institute*) (ESRI, 2011). Se trata de uno de los SIG más difundidos en el mundo y para el que se ha desarrollado una extensión específica de análisis espacial de datos.

Los resultados del estudio se han representado mediante el SIG sobre la cartografía digital municipal de Andalucía y Cataluña. En el primer caso ha sido elaborada por el Instituto de Cartografía de Andalucía (ICA) (Consejería de Obras Públicas y Vivienda de Andalucía, 2012). En el segundo caso, procede del Instituto Cartográfico de Cataluña (Departament de Territori i Sostenibilitat, 2012).

La información espacial se ha representado mediante un modelo de datos vectorial. Este modelo representa los objetos del mundo real de forma discreta con límites definidos mediante elementos geométricos (puntos, líneas o polígonos). Se ha empleado un modelo de puntos para representar los núcleos de población principal de cada municipio. A partir de éste se han establecido las distancias euclídeas entre unidades espaciales (municipios) que han servido para definir las relaciones espaciales de la matriz de pesos utilizada para calcular los índices de autocorrelación espacial. La representación cartográfica de los resultados se ha realizado

mediante polígonos equivalentes a las unidades espaciales. La abstracción de los municipios en polígonos también ha permitido establecer las relaciones de contigüidad entre ellos para establecer el primer nivel de proximidad de las distancias concéntricas mencionadas.

Los resultados que se presentan en este trabajo se presentan mediante mapas de coropletas, es decir, mapas que representan la información mediante tramas de colores (blanco-gris-negro). Los resultados numéricos se han agrupado en intervalos basados en la desviación típica o estándar señalando su dispersión estadística.

### **3.5. Limitaciones de los Estudios de Demostración**

- Las bases de datos no cuentan con información sobre uno o varios centros en la serie temporal. También se han detectado pacientes a los que no se le ha asignado el diagnóstico o el municipio de residencia.
- El estudio no analiza la prevalencia real de la esquizofrenia y la depresión sino aquella que es atendida en centros públicos y queda por tanto registrada en los sistemas de información sanitaria de la Administración.
- Los municipios limítrofes con otras Comunidades Autónomas tienen unidades espaciales vecinas con valores de prevalencia desconocidos que afectarían a sus relaciones espaciales y por tanto a su dependencia espacial (efecto frontera).
- La *falacia ecológica*, común a los estudios ecológicos, señala que no es posible inferir las relaciones entre factores de riesgo y una enfermedad a nivel individual a partir de las relaciones observadas a nivel agregado.
- También es común a este tipo de estudios el problema la *unidad espacial modificable*, que se refiere a la variación que se produce en los resultados y en su interpretación en función al tamaño del área geográfica de agregación de datos individuales que se esté analizando.

### **3.6. Valoración del Modelo AEMO/AE**

Dada la difusión del proyecto en proyectos de investigación y publicaciones científicas se ha realizado una encuesta entre expertos en investigación, planificación y gestión de servicios de salud mental para conocer su opinión sobre el modelo AEMO/AE (Anexo 1). Las preguntas



cuestionaban por la utilidad y relevancia del modelo AEMO/AE para la toma de decisiones en política y planificación sanitaria. La valoración de la utilidad y la relevancia se realizó mediante dos escalas. Una escala numérica de 1 (“Muy baja”) a 10 (“Alta”) y otra ordinal con cuatro categorías “Muy baja”, “Baja”, “Media” y “Alta”.

La encuesta se remitió a 25 expertos procedentes de las Comunidades Autónomas de Andalucía, Cantabria, Cataluña y Madrid, y de las Diputaciones forales de Guipúzcoa y Vizcaya. En la mayor parte de estos ámbitos territoriales se han realizado análisis de la prevalencia mediante el modelo o se van a realizar en el futuro aunque todos los expertos conocen el modelo. Han participado un total de 17 expertos de los contactados (68%), nueve de ellos son expertos en investigación de servicios de salud, cuatro son expertos en política y planificación de servicios de salud y otros cuatro son expertos en ambos campos. Por territorio de origen, Cantabria, Madrid, Guipúzcoa y Vizcaya están representadas por un experto cada una, otros siete proceden de Cataluña y los seis restantes de Andalucía.

## IV. RESULTADOS

### 4.1. El Algoritmo Evolutivo Multi-Objetivo de Análisis Espacial (AEMO/AE)

#### 4.1.1. Fases y procedimientos del AEMO/AE

El diseño del procedimiento del modelo AEMO/AE se muestran en la Tabla 5. Los Pre-procesos del 1 al 6, previos a al funcionamiento del algoritmo evolutivo, han sido explicados en el apartado de metodología. En los siguientes epígrafes se explicarán los Procesos que incluyen el planteamiento del problema multi-objetivo, el diseño del algoritmo y la selección de resultados finales. Existe una tercera fase denominada post-proceso destinada al análisis de resultados que se realiza en los capítulos de Resultados y Comentarios.

Tabla 5: Estructura del AEMO/AE. Descripción de fases y procedimientos

	Fases	Procedimientos
Pre-proceso 1	Análisis estadístico de la prevalencia Análisis espacial global	Análisis estadístico básico Test de Kolmogorov-Smirnov <i>I</i> de Moran y <i>G</i> de Getis y Ord global
Pre-proceso 2	Representación cartográfica de la prevalencia	Sistemas de Información Geográfica
Pre-proceso 3	Análisis de vecindad de las <i>unidades espaciales</i>	Procedimiento estadístico basado en el error cuadrático medio de la distancia euclídea
Pre-proceso 4	Determinación de la distancia relativa entre <i>unidades espaciales</i>	Procedimiento estadístico diseñado a tal fin
Pre-proceso 5	Análisis espacial de la prevalencia basado en la vecindad del paso 3 (sólo para el Estudio de Demostración A)	<i>I</i> de Moran, <i>C</i> de Geary, <i>G</i> de Getis y Ord y CAR Bayesiano
Pre-proceso 6	Estandarización de los índices del paso 5 a [0,1-0,9]	Interpolación estadística estándar
	Análisis espacial global Representación cartográfica (sólo para el Estudio de Demostración A)	<i>I</i> de Moran y <i>G</i> de Getis y Ord global Sistemas de Información Geográfica
Proceso 1	A. La media y desviación estándar de los índices del paso 5 y la distancia relativa del paso 4 son optimizados mediante un Algoritmo Evolutivo tipo SPEA2 B. La media de la prevalencia y la distancia relativa del paso 4 son optimizados mediante un Algoritmo Evolutivo tipo SPEA2	AEMO/AE (Los resultados son grupos de unidades espaciales con un número predefinido de elementos)
Proceso 2	Determinación de la probabilidad de que cada <i>unidad espacial</i> de las soluciones sea parte de un <i>hot-spot</i> o <i>cold-spot</i>	Análisis de frecuencias
Proceso 3	Selección de unidades espaciales que forman parte de los <i>hot-spot</i> o <i>cold-spot</i>	Estadística de los extremos QQ-Plots
Proceso 4	Representación cartográfica de los <i>hot-spot</i> o <i>cold-spot</i>	Sistemas de Información Geográfica

	Fases	Procedimientos
Post-proceso	Análisis de los resultados Definición de agrupaciones espaciales de resultados	Análisis estadístico básico y estadísticas no paramétricas de comparación de distribuciones y medianas (Test <i>U</i> y test de las medianas)

Fuente: adaptado de García-Alonso y colaboradores, 2010

#### 4.1.2. El Algoritmo Evolutivo de Fuerza de Pareto v.2

El primer objetivo de este trabajo es el diseño de un AEMO destinado específicamente al análisis espacial de datos (AEMO/AE) y, concretamente, a la identificación y localización secuencial de *hot-spot* y/o *cold-spots* en el territorio. El AEMO/AE se basa en el Algoritmo Evolutivo de Fuerza de Pareto v.2 (*Strength Pareto Evolutionary Algorithm - SPEA2*) (Zitzler y Thiele, 1999; Zitzler, Laumanns, et al., 2001). El SPEA2 es un algoritmo de optimización multi-objetivo “a posteriori” basado en la jerarquización de Pareto propuesta por D. Goldberg (Goldberg, 1989). En este algoritmo la dominancia de Pareto es un criterio para el operador genético de selección. Los individuos son jerarquizados por su dominancia separando los no dominados del resto de la *población*.

Junto al SPEA2 existen diversos algoritmos basados en Pareto como el Algoritmo Genético Multi-objetivo (*Multi-Objective Genetic Algorithm - MOGA*) (Fonseca y Fleming, 1993), el Algoritmo Genético de Clasificación No Dominada (*Non-Dominated Sorting Genetic Algorithm - NSGA*) (Srinivas y Deb, 2011), el Algoritmo Genético de Clasificación No Dominada II (*Non-Dominated Sorting Genetic Algorithm II - NSGA-II*) (Deb, Pratap, et al., 2002), la Estrategia de Evolución Archivada de Pareto (*Pareto Archived Evolution Strategy - PAES*) (Knowles y Corne, 1999) o el Algoritmo Evolutivo de Fuerza de Pareto (*Strength Pareto Evolutionary Algorithm - SPEA*) (Zitzler y Thiele, 1998).

El SPEA2 es un algoritmo muy flexible que permite la utilización de diferentes *funciones de aptitud*. El algoritmo emplea un archivo externo para almacenar las soluciones no dominadas (Pareto óptimas) resultantes de cada generación del proceso iterativo de optimización denominado *ENDSF* (*External non-dominated solution file*). El conjunto de soluciones no dominadas se denomina *población externa*. Así se aplica una forma de elitismo al separarse éstas del resto de soluciones candidatas en el proceso evolutivo, aunque son incorporadas al comienzo de cada iteración para sustituir a las seleccionadas para ser eliminadas. Estas soluciones participan en el proceso de selección y son sustituidas cuando se encuentra una solución mejor. El procedimiento que sigue el AEMO/AE se expone en la Tabla 6.

Tabla 6: Procedimiento del AEMO/AE

1	Creación del archivo externo <i>ENDSF</i>	
2	<b>For</b> $i = 1$ to $S^2$ <b>do</b>	
3	Diseño de la población inicial $S$	<i>Unidades espaciales</i> aleatorias
4	<b>For</b> $i = 1$ to $g^3$ <b>do</b>	
5	Analizar la aptitud de cada individuo en $S$ y <i>ENDSF</i>	Fine-grained Strength Pareto Weighted Sum of Objectives Standard Rankings Selection Fuzzy Evaluation of Weighted Objectives
6	Preservar las soluciones no dominadas en <i>ENDSF</i>	
7	Si <i>ENDSF</i> es demasiado grande se reduce mediante truncamiento	Estimación de la densidad del vecino más próximo del frente de Pareto
8	Los registros vacíos de <i>ENDSF</i> son rellenos con soluciones dominadas	
9	Selección por torneo binario con reemplazo	Criterios de parada basados en el ECM de la aptitud de las soluciones o del <i>ENDSF</i>
10	Cruzamiento	Simple Doble
11	Mutación	Aleatoria Basada en la distancia Basada en la aptitud
12	Proceso de reparación	Factibilidad estructural y técnica
13	end for	
14	end for	
<p><math>S</math>: soluciones iniciales  <math>G</math>: generaciones  <i>ENDSF</i>: External non-dominated solution file  ECM: Error Cuadrático Medio</p>		

Fuente: adaptado de García-Alonso y colaboradores, 2010

### 4.1.3. Diseño del AEMO/AE

#### 4.1.3.1. El problema multi-objetivo

Los problemas que se plantean resolver por medio del AEMO-HS son:

#### Estudio de Demostración A:

*Hallar agrupaciones de municipios próximos con elevada dependencia o autocorrelación espacial de casos de esquizofrenia o depresión atendidos en las Unidades de Salud Mental Comunitarias de Andalucía.*

## **Estudio de Demostración B:**

*Hallar agrupaciones de municipios próximos con un número estadísticamente significativo, alto (hot-spots) o bajo (cold-spots), de casos de esquizofrenia o depresión atendidos en los Centros de Salud Mental de Cataluña.*

### **4.1.3.2. Objetivos y funciones objetivo**

#### **4.1.3.2.1. Estudio de Demostración A**

La solución al problema multi-objetivo del Estudio de Demostración A son un grupo de unidades espaciales cuyos valores en los índices espaciales señalan la existencia de una fuerte dependencia espacial en la prevalencia administrativa de la esquizofrenia y la depresión, estos valores deben ser uniformes y localizarse próximos en el espacio. En base a esta descripción las *variables de decisión* a optimizar se clasifican en tres grupos:

#### **A. Dependencia espacial**

La influencia de la localización geográfica de las unidades espaciales sobre la prevalencia administrativa de la esquizofrenia y la depresión viene expresada por los valores de los Índices Locales de Asociación Espacial (LISA) y el CAR Bayesiano. Los valores finales de estos indicadores fueron re-escalados [0.1, 0.9] para evitar la existencia de valores negativos y problemas de escala (valores muy bajos coexistiendo con valores muy altos) (García-Alonso, Guardiola, et al., 2009). Se busca que cada solución (grupos de municipios) presente los siguientes objetivos:

1. Un valor medio de la  $I$  de Moran elevado indicará la existencia de autocorrelación espacial positiva:

$$f_{o_1} = \text{Max } \bar{I} \quad (17)$$

2. Un valor medio de la  $C$  de Geary bajo indicará la existencia de autocorrelación espacial positiva:

$$fo_2 = \text{Min } \bar{C} \quad (18)$$

3. Un valor medio de la  $G^*$  de Getis y Ord elevado indicará la existencia de autocorrelación espacial positiva:

$$fo_3 = \text{Max } \bar{G} \quad (19)$$

4. Un valor medio del  $CAR$  Bayesiano elevado indicará la existencia de autocorrelación espacial positiva:

$$fo_4 = \text{Max } \overline{CAR} \quad (20)$$

### ***B. Uniformidad***

Los grupos de unidades espaciales deben tener valores de índices LISA y CAR que sean uniformes evitando grandes contrastes.

5. Una desviación estándar pequeña de la  $I$  de Moran indicará uniformidad en los valores:

$$fo_5 = \text{Min } \sigma_I \quad (21)$$

6. Una desviación estándar pequeña de la  $C$  de Geary indicará uniformidad en los valores:

$$fo_6 = \text{Min } \sigma_C \quad (22)$$

7. Una desviación estándar pequeña de la  $G^*$  de Getis y Ord indicará uniformidad en los valores:

$$fo_7 = \text{Min } \sigma_G \quad (23)$$

8. Una desviación estándar pequeña del *CAR* Bayesiano indicará uniformidad en los valores:

$$fo_8 = \text{Min } \sigma_{CAR} \quad (24)$$

### C. Proximidad

El AEMO también incorpora entre los objetivos a optimizar la minimización del recorrido mínimo entre las *unidades espaciales* buscando su proximidad. El problema consiste en encontrar la distancia mínima que permita unir las unidades espaciales de cada solución.

La distancia a minimizar se define como una distancia relativa ( $dr_{ji}$ ) que se distribuye en un rango  $[1, dr_{max}]$ , donde el nivel 1 corresponde a las unidades espaciales vecinas ( $UE_i \in UEv_j$ ) definidas previamente para el cálculo de los índices de dependencia espacial (ver la sección metodológica). Se limitó el cálculo de la distancia relativa a un máximo de 10 niveles ( $dr_{ij} \leq 10$ ) para asegurar que las unidades espaciales no estuviesen excesivamente distantes y evitar cálculos que consumieran innecesariamente tiempo de computación.

A partir del primer nivel de  $dr_{ij}$  se calculan los siguientes niveles de forma sucesiva. En primer lugar, se selecciona la distancia máxima o más larga que existe entre una *UE* y sus vecinas ( $UEv_j$ ). A esta distancia se la multiplica por 1,1 y se comprueba si en la circunferencia que tiene como radio esta nueva distancia se incluyen al menos 10 municipios. Si no los incluye se va incrementando el multiplicador en 0,2 (+  $\Delta 0.2 = 1.3, 1.5\dots$ ) hasta que alcance esta condición. Con esto se define la primera corona a partir de la que se definirán las demás. Las siguientes coronas se calculan incrementando la distancia media de la primera corona en una desviación estándar (corona  $x =$  distancia media de la corona 1 + [número de corona - 1]  $\times$  desviación estándar de las distancias de la corona 1).

Este proceso se realiza para cada *unidad espacial* por lo que todas ellas tienen asignada una distancia relativa respecto a las demás. La distancia relativa entre pares de unidades espaciales

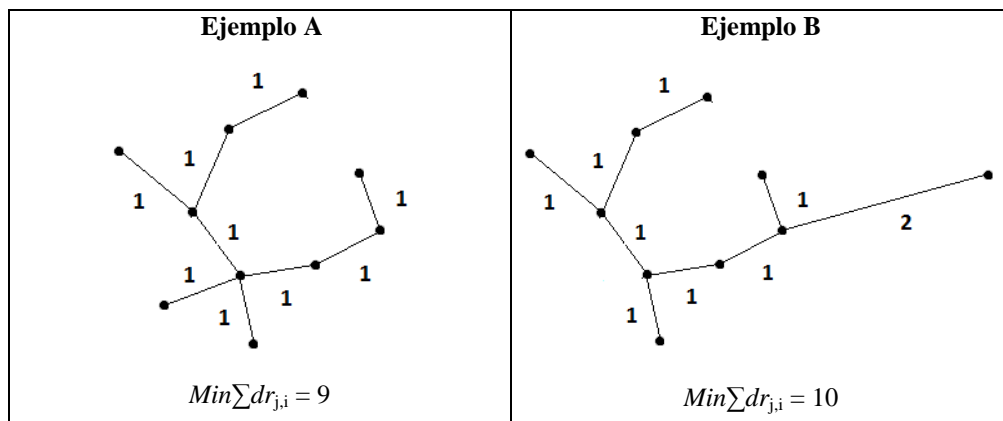
puede no ser recíproca, es decir, puede ser diferente de A a B que de B a A. Esto es debido a pequeñas variaciones que se producen en el cálculo de la distancia relativa desde A y desde B. En estos casos se calcula la media de ambas proximidades relativas para que sean recíprocas.

El AEMO/AE busca la minimización de las distancias relativas entre las unidades espaciales que componen cada solución. La distancia que se minimiza es la suma de las proximidades relativas del camino mínimo que une todas las *unidades espaciales* de la solución ( $Min\sum dr_{j,i}$ ). Este camino mínimo permite pasar varias veces por la misma *unidad espacial* (Figura 8).

9. Una distancia mínima pequeña entre todas las *unidades espaciales* indicará proximidad:

$$fo_9 = \text{Min Min } \sum dr_{ij} \quad (25)$$

Figura 8: Medición de la distancias entre unidades espaciales



Fuente: elaboración propia

#### 4.1.3.2.2. Estudio de Demostración B

En el segundo estudio las soluciones son grupos de municipios con valores de la prevalencia administrativa de la esquizofrenia y la depresión elevados (*hot-spots*) o bajos (*cold-spots*), uniformes y próximos en el espacio geográfico. Los objetivos serían los siguientes:



### A. Prevalencia

Los valores de prevalencia tienen un rango  $[0-P]$  que no precisa ser transformado para su introducción en el algoritmo. Hay una función objetivo para la búsqueda de *hot-spots* y otra para *cold-spots* que son analizadas por el AEMO/AE en dos análisis diferentes:

1. Un valor medio de la prevalencia administrativa ( $P$ ) que sea alto para los *hot-spots*:

$$fo_{1.1} = \text{Max } \bar{P} \quad (26)$$

2. Un valor medio de la prevalencia administrativa ( $P$ ) que sea bajo para los *cold-spots*:

$$fo_{1.2} = \text{Min } \bar{P} \quad (27)$$

### B. Uniformidad

La prevalencia media de las soluciones para *hot-spots* y *cold-spots* debe ser uniforme.

3. Una desviación estándar pequeña de la prevalencia administrativa ( $P$ ) indicará uniformidad en los valores:

$$fo_2 = \text{Min } \sigma_P \quad (28)$$

### C. Proximidad

La proximidad entre las unidades espaciales se define de la misma forma que para el Estudio de Demostración A (ver más arriba).

4. La proximidad viene indicada por una *Mínima Distancia* ( $\text{Min} \sum dr_{j,i}$ ) entre *unidades espaciales* pequeña:

$$f_{o_3} = \text{Min Min } \sum dr_{ij} \quad (29)$$

#### 4.1.3.3. Parámetros del AEMO/AE

El ajuste de los parámetros supone un punto crítico del diseño del AEMO/AE ya que determinarán su desempeño para encontrar soluciones óptimas o eficientes. La definición de los parámetros es uno de los temas más investigados en el campo de los algoritmos evolutivos pero no ha sido posible encontrar métodos satisfactorios para definir cada uno de los parámetros iniciales por lo éstos dependen de las preferencias del técnico que desarrolla el AEMO/AE. Los parámetros que se definen a continuación son los mismos para los dos Estudios de Demostración.

##### *Semilla.*

Se seleccionó un número suficiente de *semillas* o soluciones iniciales diseñadas al azar para garantizar la representatividad territorial de las soluciones obtenidas. La *población* fue fijada en 100 individuos o soluciones. Para iniciar el proceso se seleccionaron 5 *semillas* diferentes para evitar la posibilidad de que una única *semilla* pudiese alcanzar un óptimo prematuro o local.

##### *Individuos.*

Cada solución o *individuo* de la *población* se compone de 10 unidades espaciales con unos determinados valores de las *variables de decisión* anteriormente especificadas. Un *individuo* queda caracterizado por los valores promedio y de dispersión de las *variables de decisión*, así como por la distancia mínima entre sus unidades espaciales.

##### *Factibilidad.*

Se definieron dos criterios de factibilidad de las soluciones:

- Estructural: una solución es un conjunto de 10 unidades espaciales no repetidas ( $UE_1, UE_2, \dots, UE_{10}$ ).
- Técnica: la distancia relativa ( $dr$ ) mínima entre las *unidades espaciales* de una solución debe ser como máximo 5.

Las soluciones no factibles son mutadas por el AEMO/AE hasta que son factibles.

#### *Elitismo.*

El modelo salvaguarda las mejores soluciones para que no sean ni mutadas ni recombinadas. El elitismo comienza a considerarse a partir de la décima generación ya que antes es pronto para que las soluciones presenten valores de aptitud destacables. Las soluciones no dominadas se almacenan en el archivo *ENDSF* hasta que aparece una solución con mejor valor de aptitud. Las soluciones de esta *población externa* participan en el proceso de selección junto al resto de soluciones para mejorarlas continuamente a través de las generaciones.

Cuando el tamaño del archivo *ENDSF* es demasiado grande se reduce por truncamiento. Para ello se estima la densidad del vecino más próximo del frente de Pareto buscando agrupaciones de soluciones (*clustering*) con una aptitud alta. Las demás soluciones del archivo se sustituyen por soluciones dominadas que serán sustituidas a su vez en siguientes generaciones. De este modo, se guía al operador de selección hacia el espacio de soluciones mejor distribuidas y se promueve la diversidad en las soluciones.

#### *Criterios de parada.*

El criterio de parada se basa en el Error Cuadrático Medio (ECM) de la aptitud de las soluciones y sigue dos procedimientos. El primer procedimiento evalúa las soluciones calculando el ECM de los resultados de las *funciones de aptitud* de las soluciones más eficientes obtenidas en cada generación (excepto las primeras). El segundo procedimiento explora el archivo externo *ENDSF* de soluciones no dominadas calculando el ECM global de sus soluciones. Se ha establecido que si en al menos uno de ambos procedimientos el ECM quedaba por debajo del 2,5% durante 20 generaciones consecutivas el algoritmo evolutivo se detiene y se obtienen las soluciones finales.

#### *Operadores genéticos*

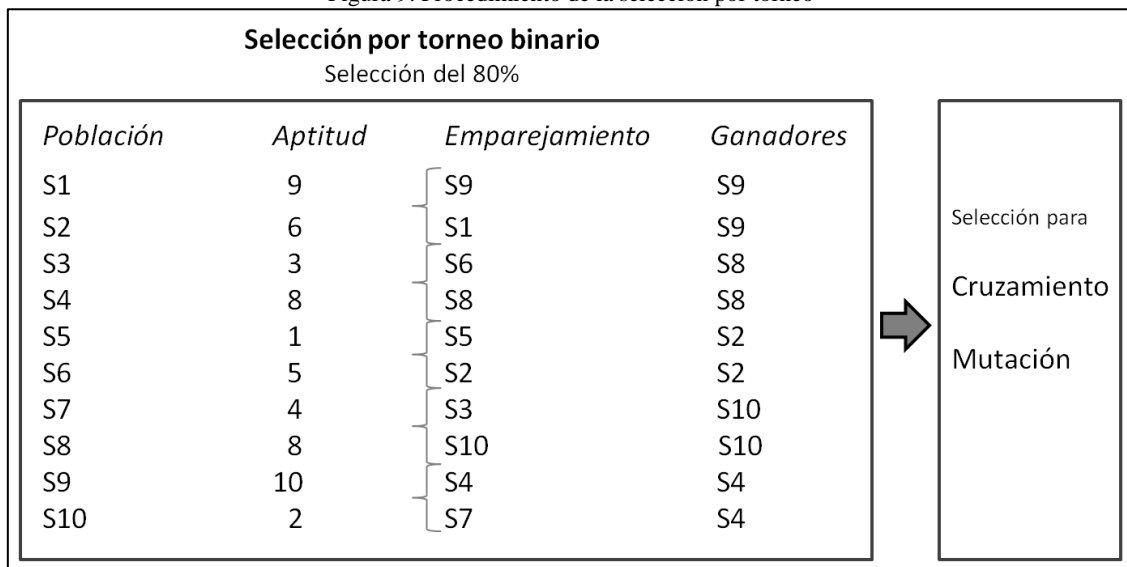
##### *Selección.*

El operador genético de selección se ejecuta mediante el método del torneo binario determinista (Coello y Lamont, 2004). La selección se efectúa en base a la comparación de la aptitud entre individuos, en la que siempre sobrevive el más apto sin opción de supervivencia del menos apto (alta presión de selección). La selección por torneo tiene como ventaja su facilidad de

implementación y eficiencia, no requiere re-escalar los valores de aptitud y permite aumentar o reducir la oportunidad de supervivencia de los menos aptos modificando el tamaño del torneo.

El método del torneo binario escoge dos soluciones al azar en base a una probabilidad o porcentaje dado, compara su aptitud y selecciona al menos apto para ser eliminado y al más apto para ser cruzado y/o mutado (Figura 9). El individuo no apto de cada torneo es sustituido por una copia del apto para mantener el tamaño de la *población* de generación en generación.

Figura 9: Procedimiento de la selección por torneo



Fuente: elaboración propia

#### *Cruzamiento.*

El AEMO/AE ha sido diseñado para que realice la recombinación de soluciones en uno (simple) o dos puntos (doble). La probabilidad de cruce simple se fijó en 0,1 y la doble en 0,005.

#### *Mutación.*

La mutación ha sido diseñada para ser aleatoria o estar basada en la distancia (*distance-based*) o en la aptitud (*fitness-based*). Estos dos últimos tipos de mutación incrementan ampliamente la intensidad de la búsqueda y si no son ajustados cuidadosamente pueden llevar a convergencias prematuras en óptimos locales (falsas soluciones). El método basado en la distancia identifica la *unidad espacial* cuya sustitución minimice la mínima distancia relativa ( $Min d_{r_{ij}}$ ) a las demás *unidades espaciales* de la solución. Por su parte, el método de mutación basado en la aptitud

identifica la *unidad espacial* cuya sustitución incrementa el valor de la *función de aptitud* de la solución. El uso de la mutación en los algoritmos genéticos es menos frecuente que el cruzamiento por lo que se fijó su probabilidad en 0,05. De las soluciones seleccionadas para mutar, se fijó que el 0,5 mutara de forma aleatoria, el 0,25 lo hiciera mediante mutación basada en la distancia y en el otro 0.25 estuviera basada en la aptitud.

#### 4.1.3.4. Funciones de aptitud

La búsqueda de soluciones factibles y no dominadas que identifican zonas altamente autocorrealadas (conjunto de Pareto) se ha realizado para ambos Estudios de Demostración mediante el empleo de cuatro funciones de aptitud diferentes. Generalmente con utilizar una única *función de aptitud* es suficiente pero se han utilizado varias para lograr unos resultados finales más robustos al estar consensuados por varias funciones.

##### 4.1.3.4.1. *Fine-grained Strength Pareto*

La *Fine-grained Strength Pareto* es la *función de aptitud* propia del Algoritmo SPEA2 (Zitzler, Laumanns, et al., 2001). Esta función opta por la estrategia de *grano fino* (*fine-grained*) para medir la fuerza (*strength*) de las soluciones. Esta estrategia parece ser la más eficiente para el problema planteado que otras estrategias como el ranking de dominancia (*dominance rank*) (Horn, Nafpliotis, et al., 1994; Zydallis, Veldhuizen, et al., 2001) o la intensidad de dominancia (*dominance depth*) (Deb, Pratap, et al., 2002). El *strength* es un valor que señala para cada solución no dominada de la *población externa* ( $P'$ ) el número de soluciones de la *población* ( $P$ ) a las que domina (mejor o igual) valorando así su utilidad. El valor de *strength* se re-escala entre  $[0, 1)$ . El valor de *strength* de una solución  $i$  de la *población externa* ( $P'$ ) se calcula como:

$$P'[i]_{strength} = \frac{|\{j \in \{1, \dots, N\} \wedge P'[i] \succ P[j]\}|}{N+1} \quad (30)$$

La *función de aptitud* de la solución  $i$  es la inversa del *strength*:

$$P'[i]_{fitness} = \frac{1}{P'[i]_{strength}} \quad (30.1)$$

Mientras que el valor de *strenght* de una solución  $j$  de la *población* considera las soluciones no dominadas que lo dominan y se calcula:

$$P[j]_{strength} = 1 + \sum_{i, P_i \geq P_j} P'[i]_{strength} \quad (31)$$

Su función de aptitud es de nuevo la inversa:

$$P[j]_{fitness} = \frac{1}{P[j]_{strength}} \quad (31.1)$$

#### 4.1.3.4.2. *Weighted Sum of Objectives*

La *Weighted Sum of Objectives* (WO) mide la aptitud de las soluciones combinando los valores de los diferentes objetivos previamente ponderados (Das y Dennis, 1997). La función consiste en la suma de las *funciones objetivo*, a donde cada objetivo se le asigna un peso en función a su importancia relativa. La ecuación sería:

$$WO(X_i) = \sum_{m=1}^M W_m f_m(X_i) \quad (32)$$

Donde  $W_m$  es el coeficiente o peso de la *función objetivo*  $f_M$  de la solución  $X_i$ . En un problema de minimización una solución  $X_i$  domina a  $X_j$  si  $WO(X_i) < WO(X_j)$ . Para aplicar esta función es necesario que los valores de las funciones objetivo sean re-escalados a un mismo rango.

#### 4.1.3.4.3. *Standard Rankings Selection*

En la *Standard Rankings Selection* (SRS) la aptitud de una solución viene dada por el sumatorio de sus posiciones jerárquicas para cada objetivo (Bentley y Wakefield, 1997). De manera que:

$$SRS(X_i) = \sum_{m=1}^M r_m f_m(X_i) \quad (33)$$

Donde  $r_M$  es la posición jerárquica de la *función objetivo*  $f_M$  de la solución  $X_i$ . La solución  $X_i$  domina a  $X_j$  cuando  $SRS(X_i) < SRS(X_j)$  en un problema de minimización. Los valores de los objetivos tienen que re-escalarsen a un mismo rango.

#### 4.1.3.4.4. *Fuzzy Evaluation of Weighted Objectives*

La última *función de aptitud* es la *Fuzzy Evaluation of Weighted Objectives* (FI). Esta función compara las soluciones considerando todas las funciones objetivo de forma conjunta y no los valores de cada función de forma independiente como las anteriores (Lee y Esbensen, 1997; Wang y Terpenney, 2005). El valor de comparación entre las soluciones  $X_i$  y  $X_j$  es la diferencia difusa, expresada como:

$$FI(X_i) = \max_{X_j \neq X_i} \prod_{m=1}^M \left[ \frac{f_m(X_i)}{f_m(X_j)} \right] \quad (34)$$

Donde,

$$\left[ \frac{a}{b} \right] = \begin{cases} a/b & a < b \\ 1 & \text{de otro modo} \end{cases} \quad (34.1)$$

De nuevo, para un problema de minimización la mejor solución será la que tenga un valor más bajo de *FI*. Cuando  $FI(X_i) < FI(X_j)$  significa que la solución  $X_i$  domina a  $X_j$ . El valor de la función objetivo debe ser normalizado al mismo rango.

### 4.1.4. Selección de las soluciones finales

#### 4.1.4.1. Soluciones del AEMO/AE

Una solución es eficiente cuando es factible y no dominada, es decir, cuando ninguna otra solución tiene todas sus *funciones objetivo* mejores que las suyas. La solución del problema planteado no es única obteniéndose varios conjuntos de unidades espaciales seleccionadas de todas las posibles combinaciones sin repeticiones de las unidades espaciales totales ( $UE_1, UE_2, \dots, UE_{10}$ ).

Las soluciones del AEMO/AE para cada *función de aptitud* son consideradas como *hot-spot* o *cold-spot potenciales* (subconjunto del conjunto de Pareto). Una vez determinadas las soluciones, cada unidad espacial se analiza de forma individual para determinar cuántas veces aparece en cualquiera de los *hot-spot* o *cold-spot potenciales* obteniéndose así un histograma de frecuencias a partir del cual se identificarán aquellas que se repiten con mayor frecuencia. Las *unidades espaciales* que aparecen con mayor frecuencia están en el extremo derecho del histograma. La distribución estadística de este extremo se analizó mediante dos *estadísticas de los extremos* (QQ-Plots) y un método de selección diseñado a tal fin para fijar el valor de repeticiones a partir del cual una unidad espacial se considera *candidata* para formar parte de los *hot-spots* o *cold-spots*.

#### **4.1.4.2. Selección mediante estadísticas de los extremos: QQ-Plots**

Los *Quantile-Quantile Plots* (QQ-Plots) (Beirlant, Goegebeur, et al., 2004) son un técnica estadística gráfica que parte de que en muchas distribuciones de datos los cuantiles están linealmente relacionados con los cuantiles de las distribuciones teóricas de probabilidad. El QQ-Plot es una técnica muy potente para comparar dos distribuciones de probabilidad o comprobar la bondad en el ajuste de una distribución a alguna de las distribuciones de probabilidad teóricas (normal, exponencial, log-normal, Pareto, etc.).

La distribución de las apariciones de las unidades espaciales en las soluciones se ha ajustado a dos distribuciones de probabilidad teóricas, por una parte a una distribución exponencial y por otra a una distribución de Pareto. Las unidades espaciales que se seleccionan como soluciones finales son aquellas localizadas en la cola de la distribución de frecuencias del número de apariciones en la soluciones del AEMO/AE que se ajustan a la linealidad exponencial y de Pareto.

El procedimiento seguido consistió en ordenar las observaciones o unidades espaciales de menor a mayor frecuencia. En segundo lugar, se descartaron las unidades espaciales que solo habían aparecido en una ocasión con la finalidad de reducir la pendiente de la curva exponencial y de Pareto y así hacer más estricta la selección de las unidades espaciales más frecuentes.

La distribución teórica exponencial de las frecuencias de las unidades espaciales es:



$$y = \frac{\sum_{i=1}^n [V_i (-1 \times \log[1 - \frac{i}{n+1}])] - V_{min} \times \sum_{i=1}^n (-1 \times \log[1 - \frac{i}{n+1}])}{\sum_{i=1}^n (-1 \times \log[1 - \frac{i}{n+1}])^2} + V_{min} \quad (35)$$

Donde,  $V_i$  es el número de veces que se repite el municipio en el conjunto de soluciones;  $i$  es el número de orden del municipio ordenado de menor a mayor por  $V_i$ ;  $V_{min}$  es el número de veces mínimo a partir del que se realiza el análisis (definido que sea a partir de 2 veces); y  $n$  es el número de observaciones (*unidades espaciales*).

Por su parte la distribución teórica de Pareto de las observaciones sería:

$$y = \frac{\sum_{i=1}^n (\log(V_i) - \log(V_{min})) \times \sum_{i=1}^n (-1 \times \log[1 - \frac{i}{n+1}])}{\sum_{i=1}^n [\log(V_i)]^2} + \log V_{min} \quad (36)$$

Donde,  $V_i$ ,  $i$ ,  $V_{min}$  y  $n$  tienen el mismo significado que en la expresión exponencial.

A continuación, se calculan los residuos de cada observación restando su valor real al valor teórico calculado mediante las expresiones anteriores siguiendo un orden de menor a mayor frecuencia. Los residuos se elevan al cuadrado para hacerlos positivos.

Por último, se evaluaron las unidades espaciales más frecuentes situadas en la cola derecha de la distribución de frecuencias. Empezando por la observación más frecuente, se calculó el error de la distribución real respecto a la teórica exponencial y de Pareto calculando la suma de los residuos al cuadrado de todas las observaciones excepto las eliminadas previamente (frecuencia igual a 1). En un segundo paso, se eliminó la última observación y se volvió a calcular los residuos al cuadrado siguiendo un proceso iterativo en base a eliminar la última observación. El punto de corte final se estableció cuando la suma de los residuos al cuadrado o error no descendió de un paso a otro. Las observaciones eliminadas hasta llegar a ese punto fueron las seleccionadas como unidades espaciales *candidatas* a constituir un *hot-spot* o un *cold-spot* debido a su elevada frecuencia. Al realizar el procedimiento de selección considerando el ajuste de la distribución a dos curvas teóricas se obtuvieron dos grupos de unidades espaciales *candidatas*.

#### 4.1.4.3. Método de selección ad hoc

El tercer método realiza la selección en base al porcentaje de apariciones de cada unidad espacial respecto al número total de unidades que componen las soluciones de una determinada *función de aptitud*. En primer lugar, se ordena la muestra de mayor a menor porcentaje de veces de aparición. A continuación, se distribuye la muestra en 30 intervalos iguales en base al porcentaje mínimo y máximo. Entonces, empezando por la izquierda de la distribución (mayor porcentaje) se establece un primer punto de corte donde se sucedan dos intervalos seguidos vacíos, es decir, sin recoger unidades espaciales. Un segundo punto de corte se delimita contando 10 intervalos desde el límite derecho de la distribución (mínimo porcentaje). Si el nuevo corte se encuentra a la izquierda del primero es seleccionado, en cambio, si se encuentra a su derecha se selecciona el primer corte. Es resumen, se seleccionaba el punto de corte situado más a la izquierda de la distribución de frecuencias. Las observaciones que quedaron a la izquierda del punto de corte se consideraron unidades espaciales *candidatas*.

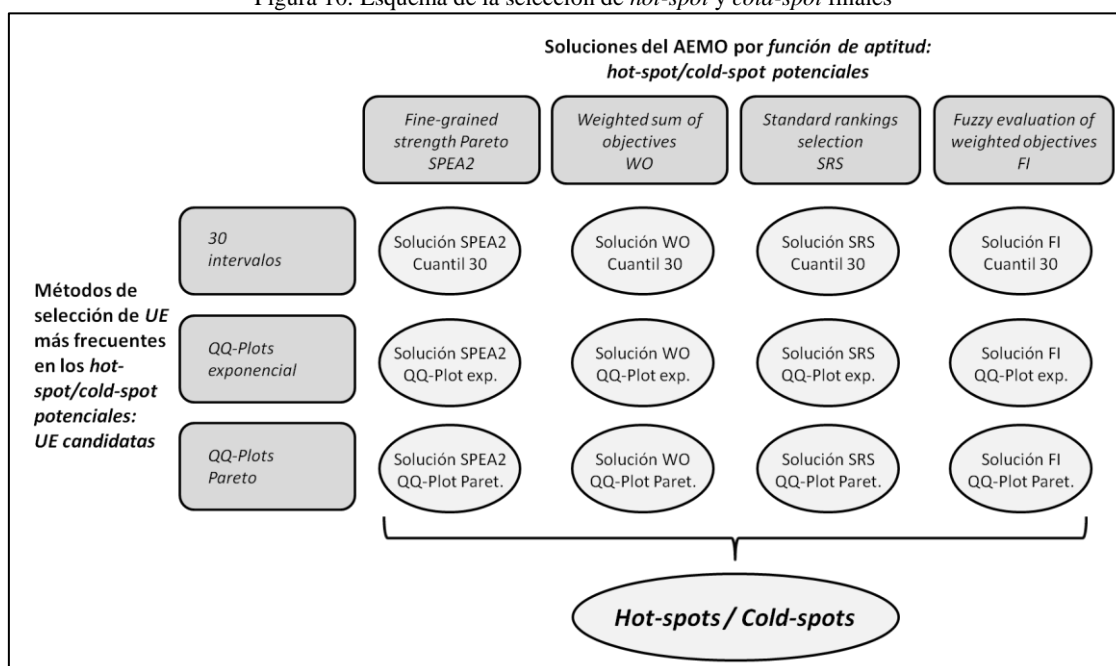
#### 4.1.4.4. Selección de hot-spots

Los tres métodos de selección aplicados sobre las soluciones de cada *función de aptitud* dieron como resultado un grupo reducido de unidades espaciales *candidatas* a formar parte de los *hot-spot* o *cold-spot*. Los grupos de unidades espaciales seleccionados por cada método para cada *función de aptitud* son muy similares, ya que recogen unidades que se sitúan en la cola superior de la distribución de frecuencias diferenciándose únicamente por el punto de corte. De esta forma, se obtienen aquellas unidades que han aparecido en mayor número de ocasiones entre las soluciones del AEMO/AE.

En este momento del procedimiento metodológico se contaba con cuatro grupos de soluciones o *hot-spot potenciales*, una para cada *función de aptitud*, con tres selecciones diferentes de unidades espaciales *candidatas* a *hot-spot* o *cold-spot*, una por cada método de selección (Figura 10).

Una *unidad espacial* forma parte de un *hot-spot* o un *cold-spot* cuando ha sido identificada por el AEMO/AE, empleando cualquiera de las cuatro *funciones de aptitud*, con una frecuencia significativamente alta señalada por alguno de los tres métodos de selección. Con esta última exigencia se consigue descartar aquellas unidades espaciales identificadas por el AEMO/AE en una única ocasión o de forma esporádica en alguno de los grupos de soluciones.

Figura 10: Esquema de la selección de *hot-spot* y *cold-spot* finales



Fuente: Elaboración propia

La selección final de las unidades espaciales que constituyen los *hot-spot* o *cold-spot* se realiza estableciendo dos niveles en función al porcentaje de *funciones de aptitud* y métodos de selección que las identifican. En el primero, una unidad espacial es considerada como solución final cuando al menos el 50% de los métodos de selección aplicados a las soluciones de las cuatro *funciones de aptitud* la han identificado como *hot-spot* o *cold-spot* (nivel 1). Mientras que el nivel 2 agrupa a aquellas unidades recogidas en las soluciones pero que no aparecen seleccionadas por al menos el 50% de los métodos aplicados a los resultados de las cuatro funciones. El resto de unidades espaciales son descartadas porque, aunque aparecen entre las soluciones, no son seleccionadas por ninguno de los tres métodos debido a su baja frecuencia.

Los *hot-spot* y *cold-spot* se componen de las unidades espaciales incluidas en los dos niveles señalados por lo que existe una graduación dentro de las soluciones. El segundo nivel recoge unidades espaciales que no han aparecido entre las soluciones de alguna de las *funciones de aptitud* o no lo han hecho con la suficiente frecuencia para ser señaladas por los métodos de selección. En todo caso se tratan de soluciones altamente eficientes al problema multi-objetivo planteado. Una vez identificadas las *unidades espaciales* consideradas como *hot-spots* y *cold-spots* pueden ser localizados geográficamente mediante un Sistema de Información Geográfica (SIG).

## 4.2. Estudio de Demostración A: aplicación del AEMO/AE sobre datos de Enfermedades Mentales en Andalucía

### 4.2.1. Caso 1: *Hot-spots* de prevalencia asistida de la esquizofrenia en Andalucía durante una serie temporal de cuatro años

#### 4.2.1.1. Distribución geográfica de la prevalencia administrativa de la esquizofrenia

En la Tabla 7 se exponen las estadísticas descriptivas de la prevalencia administrativa de la esquizofrenia por unidad espacial para cada año analizado. Se aprecia que estas estadísticas se mantienen en el tiempo con cambios mínimos, con una prevalencia media que se mueve entre 2,77 y 3,09 casos atendidos por 1.000 habitantes, y una desviación estándar que oscila entre 2,3 y 2,38. El valor máximo del rango de las observaciones fue mayor en 2004 con 23,44 casos por 1.000 para estabilizarse en los siguientes años entre 16,95 y 18,99 casos por 1.000. El valor mínimo de la población corresponde a municipios sin casos de esquizofrenia. El número de municipios sin casos se mueve entre los 73 de 2007 y los 95 de 2008. Los valores 0 corresponden en muchos casos a municipios de las USMC que no informaron a SISMA y en otros casos a municipios poco poblados.

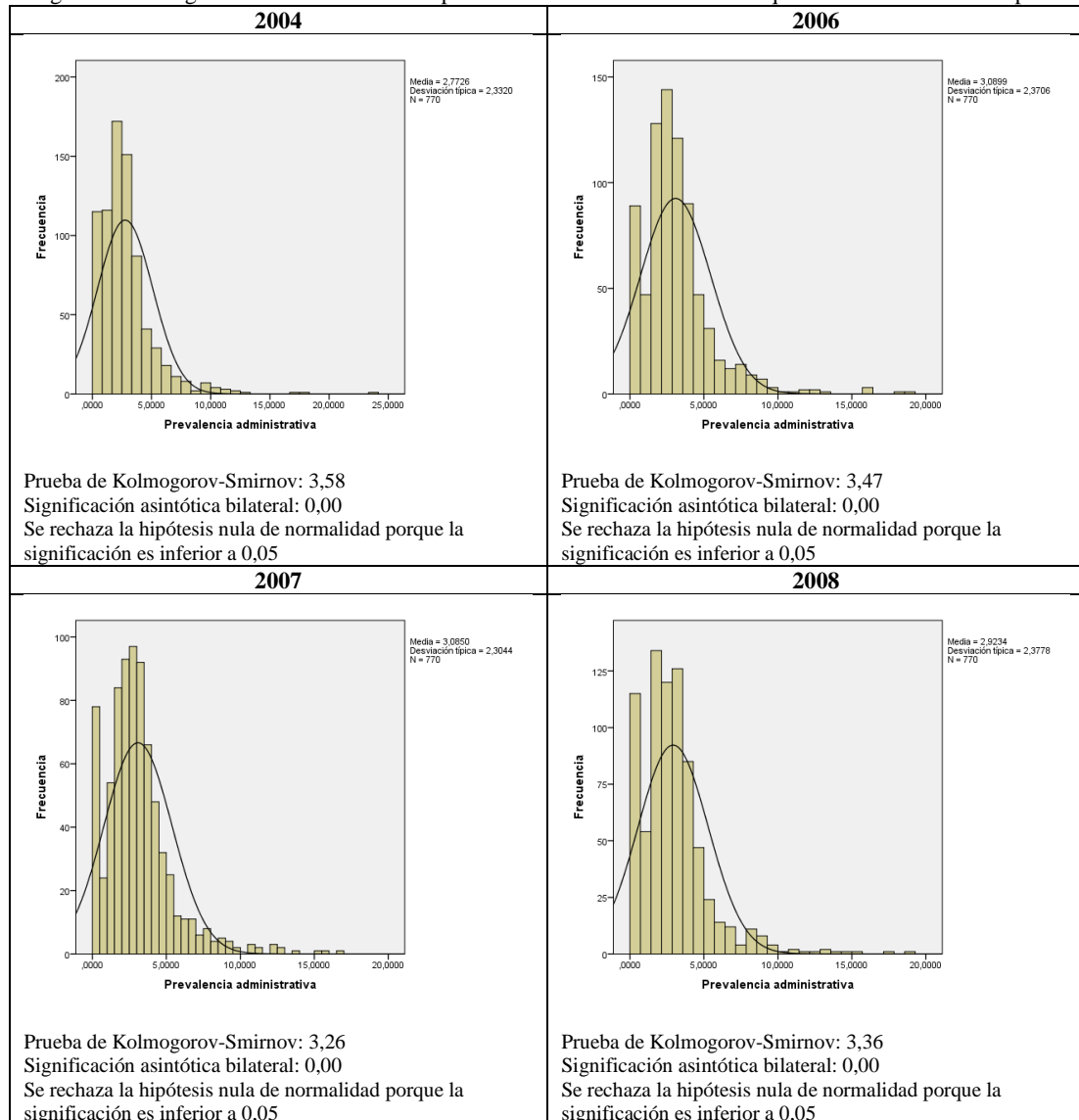
Tabla 7: Estadísticas descriptivas de la prevalencia administrativa de la esquizofrenia en Andalucía

	2004	2006	2007	2008
Recuento	770	770	770	770
Media	2,77	3,09	3,08	2,92
Desviación estándar	2,33	2,37	2,30	2,38
Mediana	2,43	2,75	2,77	2,62
Mínimo	0 (88 UE)	0 (79 UE)	0 (73 UE)	0 (95 UE)
Máximo	23,44	18,75	16,95	18,99

Fuente: SISMA

En la Figura 11 se presentan los histogramas de las 4 series de prevalencia administrativa de la esquizofrenia. Se ha efectuado el test de Kolmogorov-Smirnov sobre las mismas para comprobar su bondad de ajuste a una distribución normal. En todos los casos se ha rechazado la hipótesis nula de normalidad al estar la significación de la prueba por debajo de 0,05. El elevado número de casos con valor 0 distorsiona la forma de la distribución que parece aproximarse más a una curva  $\beta$ .

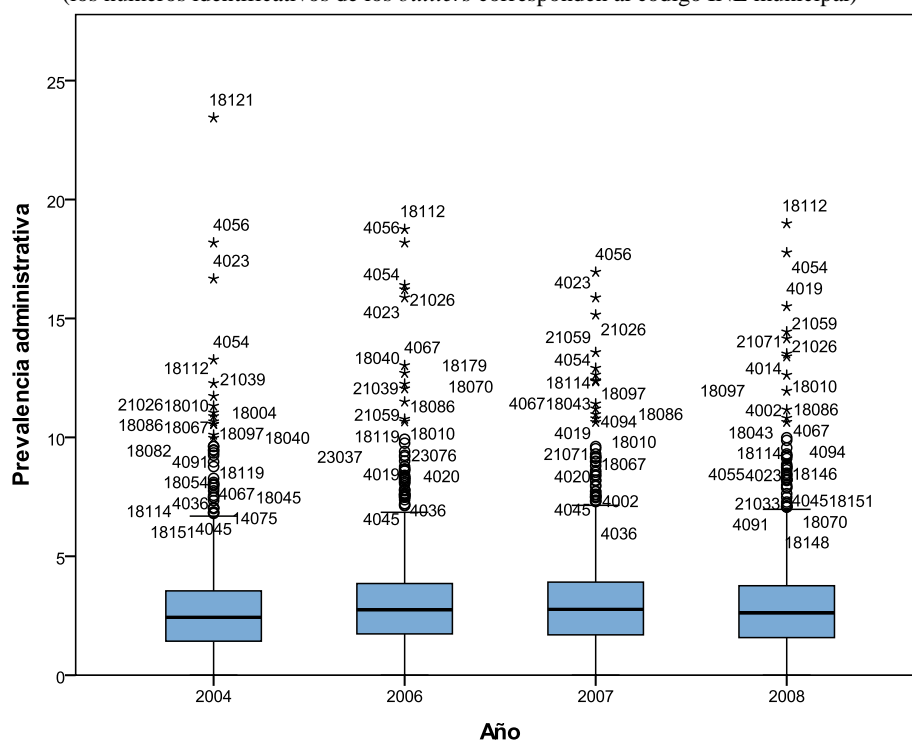
Figura 11: Histogramas de frecuencia de la prevalencia administrativa de la esquizofrenia en la serie temporal



Fuente: SISMA

La dispersión de las cuatro poblaciones se puede comprobar mediante los gráficos *box-plot* en el que se representa la mediana, cuartiles y valores atípicos (Figura 12). En los mismos se observa la semejanza de las cuatro distribuciones y la existencia de numerosos valores atípicos u *outliers* de valores elevados de prevalencia. Los valores atípicos ( $V_{Amin}$  y  $V_{Amax}$ ) son aquellos valores por encima de 1,5 veces el recorrido intercuartílico ( $R_Q$ ) entre el primer y el tercer cuartil ( $Q_3 - Q_1$ ) ya sea por la izquierda o la derecha de la distribución ( $V_{Amin} = Q_1 - 1,5 R_Q$  y  $V_{Amax} = Q_3 + 1,5 R_Q$ ). Por su parte, los valores extremos ( $V_{Emin}$  y  $V_{Emax}$ ) superan 3 veces el recorrido intercuartílico ( $V_{Emin} = Q_1 - 3 R_Q$  y  $V_{Emax} = Q_3 + 3 R_Q$ ).

Figura 12: *Box-plot* de las poblaciones de prevalencia administrativa de la esquizofrenia en la serie temporal (los números identificativos de los *outliers* corresponden al código INE municipal)



Fuente: SISMA

La prevalencia administrativa de la esquizofrenia para la serie de cuatro años ha sido representada en mapas para facilitar el conocimiento de su distribución espacial (Mapa 3). En los mapas se puede observar que las series de esquizofrenia muestran cierta estabilidad temporal en la distribución de las prevalencias por áreas geográficas. El mapa representa los datos mediante una distribución de frecuencias que considera las veces que se aleja cada valor de la media en función de la desviación estándar (Tabla 8).

Tabla 8: Distribución de frecuencias de la prevalencia administrativa de la esquizofrenia en Andalucía

	2004	2006	2007	2008
-1,5 a -0,5 Desv. Est.	218	218	227	216
-0,5 a 0,5 Desv. Est.	401	400	387	404
0,5 a 1,5 Desv. Est.	101	101	102	99
1,5 a 2,5 Desv. Est.	28	34	32	32
>2,5 Desv. Est.	22	17	22	19

Fuente: SISMA

Los valores significativamente elevados, superiores a la media de la distribución en 2,5 veces la desviación estándar, corresponden a cerca de 20 municipios cada año. Los municipios que aparecen entre los de mayor prevalencia de esquizofrenia a lo largo de la serie temporal son: Aldeire y Gorafe en la comarca de Guadix (Granada), Juviles y Torvizcón en la Alpujarra granadina, Cortelazor y Puerto Moral en Aracena (Huelva), e Íllar en la Alpujarra almeriense.

Los valores más pequeños de prevalencia de esquizofrenia corresponden a municipios sin pacientes. Es preciso recordar que existen USMC que no informaron al SISMA y esta limitación se materializa en las concentraciones de valores bajos en la provincia de Málaga en torno a Antequera en los cuatro años y en las provincias de Cádiz y Granada en 2008. Considerando las salvedades anteriores, se puede observar que las unidades espaciales con prevalencia baja se distribuyen por todo el territorio andaluz.

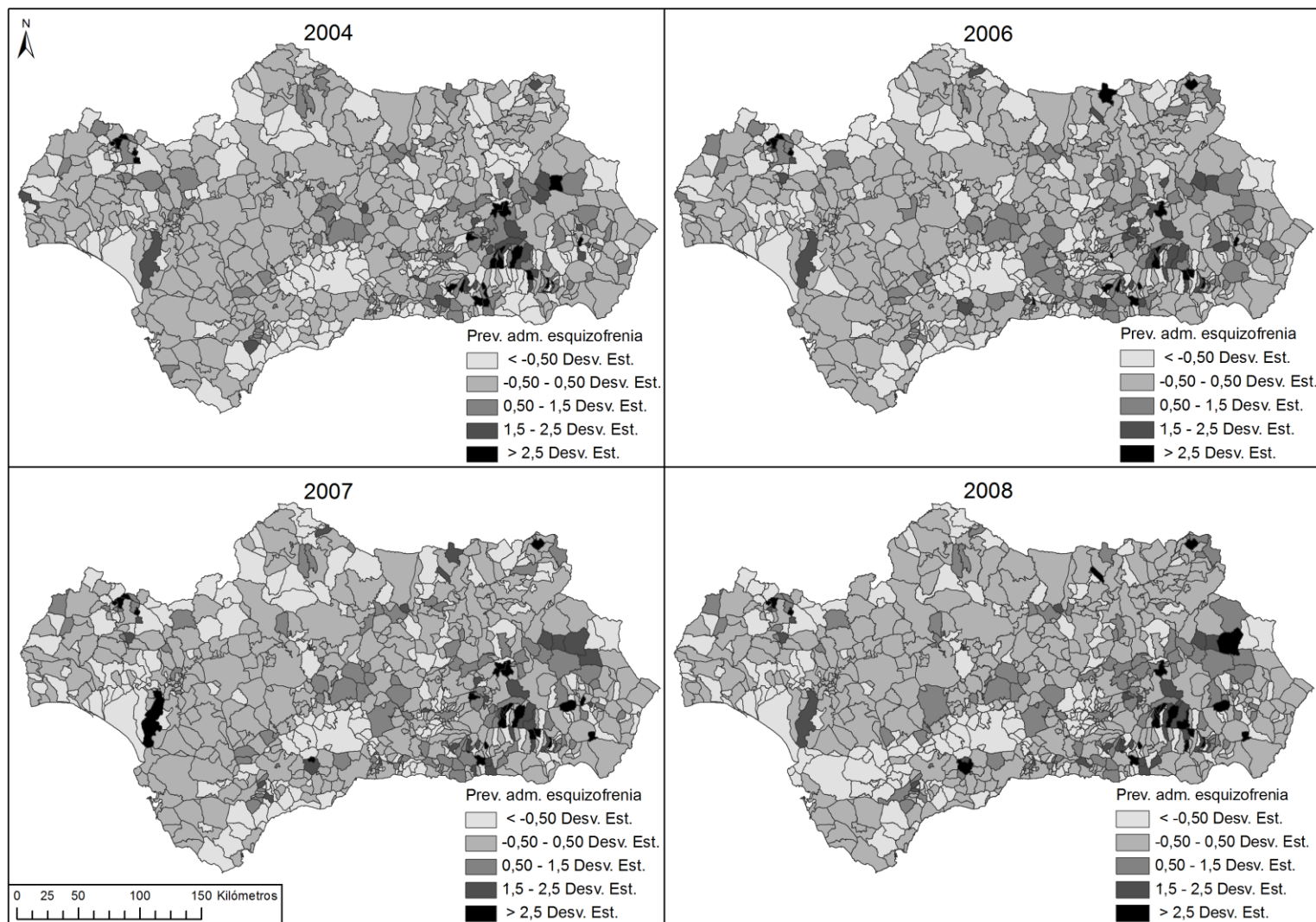
El análisis espacial exploratorio realizado sobre la prevalencia administrativa de la esquizofrenia, mediante la *I* de Moran y la *G* de Getis y Ord globales, señala la existencia de dependencia espacial con agrupaciones o clústeres de valores elevados de prevalencia en los cuatro años de la serie analizada (Tabla 9). Este análisis descarta la hipótesis nula de que la prevalencia se distribuye de forma aleatoria ( $p = 0,01$ ).

Tabla 9: Análisis de la autocorrelación espacial global de la prevalencia asistida de la esquizofrenia en la serie temporal (Índices globales de la *I* de Moran y la *G* de Getis y Ord)

	2004	2006	2007	2008
<i>I</i> de Moran global	Existe agrupación espacial de la prevalencia  $I = 0,04$ Valor $z = 13,46$ d.e. Valor $p = 0,01$	Existe agrupación espacial de la prevalencia  $I = 0,03$ Valor $z = 10,99$ d.e. Valor $p = 0,01$	Existe agrupación espacial de la prevalencia  $I = 0,03$ Valor $z = 11,39$ d.e. Valor $p = 0,01$	Existe agrupación espacial de la prevalencia  $I = 0,04$ Valor $z = 14,92$ d.e. Valor $p = 0,01$
<i>G</i> de Getis y Ord global	Agrupaciones de valores altos de prevalencia  $G = 0$ Valor $z = 8,30$ d.e. Valor $p = 0,01$	Agrupaciones de valores altos de prevalencia  $G = 0$ Valor $z = 6,33$ d.e. Valor $p = 0,01$	Agrupaciones de valores altos de prevalencia  $G = 0$ Valor $z = 5,44$ d.e. Valor $p = 0,01$	Agrupaciones de valores altos de prevalencia  $G = 0$ Valor $z = 7,43$ d.e. Valor $p = 0,01$
d.e. = desviación estándar.				

Fuente: elaboración propia

Mapa 3: Prevalencia administrativa de la esquizofrenia en los municipios de Andalucía



Fuente: Elaboración propia a partir del SISMA (2004, 2006, 2007 y 2008)



#### 4.2.1.2. Dependencia espacial de la prevalencia administrativa de la esquizofrenia

Como se ha explicado en la metodología, se han aplicado tres índices de autocorrelación espacial y un modelo CAR Bayesiano a los valores municipales de prevalencia administrativa. La existencia de dependencia espacial positiva se manifiesta en los valores más altos de los mismos, excepto para la  $C$  de Geary que se interpreta de forma opuesta. Los valores más elevados de la  $G^*$  de Getis y Ord y el riesgo relativo bayesiano además de dependencia espacial positiva también indican la asociación de prevalencias significativamente elevadas.

Las estadísticas descriptivas básicas de cada uno de los cuatro métodos de análisis espacial antes de re-escalar sus valores se ofrecen en la Tabla 10, Tabla 11, Tabla 12 y Tabla 13. Las diferencias interanuales observadas en las estadísticas descriptivas son mínimas y presentan un comportamiento uniforme. La dispersión de las 16 series de índices espaciales se muestra en los *box-plots* (Figura 13, Figura 14, Figura 15 y Figura 16). Las medianas y valores máximos y mínimos son semejantes para cada índice durante los cuatro años. También se aprecia la existencia de numerosos valores atípicos y extremos.

Los cuatro métodos de análisis de la dependencia espacial que se han empleado ofrecen diferencias en cuanto a su distribución territorial a lo largo de la serie temporal, como se puede comprobar en el Mapa 4, Mapa 5, Mapa 6, y Mapa 7. Es posible señalar algunas coincidencias en los métodos por grandes áreas geográficas donde se detectan agrupaciones de valores elevados y similares en las áreas de Guadix, La Alpujarra granadina y La Alpujarra almeriense, todas ellas vecinas, y la Sierra de Aracena. Estas agrupaciones de valores ya se intuían en el análisis de las tasas de prevalencia anterior. Los métodos no coinciden en identificar la existencia de dependencia espacial en el área de la Costa Tropical granadina o la Marisma del Guadalquivir en Sevilla.

Tabla 10: Estadísticas descriptivas de la  $I$  de Moran  
(prevalencia administrativa de la esquizofrenia)

	2004	2006	2007	2008
Recuento	770	770	770	770
Media	0,16	0,11	0,13	0,17
Desviación estándar	0,66	0,55	0,63	0,66
Mediana	0,038	0,019	0,028	0,030
Mínimo	-2,3	-2,3	-2,5	-2,5
Máximo	6,6	4	5,5	7,4

Fuente: Elaboración propia

Tabla 11: Estadísticas descriptivas de la *C* de Geary  
(prevalencia administrativa de la esquizofrenia)

	2004	2006	2007	2008
Recuento	770	770	770	770
Media	9,2	10	9,4	9,5
Desviación estándar	21	19	16	19
Mediana	3,5	4,4	4,0	3,8
Mínimo	0,0033	0,00075	0,00073	0,00014
Máximo	370	250	220	260

Fuente: Elaboración propia

Tabla 12: Estadísticas descriptivas de la *G\** de Getis y Ord  
(prevalencia administrativa de la esquizofrenia)

	2004	2006	2007	2008
Recuento	770	770	770	770
Media	0,060	0,060	0,059	0,059
Desviación estándar	0,022	0,019	0,018	0,020
Mediana	0,058	0,059	0,057	0,056
Mínimo	0,020	0,021	0,022	0,017
Máximo	0,15	0,14	0,13	0,14

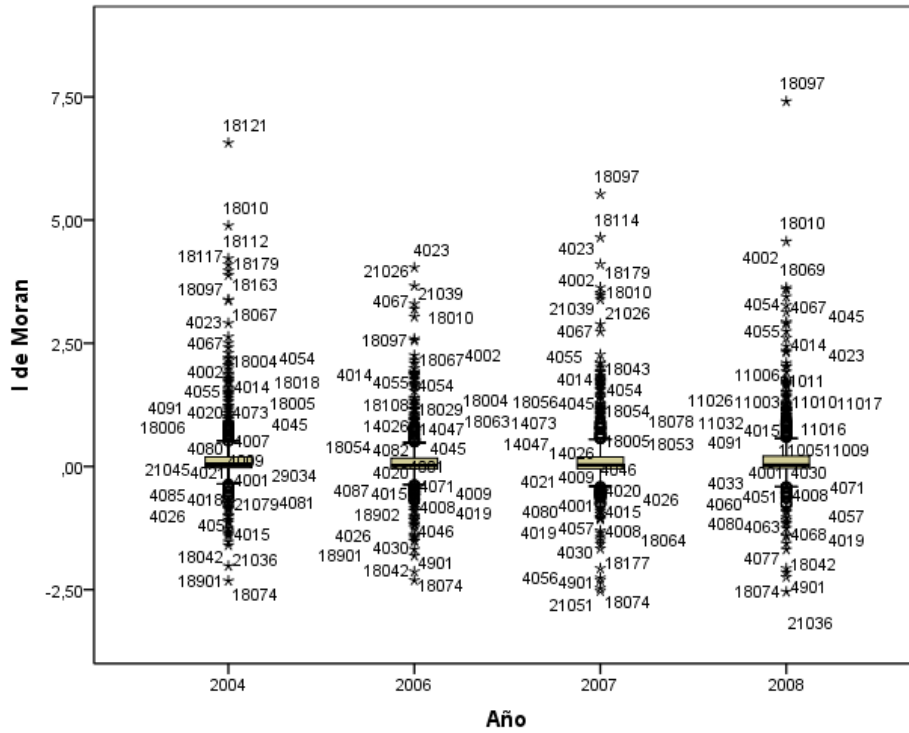
Fuente: Elaboración propia

Tabla 13: Estadísticas descriptivas del CAR Bayesiano  
(prevalencia administrativa de la esquizofrenia)

	2004	2006	2007	2008
Recuento	770	770	770	770
Media	0,97	0,98	0,98	0,99
Desviación estándar	0,37	0,33	0,34	0,47
Mediana	0,92	0,94	0,95	0,94
Mínimo	0,087	0,079	0,073	0,017
Máximo	2,8	2,3	2,9	3,7

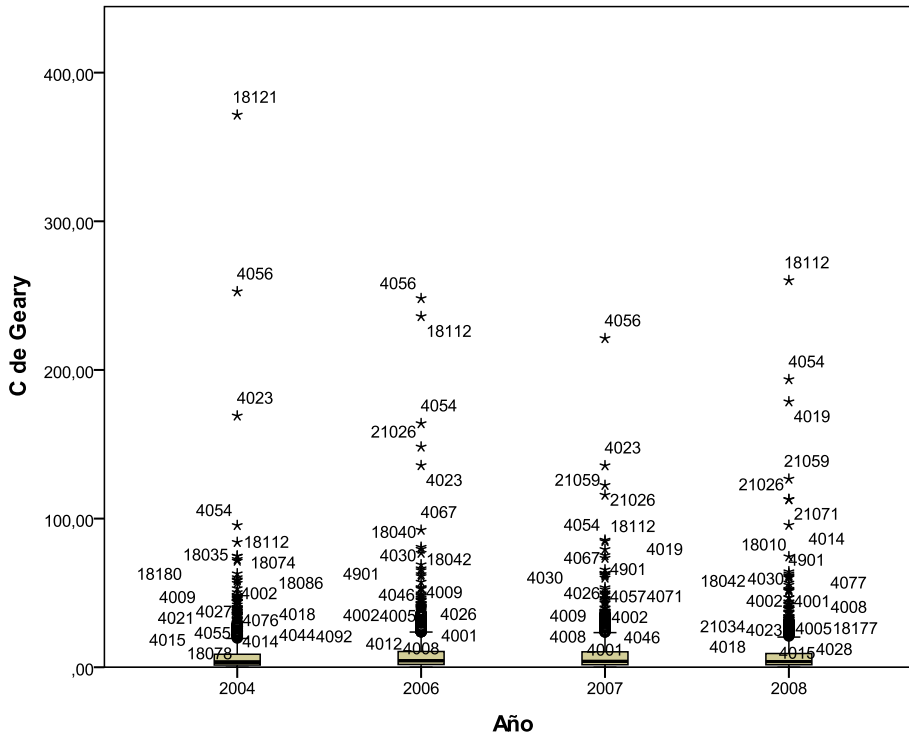
Fuente: Elaboración propia

Figura 13: *Box-Plot* de los valores de la *I* de Moran para la prevalencia de la esquizofrenia (los números identificativos de los *outliers* corresponden al código INE municipal)



Fuente: Elaboración propia

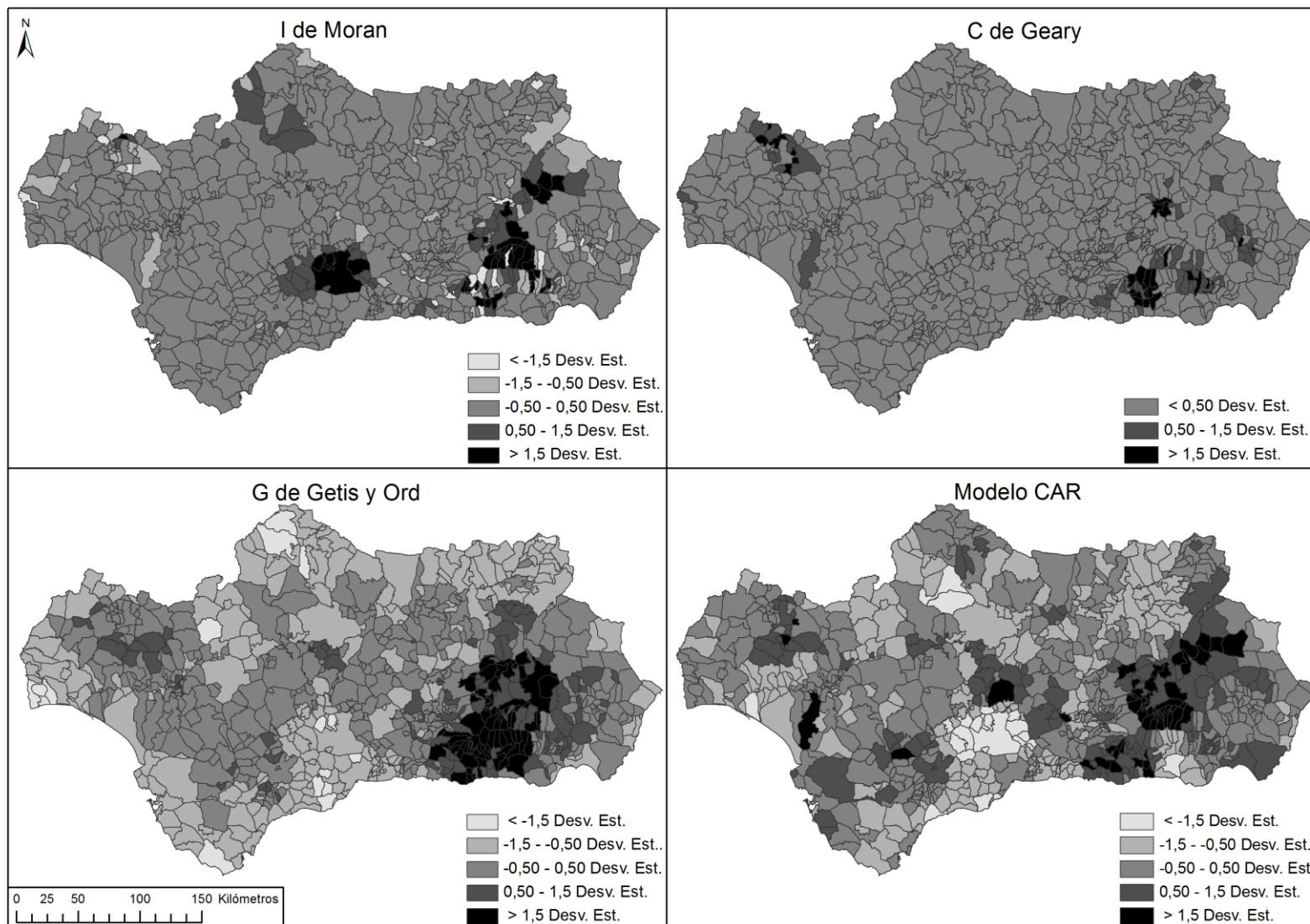
Figura 14: *Box-Plot* de los valores de la *C* de Geary para la prevalencia de la esquizofrenia (los números identificativos de los *outliers* corresponden al código INE municipal)



Fuente: Elaboración propia

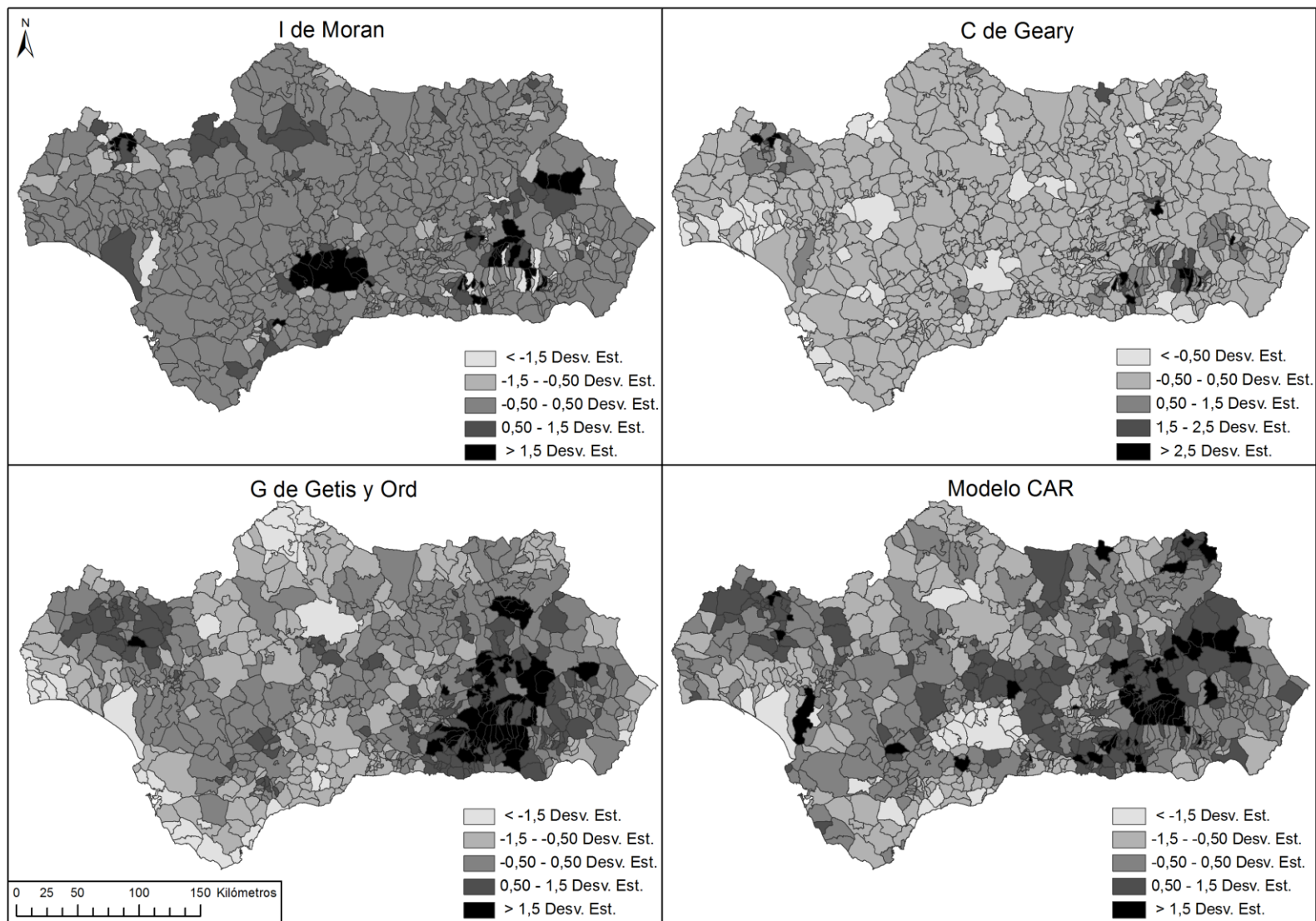


Mapa 4: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la esquizofrenia en los municipios de Andalucía (2004)



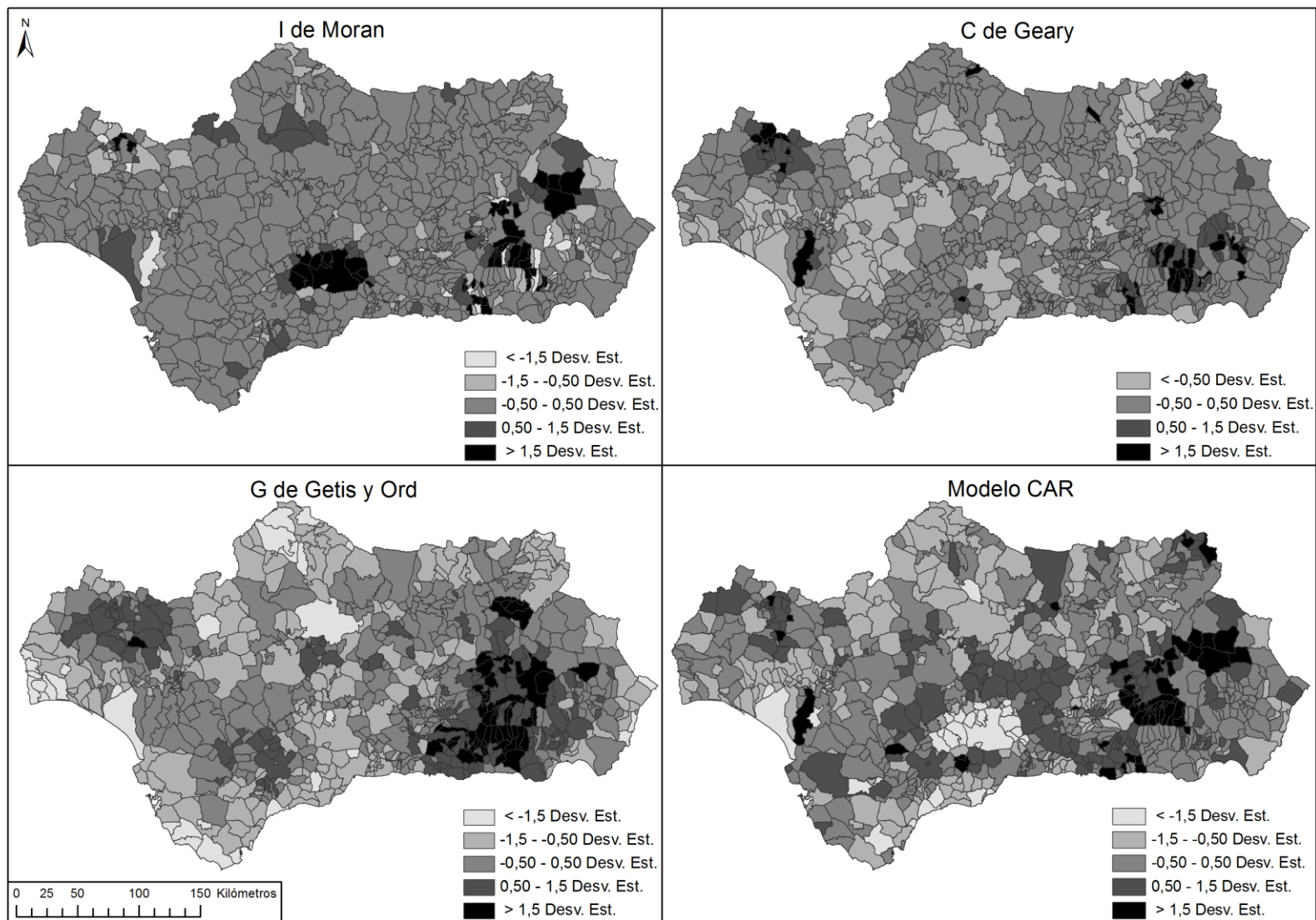
Fuente: Elaboración propia a partir del SISMA, 2004

Mapa 5: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la esquizofrenia en los municipios de Andalucía (2006)



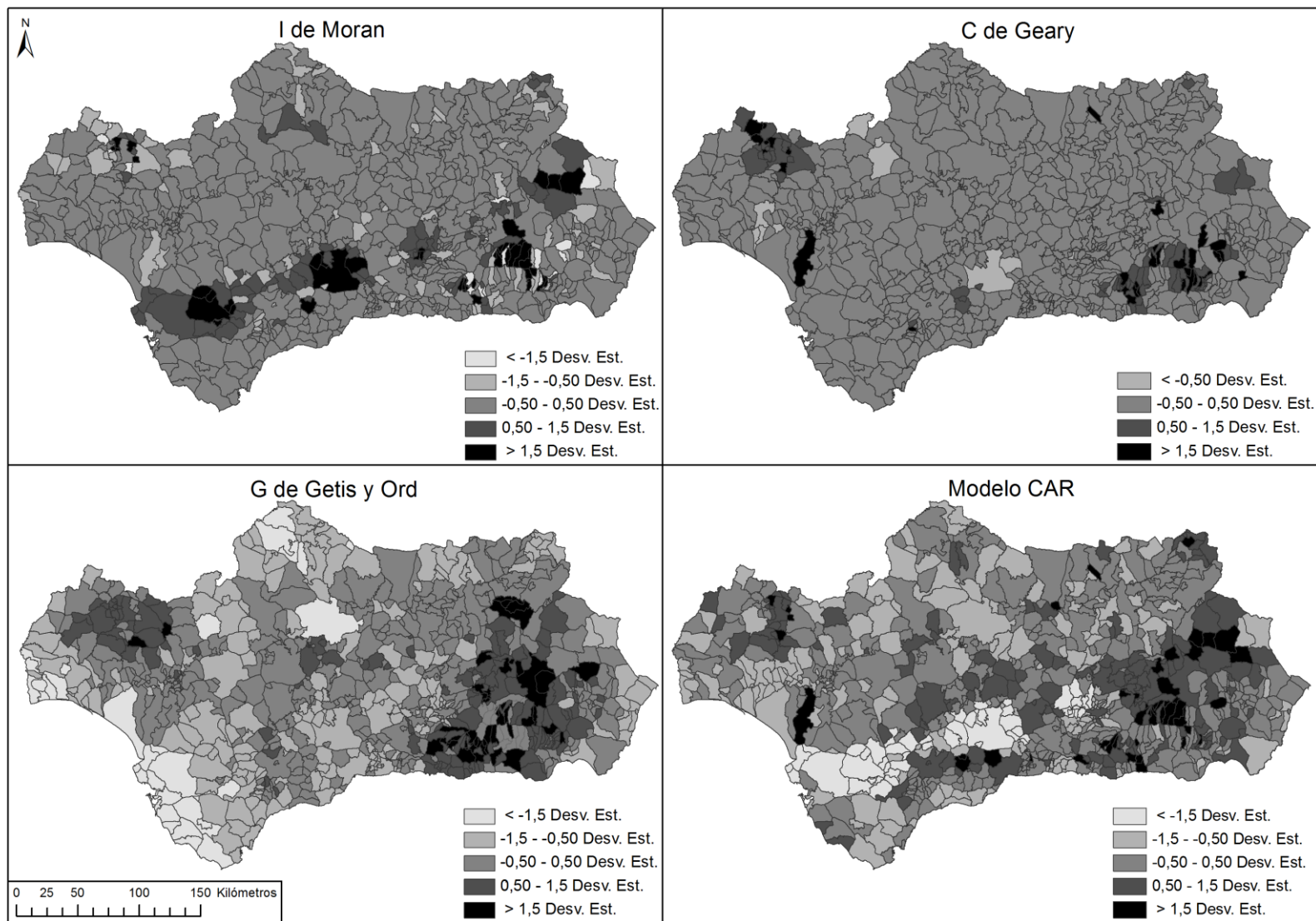
Fuente: Elaboración propia a partir del SISMA, 2006

Mapa 6: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la esquizofrenia en los municipios de Andalucía (2007)



Fuente: Elaboración propia a partir del SISMA, 2007

Mapa 7: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la esquizofrenia en los municipios de Andalucía (2008)



Fuente: Elaboración propia a partir del SISMA, 2008



### 4.2.1.3. Hot-spots de prevalencia administrativa de la esquizofrenia

#### Descripción estadística

La selección de los *hot-spots* finales es el resultado de una solución de consenso entre los resultados de los tres indicadores locales de asociación espacial y el modelo CAR Bayesiano. La aplicación del AEMO/AE ha permitido combinar los cuatro métodos para obtener *hot-spots* formados por municipios próximos y con una elevada dependencia espacial en los valores de prevalencia administrativa de la esquizofrenia. Como se explicó en la metodología, se han aplicado cuatro funciones de aptitud diferentes para evaluar las soluciones dadas por el Algoritmo en cada generación: *Fine-grained strength Pareto (SPEA2)*, *Weighted sum of objectives (WO)*, *Standard Rankings Selection (SRS)* y *Fuzzy evaluation of weighted objectives (FI)*. Los resultados de las *funciones de aptitud* y los *hot-spot finales* han sido cartografiados mediante el SIG.

En la Tabla 14, Tabla 15, Tabla 16 y Tabla 17 se muestran las estadísticas descriptivas de los *hot-spots* referidas a los valores de los índices de dependencia espacial. También se muestran las estadísticas de la prevalencia de la esquizofrenia en la Tabla 18. Los histogramas de frecuencia de cada grupo de soluciones se representan en la Figura 17.

Tabla 14: Estadísticas descriptivas de los *hot-spots* de esquizofrenia en 2004  
(LISA y CAR Bayesiano)

	<i>I</i> de Moran	<i>C</i> de Geary	<i>G</i> * de Getis	CAR
Media	1,04	11,28	0,09	1,52
Desviación estándar	1,59	11,85	0,03	0,63
Mediana	0,17	8,57	0,08	1,24
Mínimo	-0,49	0,32	0,06	0,68
Máximo	4,89	37,61	0,15	2,81

Fuente: Elaboración propia

Tabla 15: Estadísticas descriptivas de los *hot-spots* de esquizofrenia en 2006  
(LISA y CAR Bayesiano)

	<i>I</i> de Moran	<i>C</i> de Geary	<i>G</i> * de Getis	CAR
Media	1,21	18,10	0,10	1,70
Desviación estándar	1,19	11,36	0,02	0,45
Mediana	1,05	17,11	0,10	1,79
Mínimo	-0,43	3,69	0,08	1,04
Máximo	3,30	40,17	0,13	2,30

Fuente: Elaboración propia

Tabla 16: Estadísticas descriptivas de los *hot-spots* de esquizofrenia en 2007  
(LISA y CAR Bayesiano)

	<i>I</i> de Moran	<i>C</i> de Geary	<i>G</i> * de Getis	CAR
Media	1,76	24,11	0,98	1,80
Desviación estándar	2,09	16,18	0,02	0,63
Mediana	1,17	22,11	0,10	1,81
Mínimo	-0,61	5,77	0,07	0,95
Máximo	5,53	61,53	0,13	2,92

Fuente: Elaboración propia

Tabla 17: Estadísticas descriptivas de los *hot-spots* de esquizofrenia en 2008  
(LISA y CAR Bayesiano)

	<i>I</i> de Moran	<i>C</i> de Geary	<i>G</i> * de Getis	CAR
Media	0,96	14,62	0,09	1,56
Desviación estándar	1,80	15,59	0,02	0,74
Mediana	0,10	11,99	0,08	1,25
Mínimo	-0,43	0,35	0,04	0,89
Máximo	7,41	61,22	0,14	3,73

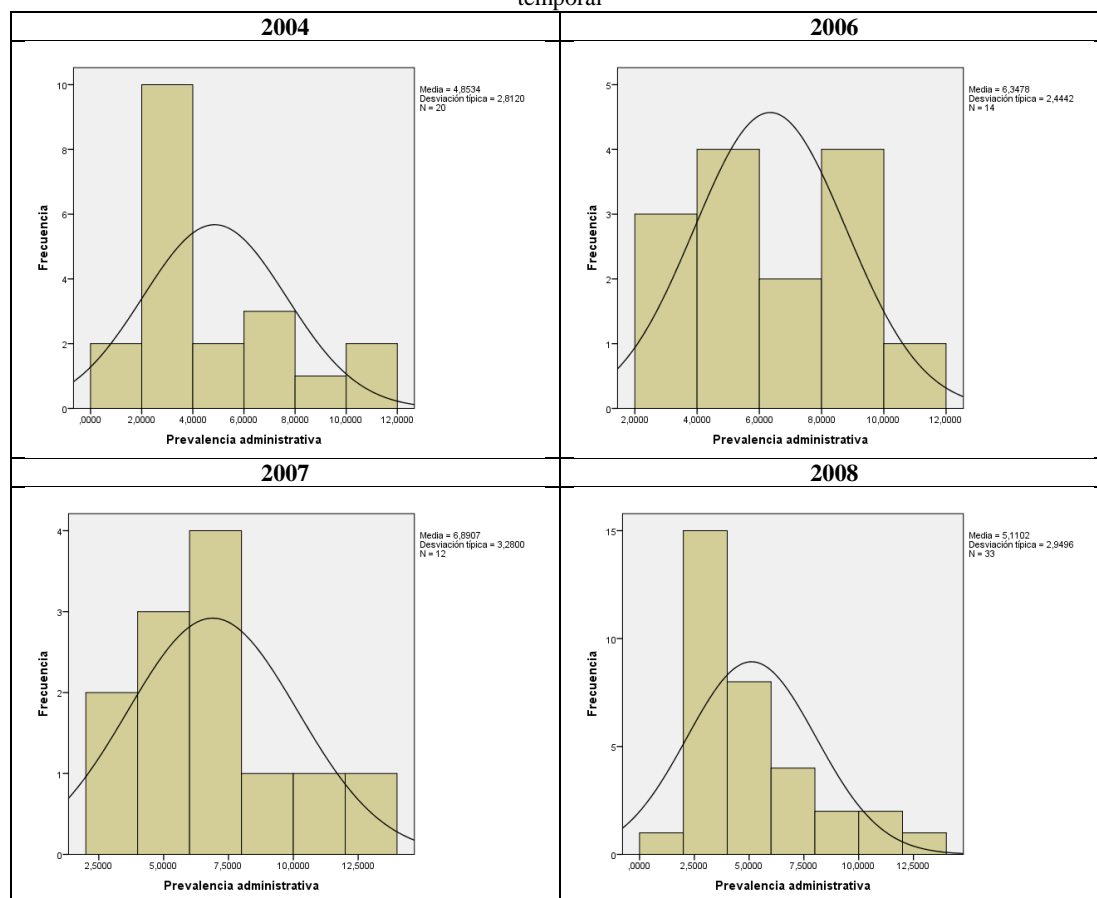
Fuente: Elaboración propia

Tabla 18: Estadísticas descriptivas de los *hot-spots* de esquizofrenia  
(prevalencia administrativa de la esquizofrenia)

	2004	2006	2007	2008
Recuento	20	14	12	33
Media	4,85	6,35	6,89	5,11
Desviación estándar	2,81	2,44	3,28	2,95
Mediana	3,88	6,57	7,47	4,16
Mínimo	1,15	2,26	2,18	1,52
Máximo	10,66	10,77	12,91	12,61

Fuente: Elaboración propia

Figura 17: Histogramas de frecuencia de la prevalencia administrativa de la esquizofrenia de los *hot-spots* en la serie temporal



Fuente: SISMA

Las unidades seleccionadas como *hot-spots* han sido agrupadas por vecindad y numeradas para facilitar su seguimiento en el tiempo. A cada unidad espacial se le ha asignado un grupo numerado que se mantiene durante los cuatro años denominado con la abreviatura HSe (*hot-spot* esquizofrenia) y números consecutivos. En la Tabla 19 se muestran los municipios pertenecientes a cada grupo de *hot-spots* junto a la provincia y a la Unidad de Salud Mental Comunitaria a la que pertenece. También se puede observar la persistencia temporal de los mismos. Los *hot-spots* han aparecido en 13 de las 75 áreas de salud mental andaluzas.

Tabla 19: Municipios que forman parte de los hot-spots de esquizofrenia durante la serie temporal

HSe	Municipio	2004	2006	2007	2008	Prov.	Área SM
HSe1	Abla	x	x	x	x	AL	Almería
HSe1	Abrucena	x	x	x	x	AL	Almería
HSe1	Alboloduy	x			x	AL	Almería
HSe1	Alhabia			x		AL	Almería

HSe	Municipio	2004	2006	2007	2008	Prov.	Área SM
HSe1	Alhama de Almería	x			x	AL	Almería
HSe1	Alicún	x				AL	Almería
HSe1	Fiñana	x	x	x	x	AL	Almería
HSe1	Ohanes				x	AL	Almería
HSe1	Santa Fe de Modújar				x	AL	Almería
HSe1	Alcolea	x	x	x	x	AL	El Ejido
HSe1	Bayárcal		x	x	x	AL	El Ejido
HSe1	Berja				x	AL	El Ejido
HSe1	Dalías				x	AL	El Ejido
HSe1	El Ejido				x	AL	El Ejido
HSe1	Fondón				x	AL	El Ejido
HSe1	Paterna del Río		x	x	x	AL	El Ejido
HSe1	Albuñán		x	x	x	GR	Guadix
HSe1	Aldeire	x	x	x	x	GR	Guadix
HSe1	Alquife		x	x		GR	Guadix
HSe1	Cogollos de Guadix	x		x		GR	Guadix
HSe1	Diezma		x			GR	Guadix
HSe1	Dólar				x	GR	Guadix
HSe1	Huéneja	x	x	x	x	GR	Guadix
HSe1	Jerez del Marquesado		x			GR	Guadix
HSe1	La Calahorra			x	x	GR	Guadix
HSe1	Lanteira	x	x			GR	Guadix
HSe1	Valle del Zalabí	x	x		x	GR	Guadix
HSe1	La Mojonera				x	AL	Roquetas
HSe2	Albánchez				x	AL	Albox
HSe2	Albox	x			x	AL	Albox
HSe2	Arboleas				x	AL	Albox
HSe2	Bédar				x	AL	Huércal-Overa
HSe3	Alcalá del Valle	x				CA	Villamartín
HSe3	Bornos	x				CA	Villamartín
HSe3	Espera	x				CA	Villamartín
HSe3	Olvera	x				CA	Villamartín
HSe3	Prado del Rey	x				CA	Villamartín
HSe3	Puerto Serrano	x				CA	Villamartín
HSe3	Villamartín	x				CA	Villamartín
HSe4	La Algaba				x	SE	Macarena Norte
HSe5	Palma del Río				x	CO	Palma del Río
HSe6	Larva				x	JA	Úbeda
HSe7	Gobernador				x	GR	Cartuja
HSe7	Píñar				x	GR	Cartuja
HSe8	Villamena				x	GR	Motril
HSe8	Otívar				x	GR	Zaidín

Fuente: Elaboración propia

## Comparación de las distribuciones y medianas

Se ha aplicado la Prueba  $U$  de Mann-Whitney y la Prueba de medianas (Tabla 20), en ambos casos para muestras independientes, a los valores de los índices de dependencia espacial y a la prevalencia administrativa. La finalidad es comparar las distribuciones y las medianas entre los diferentes grupos de *hot-spots* y el resto *unidades espaciales*.

En el caso de los índices, se observa que las distribuciones entre los *hot-spots* y el resto no son iguales en ningún año. En la comparación de las medianas se aprecian distintas situaciones aunque mayoritariamente se rechaza la hipótesis nula de igualdad excepto para la  $I$  de Moran en los cuatro años.

En cuanto a la prevalencia administrativa se verifica que las distribuciones y medianas son diferentes entre el grupo de *unidades espaciales* identificadas como *hot-spots* de dependencia espacial de esquizofrenia y el grupo de aquellas que no lo son (significación del 0,05). Es decir, la prevalencia tiene un comportamiento desigual dentro de los *hot-spots* y fuera de ellos.

Tabla 20: Comparación de las distribuciones y medianas de las unidades espaciales que forman los diferentes *hot-spots* de esquizofrenia y aquellas que no forman parte de ninguno

	<b>Prueba</b>	<b>Hipótesis nula</b>	<b>2004</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>
Dependencia espacial	Prueba Kruskal-Wallis de muestras independientes	H <sub>0</sub> : la distribución de la prevalencia es la misma en las unidades espaciales identificadas como hot-spots (HSe1, HSe2, HSe3, HSe4, HSe5, HSe6, HSe7 y HSe8) y las que no lo han sido.	<i>I</i> <b>Rechazar</b> (sig. 0,003 < 0,05)	<i>I</i> <b>Rechazar</b> (sig. 0,002 < 0,05)	<i>I</i> <b>Rechazar</b> (sig. 0,017 < 0,05)	<i>I</i> <b>Rechazar</b> (sig. 0,019 < 0,05)
			<i>C</i> <b>Rechazar</b> (sig. 0,001 < 0,05)	<i>C</i> <b>Rechazar</b> (sig. 0,001 < 0,05)	<i>C</i> <b>Rechazar</b> (sig. 0,00 < 0,05)	<i>C</i> <b>Rechazar</b> (sig. 0,00 < 0,05)
			<i>G</i> * <b>Rechazar</b> (sig. 0,00 < 0,05)	<i>G</i> * <b>Rechazar</b> (sig. 0,00 < 0,05)	<i>G</i> * <b>Rechazar</b> (sig. 0,00 < 0,05)	<i>G</i> * <b>Rechazar</b> (sig. 0,00 < 0,05)
			<i>CAR</i> <b>Rechazar</b> (sig. 0,00 < 0,05)	<i>CAR</i> <b>Rechazar</b> (sig. 0,00 < 0,05)	<i>CAR</i> <b>Rechazar</b> (sig. 0,00 < 0,05)	<i>CAR</i> <b>Rechazar</b> (sig. 0,00 < 0,05)
	Prueba de medianas de muestras independientes	H <sub>0</sub> : la mediana de la distribución de la prevalencia es la misma en las unidades espaciales identificadas como	<i>I</i> Retener (sig. 0,156 > 0,05)	<i>I</i> Retener (sig. 0,144 > 0,05)	<i>I</i> Retener (sig. 0,380 > 0,05)	<i>I</i> Retener (sig. 0,354 > 0,05)

	<b>Prueba</b>	<b>Hipótesis nula</b>	<b>2004</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>
		hot-spots (HSe1, HSe2, HSe3, HSe4, HSe5, HSe6, HSe7 y HSe8) y las que no lo han sido.	<i>C</i> <b>Rechazar</b> (sig. 0,011 < 0,05)	<i>C</i> <b>Rechazar</b> (sig. 0,022 < 0,05)	<i>C</i> <b>Rechazar</b> (sig. 0,001 < 0,05)	<i>C</i> <b>Rechazar</b> (sig. 0,006 < 0,05)
			<i>G*</i> <b>Rechazar</b> (sig. 0,00 < 0,05)	<i>G*</i> <b>Rechazar</b> (sig. 0,001 < 0,05)	<i>G*</i> <b>Rechazar</b> (sig. 0,001 < 0,05)	<i>G*</i> <b>Rechazar</b> (sig. 0,00 < 0,05)
			<i>CAR</i> <b>Rechazar</b> (sig. 0,00 < 0,05)	<i>CAR</i> <b>Rechazar</b> (sig. 0,001 < 0,05)	<i>CAR</i> <b>Rechazar</b> (sig. 0,001 < 0,05)	<i>CAR</i> <b>Rechazar</b> (sig. 0,00 < 0,05)
Prevalencia administrativa	Prueba Kruskal-Wallis de muestras independientes	H <sub>0</sub> : la distribución de la prevalencia es la misma en las unidades espaciales identificadas como hot-spots (HSe1, HSe2, HSe3, HSe4, HSe5, HSe6, HSe7 y HSe8) y las que no lo han sido.	<b>Rechazar</b> (sig. 0,00 < 0,05)	<b>Rechazar</b> (sig. 0,00 < 0,05)	<b>Rechazar</b> (sig. 0,00 < 0,05)	<b>Rechazar</b> (sig. 0,00 < 0,05)
	Prueba de medianas de muestras independientes	H <sub>0</sub> : la mediana de la distribución de la prevalencia es la misma en las unidades espaciales identificadas como hot-spots (HSe1, HSe2, HSe3, HSe4, HSe5, HSe6, HSe7 y HSe8) y las que no lo han sido.	<b>Rechazar</b> (sig. 0,006 < 0,05)	<b>Rechazar</b> (sig. 0,022 < 0,05)	<b>Rechazar</b> (sig. 0,041 < 0,05)	<b>Rechazar</b> (sig. 0,00 < 0,05)

Fuente: elaboración propia

## Descripción de los *hot-spots*

### *Año 2004*

El *hot-spot* principal (HSe1) de 2004 en cuanto a superficie y acuerdo entre las *funciones de aptitud* ha sido localizado en una zona continua que abarca las áreas de salud mental de Guadix (Granada), el Ejido y Almería (Almería) (Mapa 8 y Mapa 9). En la zona oriental de Andalucía también aparece una *unidad espacial* (HSe2) identificada como *hot-spot* de segundo nivel en el área de Albox. Por último, existe un tercer *hot-spot* (HSe3) de segundo nivel que recoge varias unidades espaciales en la Sierra de Cádiz correspondientes al área de Villamartín.

#### *Año 2006*

El HSe1 vuelve a aparecer en el área limítrofe entre la comarca de Guadix (Granada) y la comarca Los Filabres-Tabernas atendidas por los USMC de el Ejido, Guadix y Almería (Almería) (Mapa 10 y Mapa 11). Junto a éstos aparecen unidades espaciales de segundo nivel que refuerzan los altos valores de dependencia encontrados por el AEMO/AE en este ámbito. Los *hot-spot* del área de Villamartín en la Sierra de Cádiz (HSe3) y de Albox en el Valle del Almanzora (HSe2) no aparecen en 2006.

#### *Año 2007*

En 2007, el HSe1 no solo se mantiene sino que se intensifica y ahora es de primer nivel en toda su extensión (Mapa 12 y Mapa 13). Estas unidades espaciales son de primer nivel porque han sido identificadas como *hot-spot* por al menos el 50% de los métodos de selección aplicados a los resultados de las cuatro *funciones de aptitud*. El AEMO/AE no identifica a los otros dos *hot-spots* de 2004.

#### *Año 2008*

En el año 2008 aparecen nuevos ámbitos aunque son pequeños y atañen a *unidades espaciales* aisladas (Mapa 14 y Mapa 15). Las nuevas unidades que abarca son *hot-spots* de segundo nivel y, por tanto, presentan un menor consenso entre *funciones de aptitud*. El HSe1 no solo sigue siendo el más potente sino que se extiende a comarcas vecinas y alcanza la costa a través del área de salud mental de el Ejido. El HSe2 del área de Albox que aparecía en 2004 también aparece en 2008 aunque con mayor extensión. Nuevos *hot-spots* son identificados en las áreas de Motril (HSe8) y Granada Norte (HSe7) en Granada, Úbeda en Jaén (HSe6), la Palma del Río en Córdoba (HSe5) y Macarena Norte en Sevilla (HSe4). En 2008 tampoco aparece el HSe3 de Villamartín que apareció previamente en 2004.

#### *Comentario de la serie*

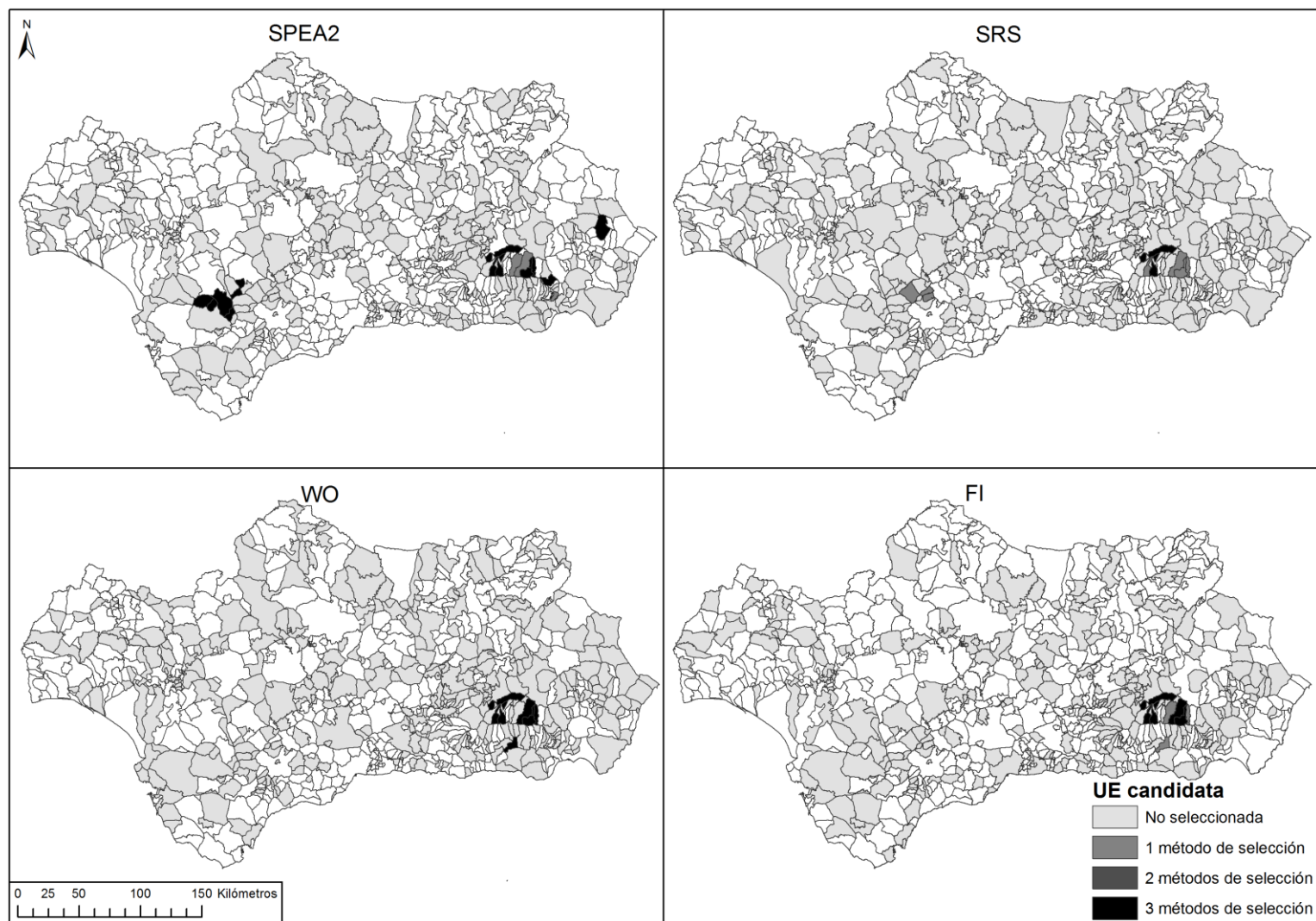
El *hot-spot* de esquizofrenia (HSe1) se localizan de forma persistente a lo largo de los cuatro años analizados en el área limítrofe entre la cara norte de Sierra Nevada, que incluye las comarcas de Guadix (Granada) y Los Filabres-Tabernas (Almería), y la cara sur que constituye La Alpujarra almeriense (Tabla 19). En 2004, 2006 y 2008 también aparece un *hot-spot* (HSe2) en el Valle del Almanzora en Almería. El resto de *hot-spots* aparecen en ciertos años como el de

la Sierra de Cádiz (HSe3) en 2004 y los de las comarcas de Granada, Jaén, Córdoba y Sevilla en 2008 (HSe4, HSe5, HSe6 y HSe8).

Si se comparan los resultados del AEMO/AE con los obtenidos por los métodos locales de autocorrelación clásicos y el modelo Bayesiano destaca la exclusión como *hot-spot* de la Sierra de Aracena, que presenta una clara dependencia espacial en varios índices pero como demuestra el AEMO no de forma consensuada. En cambio el *hot-spot* principal del eje Sierra Nevada-La Alpujarra (HSe1) si es identificado por los métodos de análisis de la dependencia espacial.

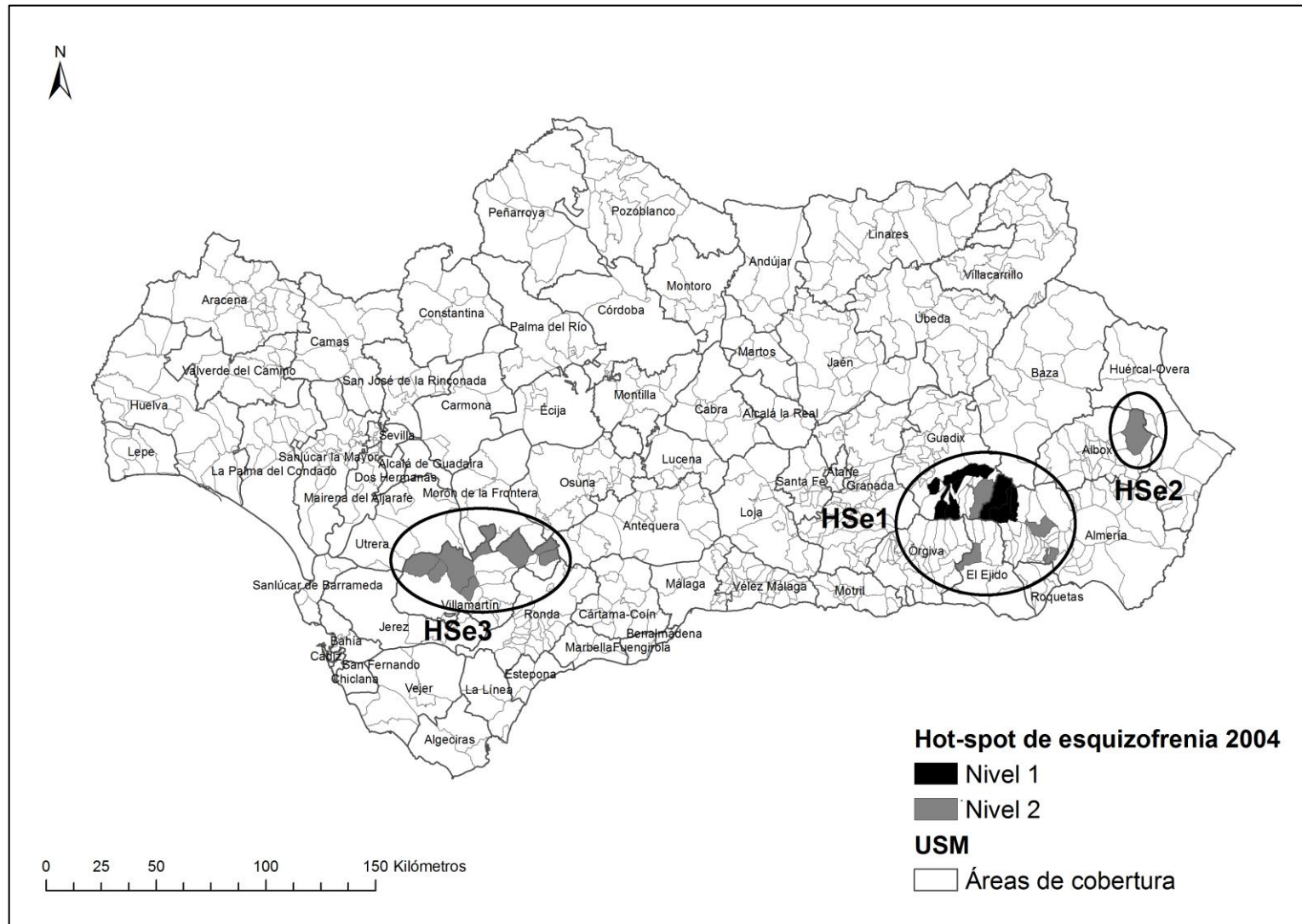


Mapa 8: Resultados del AEMO/AE por *función de aptitud* para los *hot-spot* de esquizofrenia en Andalucía (2004).  
Unidades espaciales (UE) candidatas y número de métodos de selección que la han identificado como *hot-spot* potencial



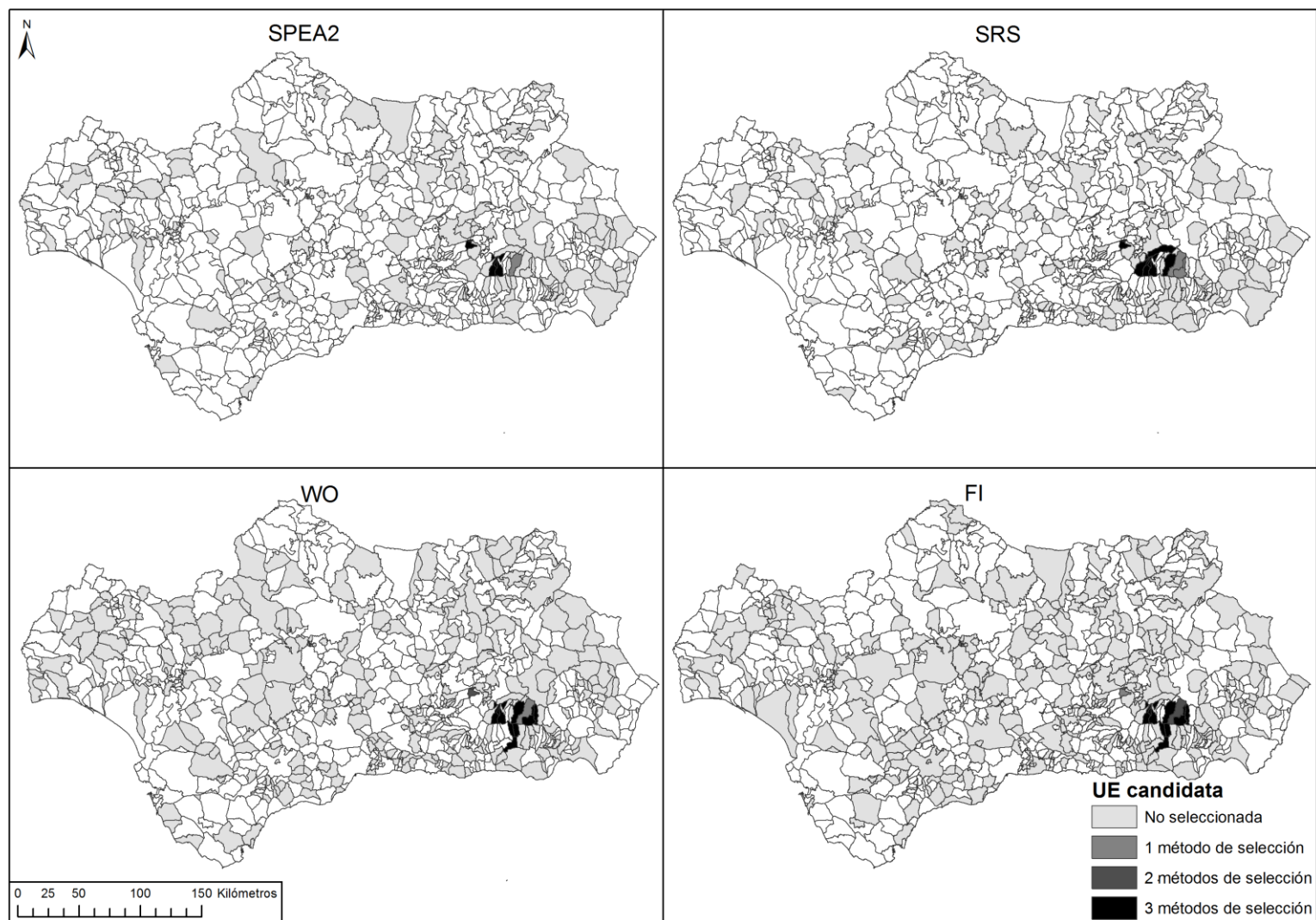
Fuente: Elaboración propia a partir del SISMA, 2004

Mapa 9: Hot-spots de dependencia espacial de la esquizofrenia en Andalucía (2004)



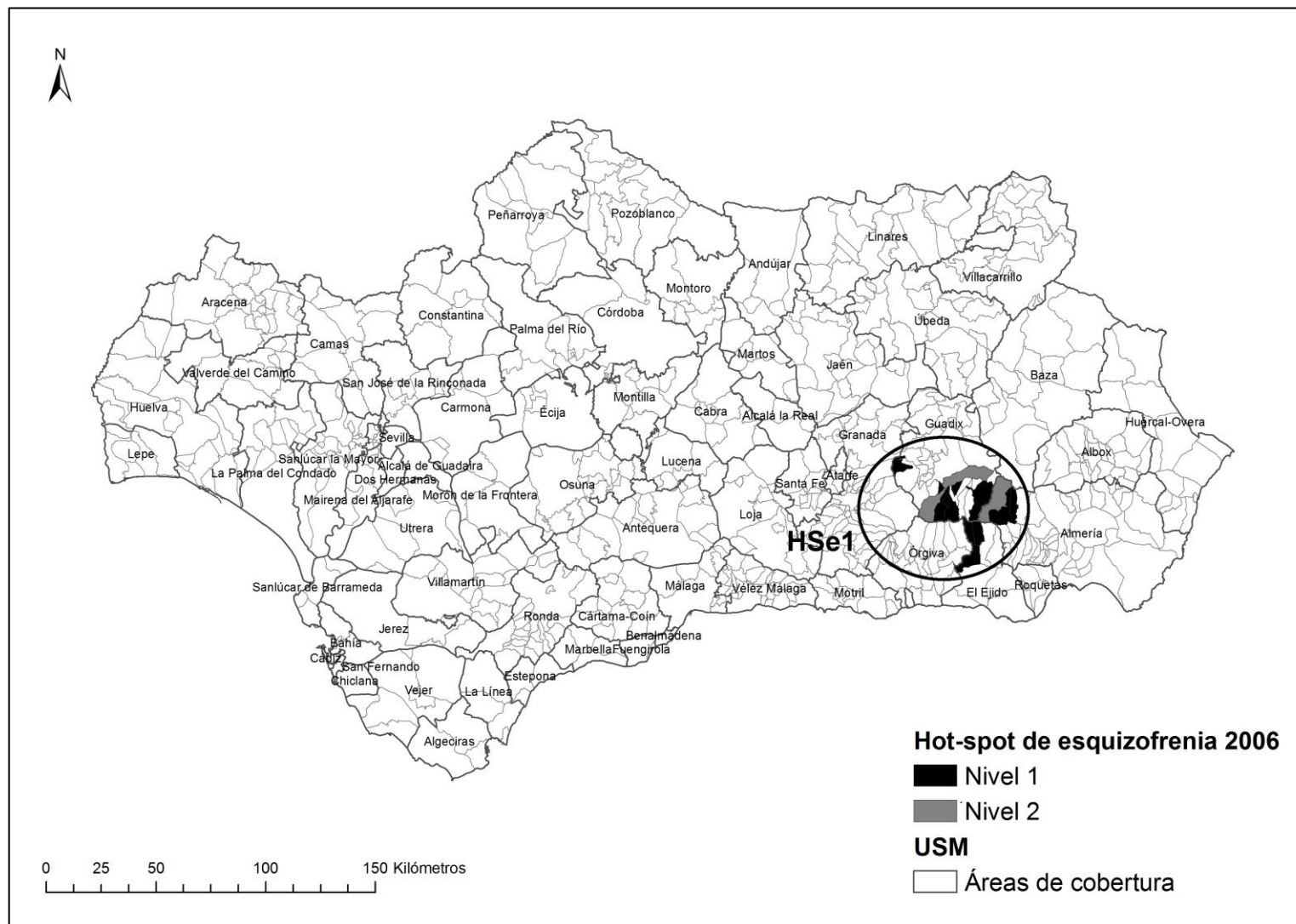
Fuente: Elaboración propia a partir del SISMA, 2004

Mapa 10: Resultados del AEMO/AE por *función de aptitud* para los *hot-spot* de esquizofrenia en Andalucía (2006)  
 Unidades espaciales (*UE*) candidatas y número de métodos de selección que la han identificado como *hot-spot* potencial



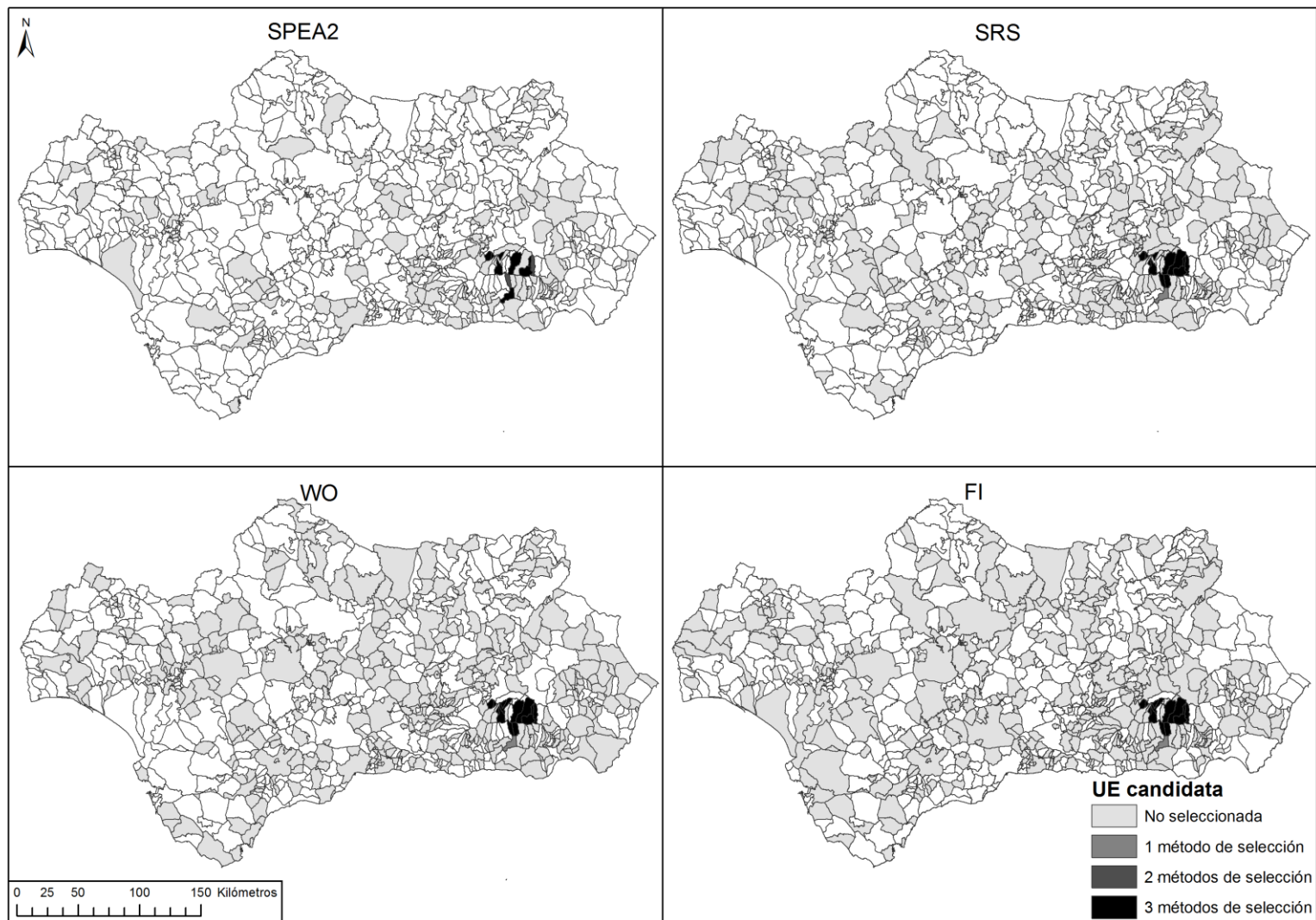
Fuente: Elaboración propia a partir del SISMA, 2006

Mapa 11: *Hot-spots* de dependencia espacial de la esquizofrenia en Andalucía (2006)



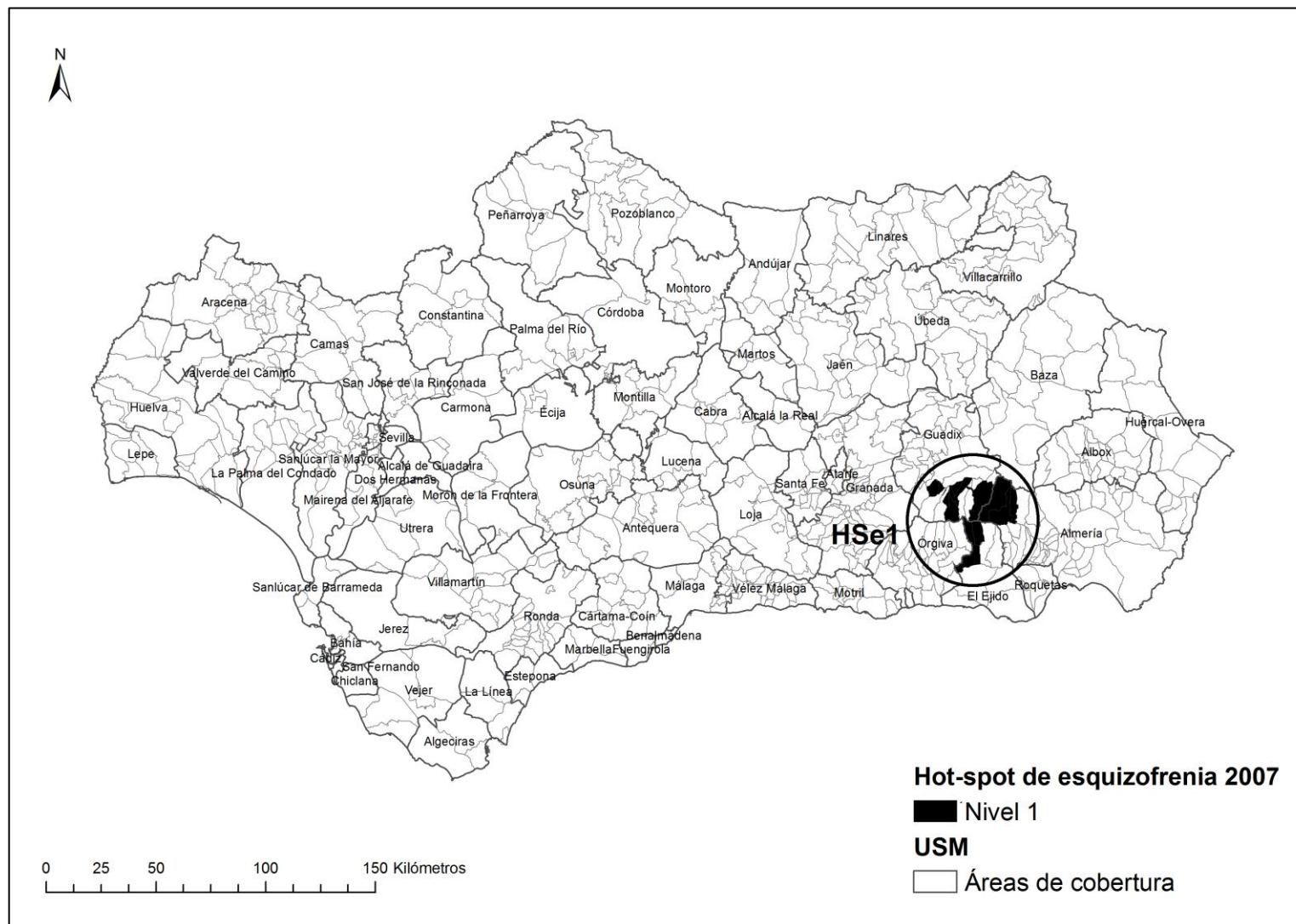
Fuente: Elaboración propia a partir del SISMA, 2006

Mapa 12: Resultados del AEMO/AE por *función de aptitud* para los *hot-spot* de esquizofrenia en Andalucía (2007)  
 Unidades espaciales (UE) candidatas y número de métodos de selección que la han identificado como *hot-spot* potencial



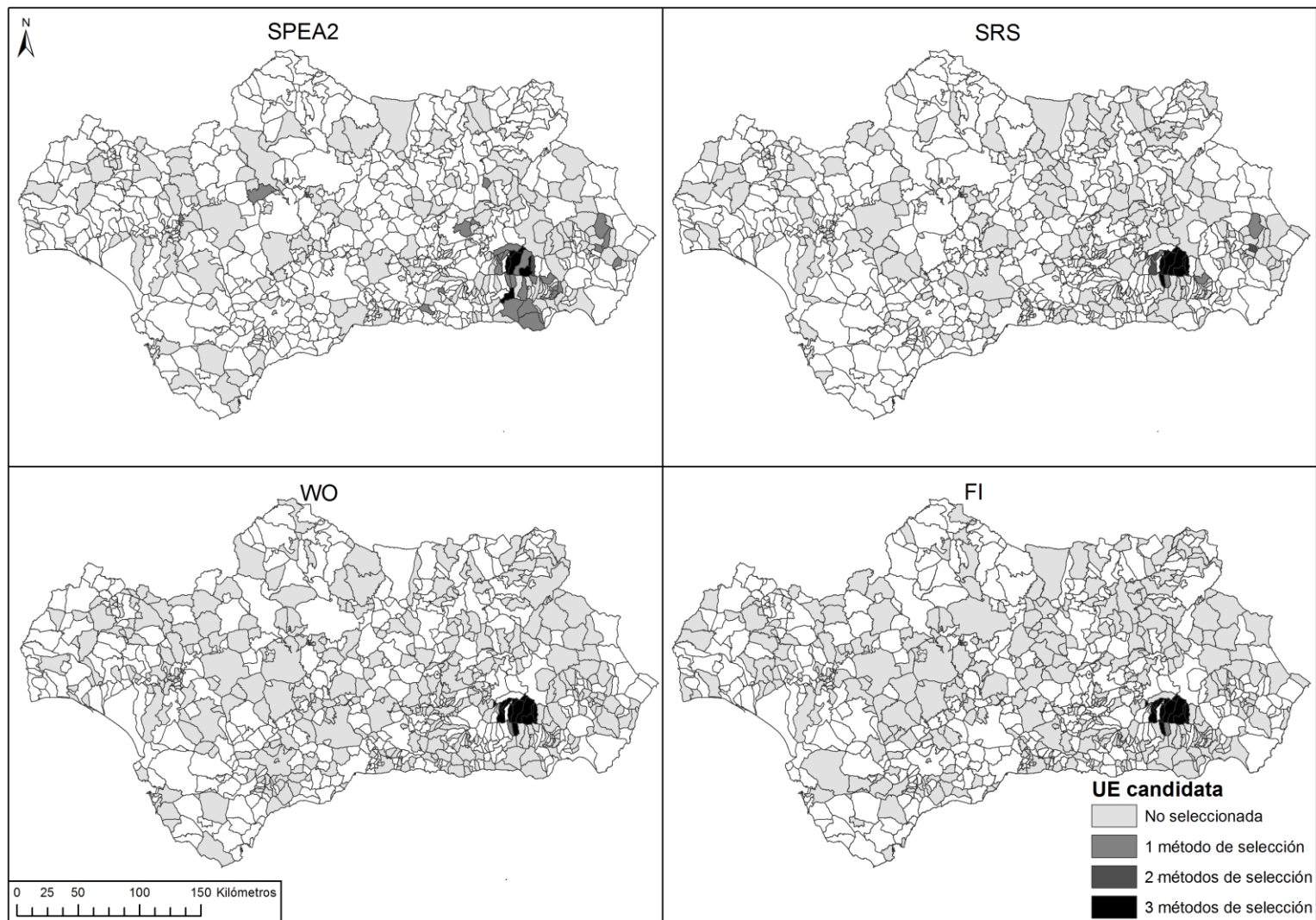
Fuente: Elaboración propia a partir del SISMA, 2007

Mapa 13: *Hot-spots* de dependencia espacial de la esquizofrenia en Andalucía (2007)



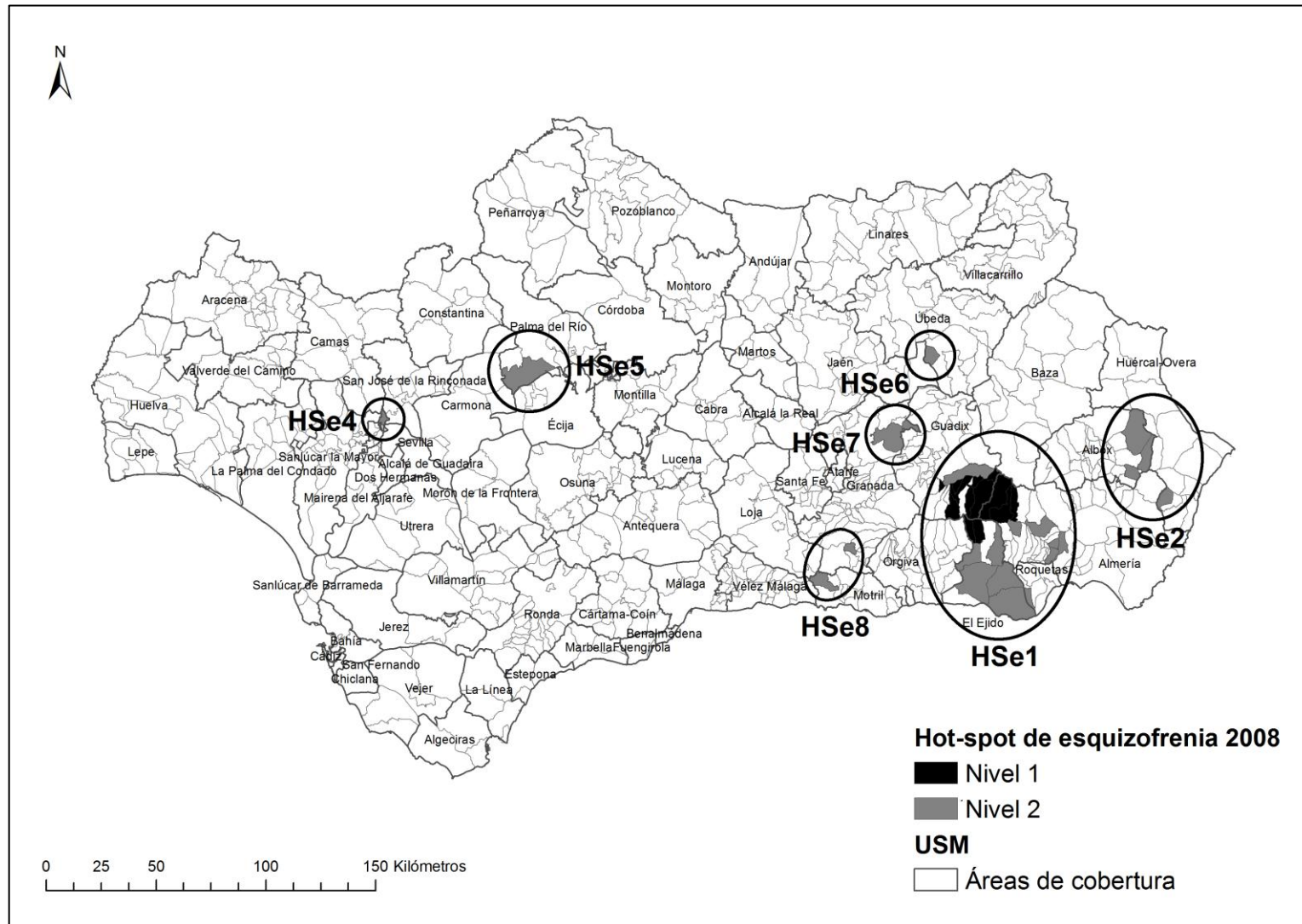
Fuente: Elaboración propia a partir del SISMA, 2007

Mapa 14: Resultados del AEMO/AE por *función de aptitud* para los *hot-spot* de esquizofrenia en Andalucía (2008)  
Unidades espaciales (*UE*) candidatas y número de métodos de selección que la han identificado como *hot-spot* potencial



Fuente: Elaboración propia a partir del SISMA, 2008

Mapa 15: Hot-spots de dependencia espacial de la esquizofrenia en Andalucía (2008)



Fuente: Elaboración propia a partir del SISMA, 2008



#### 4.2.2. Caso 2: *Hot-spots* de prevalencia asistida de la depresión en Andalucía durante una serie temporal de cuatro años

##### 4.2.2.1. Distribución geográfica de la prevalencia administrativa de la depresión

En la Tabla 21 se muestran las estadísticas descriptivas de la prevalencia administrativa de la depresión durante la serie temporal. Se observa que, al igual que ocurre con la esquizofrenia, la depresión, tiene un comportamiento similar en los cuatro años con una variación insignificante de la media entre 3,1 y 3,3 casos por 1.000 habitantes, y de la desviación estándar entre 2,6 y 2,8 por 1.000. El valor mínimo es 0 y corresponde a municipios atendidos por USMC que no informaron al SISMA o a municipios muy pequeños que no presentaron casos. El número de municipios con valor 0 de prevalencia se mueve entre 83 y 104 a lo largo de la serie temporal. Por su parte, la variación entre los valores máximos de prevalencia para cada año oscila entre 19,2 y 25,0 casos por 1.000 habitantes.

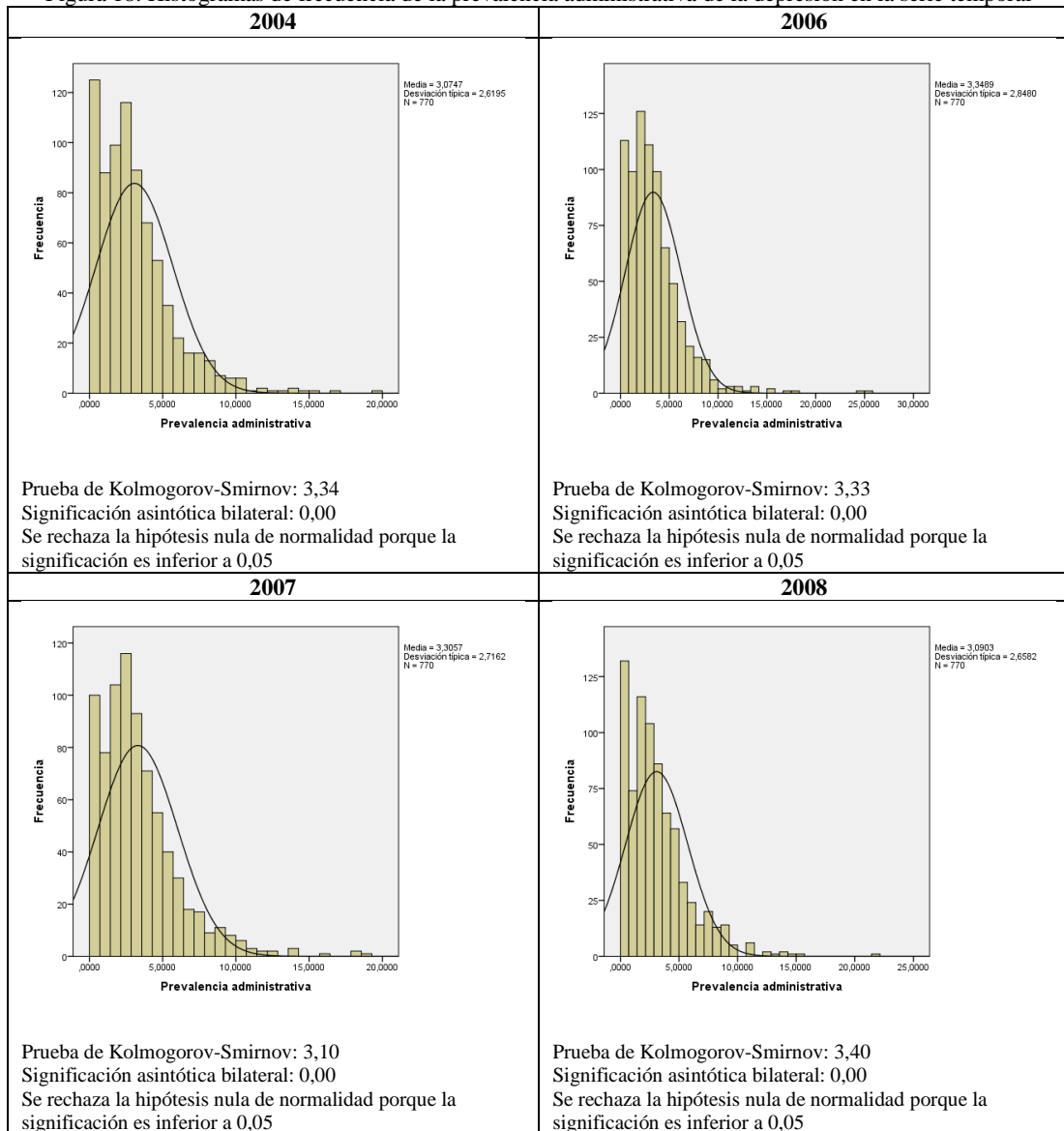
Tabla 21: Estadísticas descriptivas de la prevalencia administrativa de la depresión en Andalucía

	2004	2006	2007	2008
Recuento	770	770	770	770
Media	3,07	3,35	3,31	3,09
Desviación estándar	2,61	2,85	2,72	2,66
Mediana	2,53	2,85	2,77	2,50
Mínimo	0 (95 UE)	0 (85 UE)	0 (83 UE)	0 (104 UE)
Máximo	19,32	25,00	19,16	21,98

Fuente: SISMA

La Figura 18 contiene los histogramas de frecuencias de las poblaciones de prevalencia administrativa de la depresión para los cuatro años. El test de Kolmogorov-Smirnov muestra que su distribución estadística no se ajusta a una distribución normal, rechazándose la hipótesis nula, con una significación por debajo de 0,05.

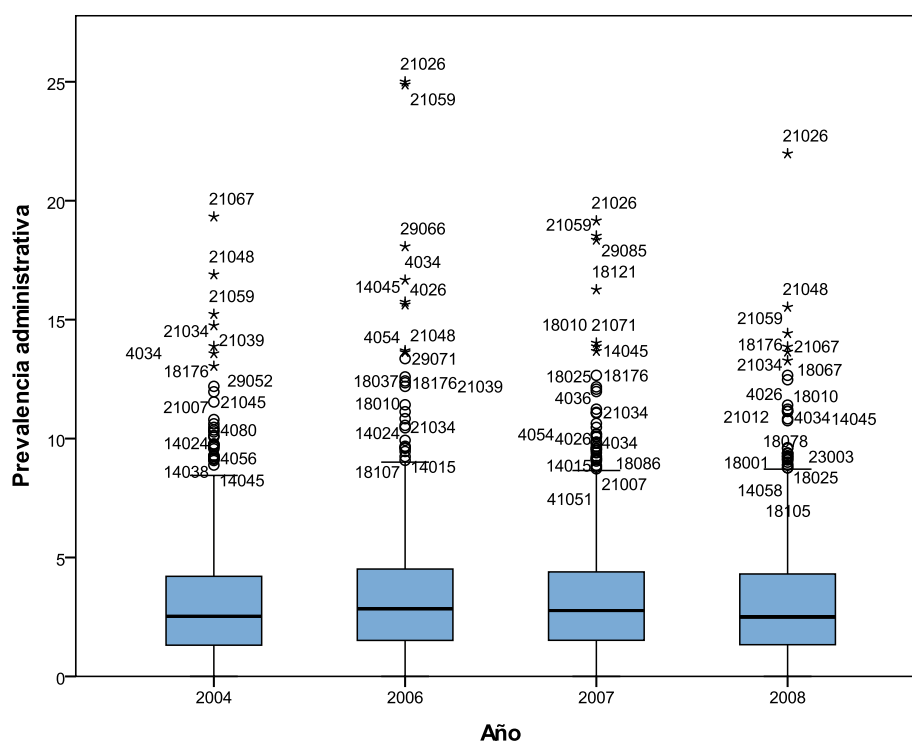
Figura 18: Histogramas de frecuencia de la prevalencia administrativa de la depresión en la serie temporal



Fuente: SISMA

Los *box-plot* de las series de prevalencia administrativa de la depresión muestran el grado de dispersión de la misma. Se observa la semejanza en el comportamiento de las poblaciones a través de la serie temporal. Existen numerosos casos atípicos con valores que se encuentran por encima de 1,5 veces el rango intercuartílico y casos extremos por encima de 3 veces el mismo.

Figura 19: *Box-plot* de las poblaciones de prevalencia administrativa de la depresión en la serie temporal (los números identificativos de los outliers corresponden al código INE municipal)



Fuente: SISMA

La prevalencia administrativa de la depresión para la serie temporal está representada en el Mapa 16. Los municipios con un valor de la prevalencia de la depresión significativamente alto, aquellos superiores a la media en 2,5 veces la desviación estándar, varían entre 14 y 20. Los municipios con valores muy elevados para la serie temporal completa fueron: Galaroza y Puerto Moral en Aracena (Huelva), Córdar en el Valle de Almanzora (Almería) y Soportújar en la Alpujarra granadina. La distribución territorial de los municipios con valores bajos, por debajo de -0,5 veces la desviación estándar y los valores 0, tienen que ser interpretados considerando las USMC que no informaron al SISMA. Así, al igual que con la esquizofrenia, encontramos valores bajos en las provincias de Cádiz, Granada y Málaga que responden a estas USMC.

Tabla 22: Distribución de frecuencias de la prevalencia administrativa de la depresión en Andalucía

	2004	2006	2007	2008
-1,5 a -0,5 Desv. Est.	253	246	252	254
-0,5 a 0,5 Desv. Est.	343	345	344	333
0,5 a 1,5 Desv. Est.	113	127	115	117
1,5 a 2,5 Desv. Est.	41	35	40	52
>2,5 Desv. Est.	20	17	19	14

Fuente: SISMA

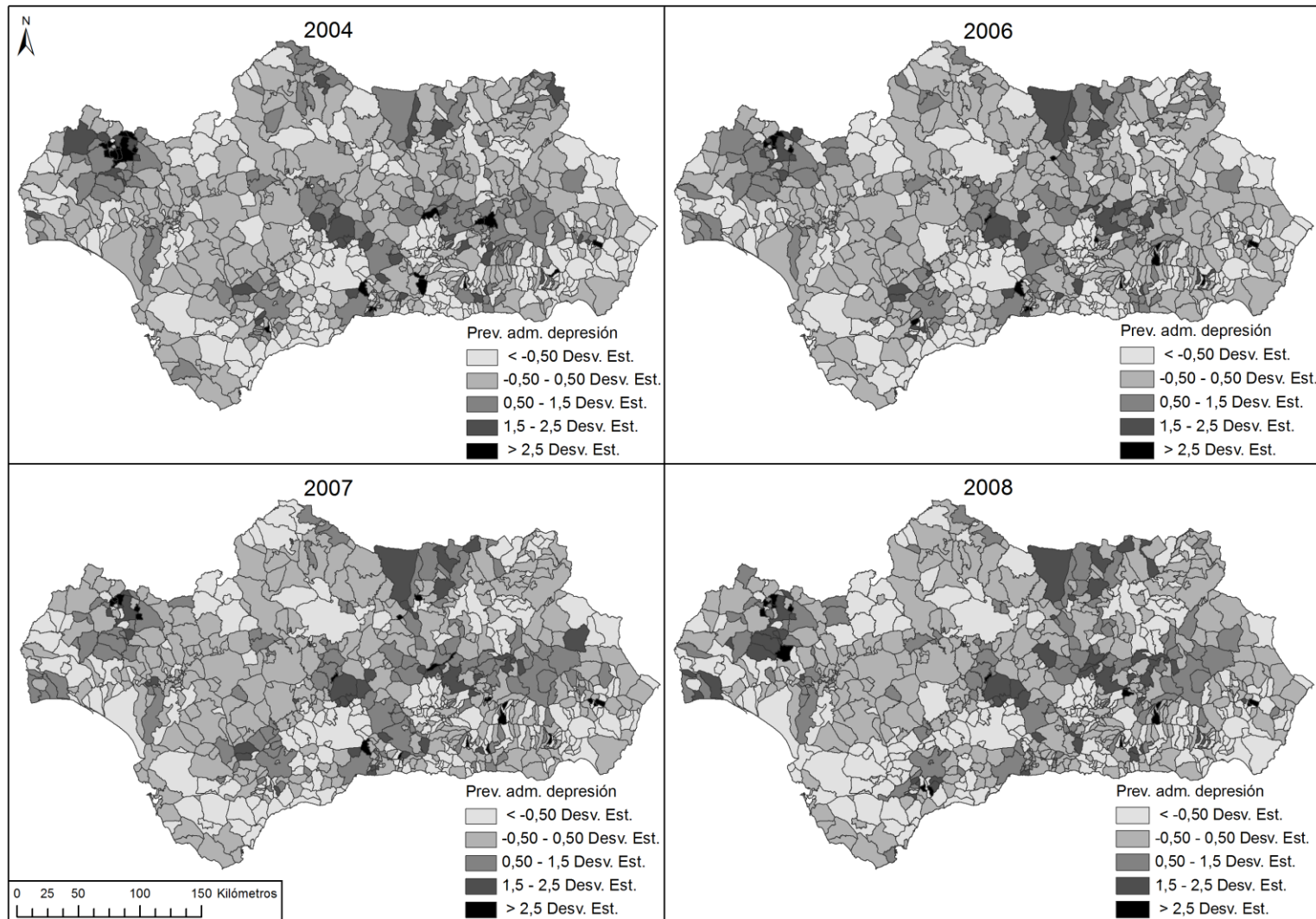
Los resultados de los índices globales de la *I* de Moran y la *G* de Getis y Ord descartan la distribución aleatoria de la prevalencia administrativa de la depresión en Andalucía ( $p = 0,01$ ). En los cuatro años el análisis espacial exploratorio ha detectado la existencia de agrupaciones espaciales de valores de prevalencia altos (Tabla 23).

Tabla 23: Análisis de la autocorrelación espacial global de la prevalencia asistida de la depresión en la serie temporal (Índices globales de la *I* de Moran y la *G* de Getis y Ord)

	2004	2006	2007	2008
<i>I</i> de Moran global	Existe agrupación espacial de la prevalencia  $I = 0,06$ Valor $z = 21,18$ d.e. Valor $p = 0,01$	Existe agrupación espacial de la prevalencia  $I = 0,04$ Valor $z = 14,78$ d.e. Valor $p = 0,01$	Existe agrupación espacial de la prevalencia  $I = 0,03$ Valor $z = 11,87$ d.e. Valor $p = 0,01$	Existe agrupación espacial de la prevalencia  $I = 0,04$ Valor $z = 14,95$ d.e. Valor $p = 0,01$
<i>G</i> de Getis y Ord global	Agrupaciones de valores altos de prevalencia  $G = 0$ Valor $z = 9,36$ d.e. Valor $p = 0,01$	Agrupaciones de valores altos de prevalencia  $G = 0$ Valor $z = 7,11$ d.e. Valor $p = 0,01$	Agrupaciones de valores altos de prevalencia  $G = 0$ Valor $z = 5,91$ d.e. Valor $p = 0,01$	Agrupaciones de valores altos de prevalencia  $G = 0$ Valor $z = 6,90$ d.e. Valor $p = 0,01$
d.e. = desviación estándar.				

Fuente: elaboración propia

Mapa 16: Prevalencia administrativa de la depresión en los municipios de Andalucía



Fuente: Elaboración propia a partir del SISMA (2004, 2006, 2007 y 2008)

#### 4.2.2.2. Dependencia espacial de la prevalencia administrativa de la depresión

Los resultados de los cuatro índices de análisis espacial aplicados sobre los valores de la prevalencia administrativa de la depresión se han representado en el Mapa 17, Mapa 18, Mapa 19 y Mapa 20. Y sus estadísticas básicas en la Tabla 24, Tabla 25, Tabla 26 y Tabla 27. Junto a estos datos también se presentan sus correspondientes gráficos *box-plot* (Figura 20, Figura 21, Figura 22 y Figura 23). Las estadísticas descriptivas y las representaciones gráficas de las mismas muestran una gran homogeneidad temporal sin grandes contrastes excepto por la aparición de *outliers*.

Los municipios que presentan dependencia espacial positiva de los valores de prevalencia se localizan fundamentalmente en la Sierra de Aracena de Huelva. La existencia de este clúster se podía intuir en los mapas que representaban las tasas de prevalencia pero ahora es cuando se ha justificado estadísticamente mediante las técnicas de análisis espacial.

Existen otras agrupaciones de municipios que también se caracterizan por sus prevalencia altas aunque en este caso no existe consenso entre los métodos, como la Subbética Cordobesa, el Valle Alto del Guadalquivir y el Valle de los Pedroches en Córdoba; el Andévalo y la Cuenca Minera onubense; las áreas de Guadix, Loja y Alhama en Granada; las áreas limítrofes de la Sierra de Cádiz y la Serranía de Ronda; Andújar y la Sierra de Segura en Jaén; y el Valle del Almanzora en Almería.

Tabla 24: Estadísticas descriptivas de la *I* de Moran  
(prevalencia administrativa de la depresión)

	2004	2006	2007	2008
Recuento	770	770	770	770
Media	0,32	0,22	0,21	0,28
Desviación estándar	1,2	0,76	0,60	0,79
Mediana	0,090	0,075	0,086	0,091
Mínimo	-2,4	-3,1	-2,3	-2,4
Máximo	14	9,7	6	8,7

Fuente: Elaboración propia

Tabla 25: Estadísticas descriptivas de la C de Geary  
(prevalencia administrativa de la depresión)

	2004	2006	2007	2008
Recuento	770	770	770	770
Media	9,5	13	12	10
Desviación estándar	14	27	20	18
Mediana	4,9	5,6	5,5	4,8
Mínimo	0,0034	0,028	0,011	0,0092
Máximo	140	390	240	270

Fuente: Elaboración propia

Tabla 26: Estadísticas descriptivas de la G\* de Getis & Ord  
(prevalencia administrativa de la depresión)

	2004	2006	2007	2008
Recuento	770	770	770	770
Media	0,059	0,060	0,060	0,060
Desviación estándar	0,022	0,020	0,020	0,022
Mediana	0,057	0,057	0,059	0,058
Mínimo	0,020	0,018	0,019	0,015
Máximo	0,15	0,14	0,12	0,14

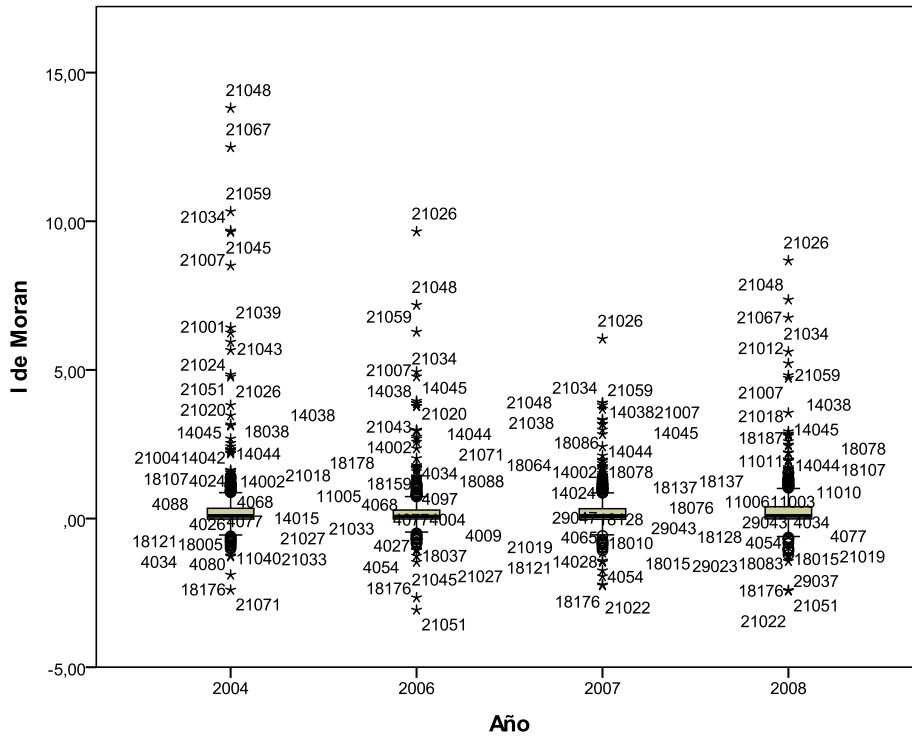
Fuente: Elaboración propia

Tabla 27: Estadísticas descriptivas del CAR Bayesiano  
(prevalencia administrativa de la depresión)

	2004	2006	2007	2008
Recuento	770	770	770	770
Media	1,0	0,96	0,99	1,0
Desviación estándar	0,57	0,50	0,51	0,62
Mediana	0,89	0,86	0,87	0,86
Mínimo	0,054	0,076	0,061	0,015
Máximo	3,9	4,3	3,8	3,8

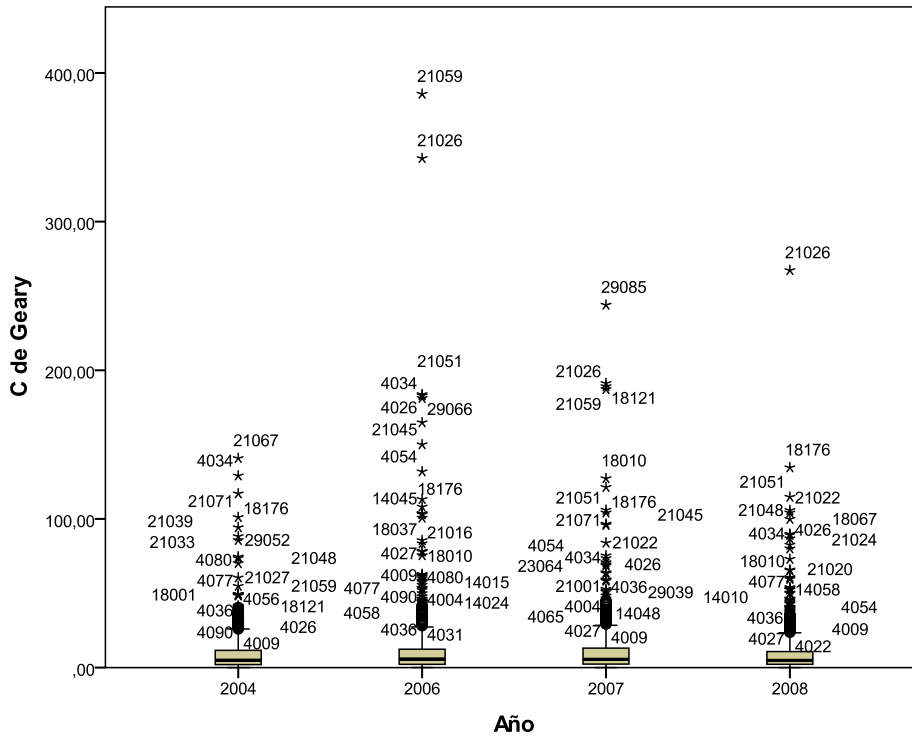
Fuente: Elaboración propia

Figura 20: *Box-Plot* de los valores de la I de Moran para la prevalencia de la depresión



Fuente: Elaboración propia

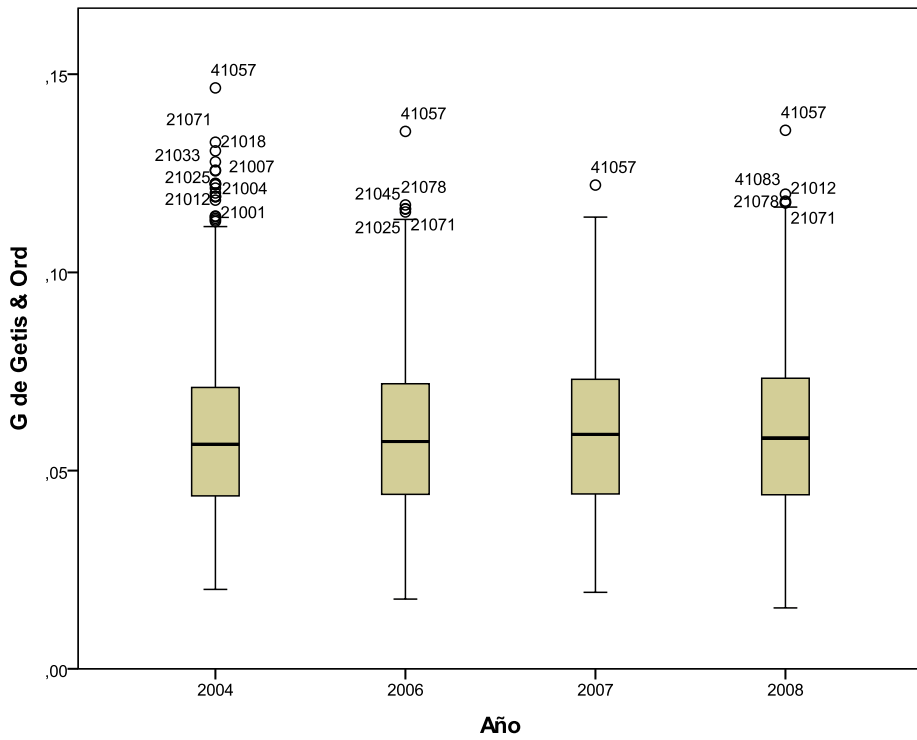
Figura 21: *Box-Plot* de los valores de la C de Geary para la prevalencia de la depresión



Fuente: Elaboración propia

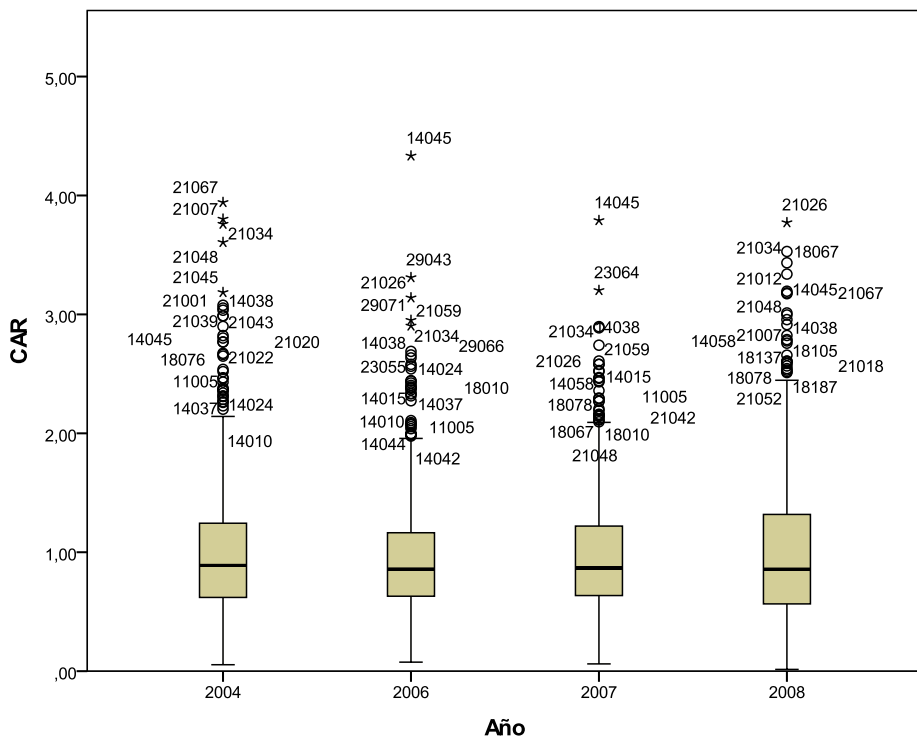


Figura 22: *Box-Plot* de los valores de la G de Getis & Ord para la prevalencia de la depresión



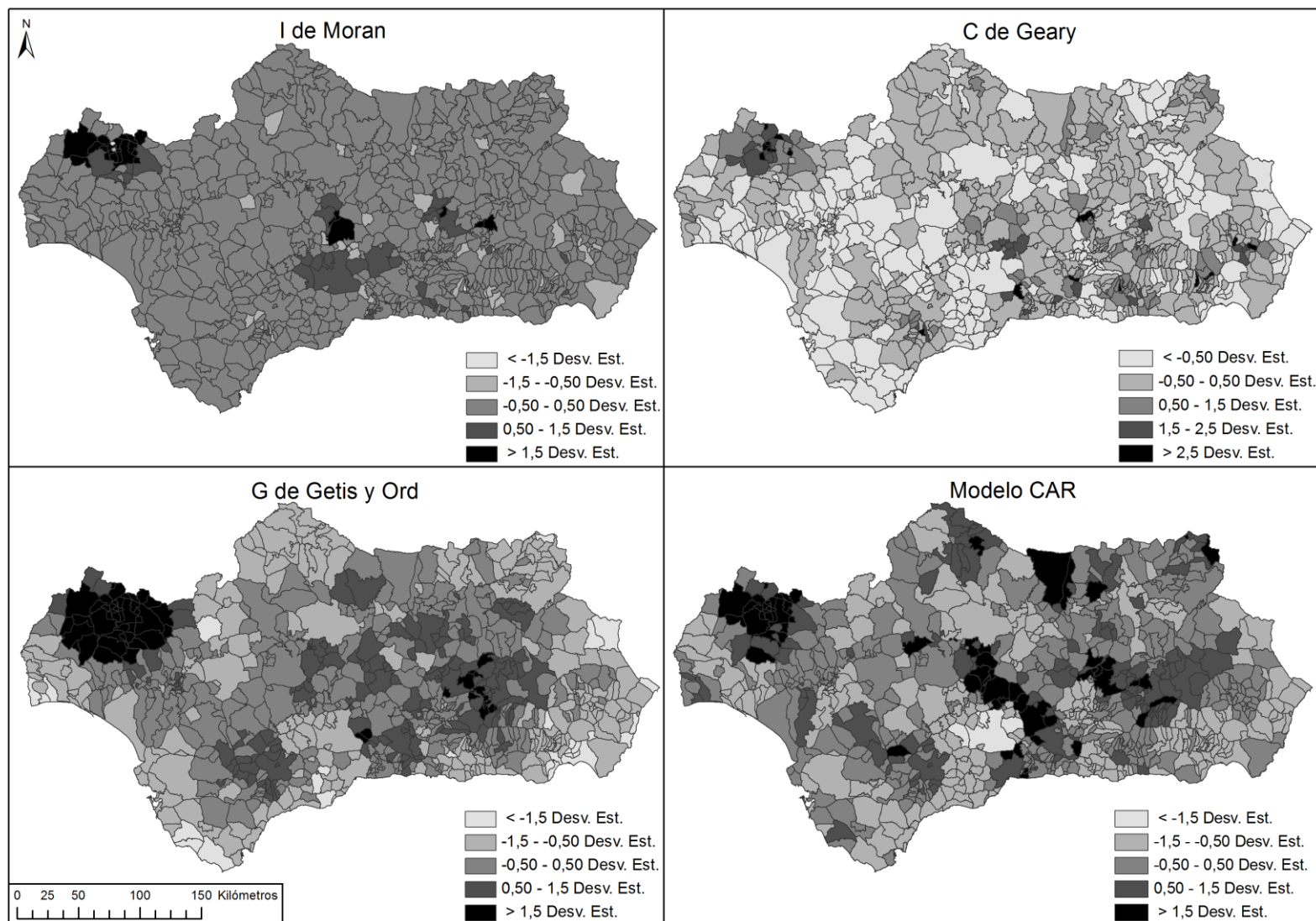
Fuente: Elaboración propia

Figura 23: *Box-Plot* de los valores del modelo CAR Bayesiano para la prevalencia de la depresión



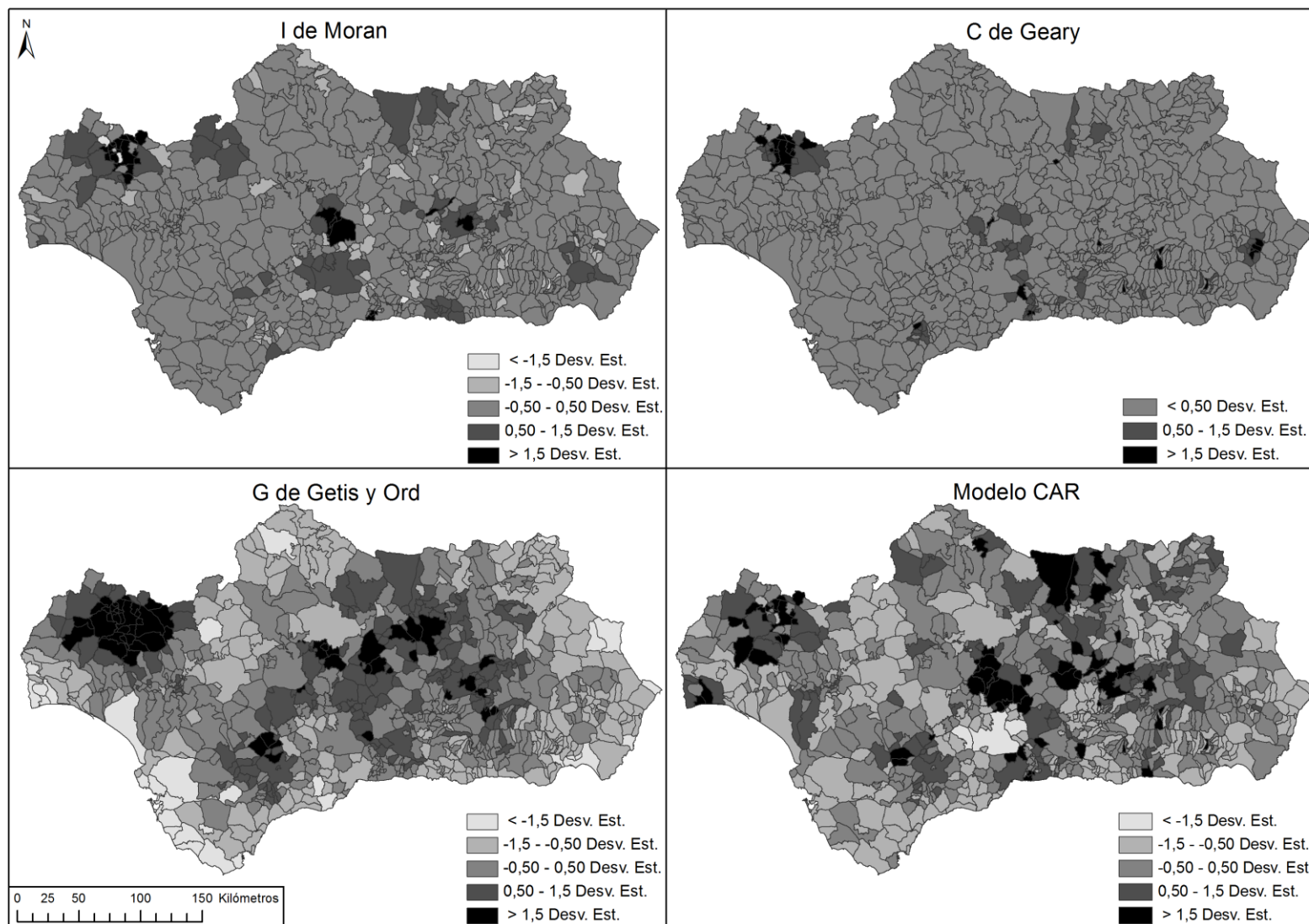
Fuente: Elaboración propia

Mapa 17: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la depresión en los municipios de Andalucía (2004)



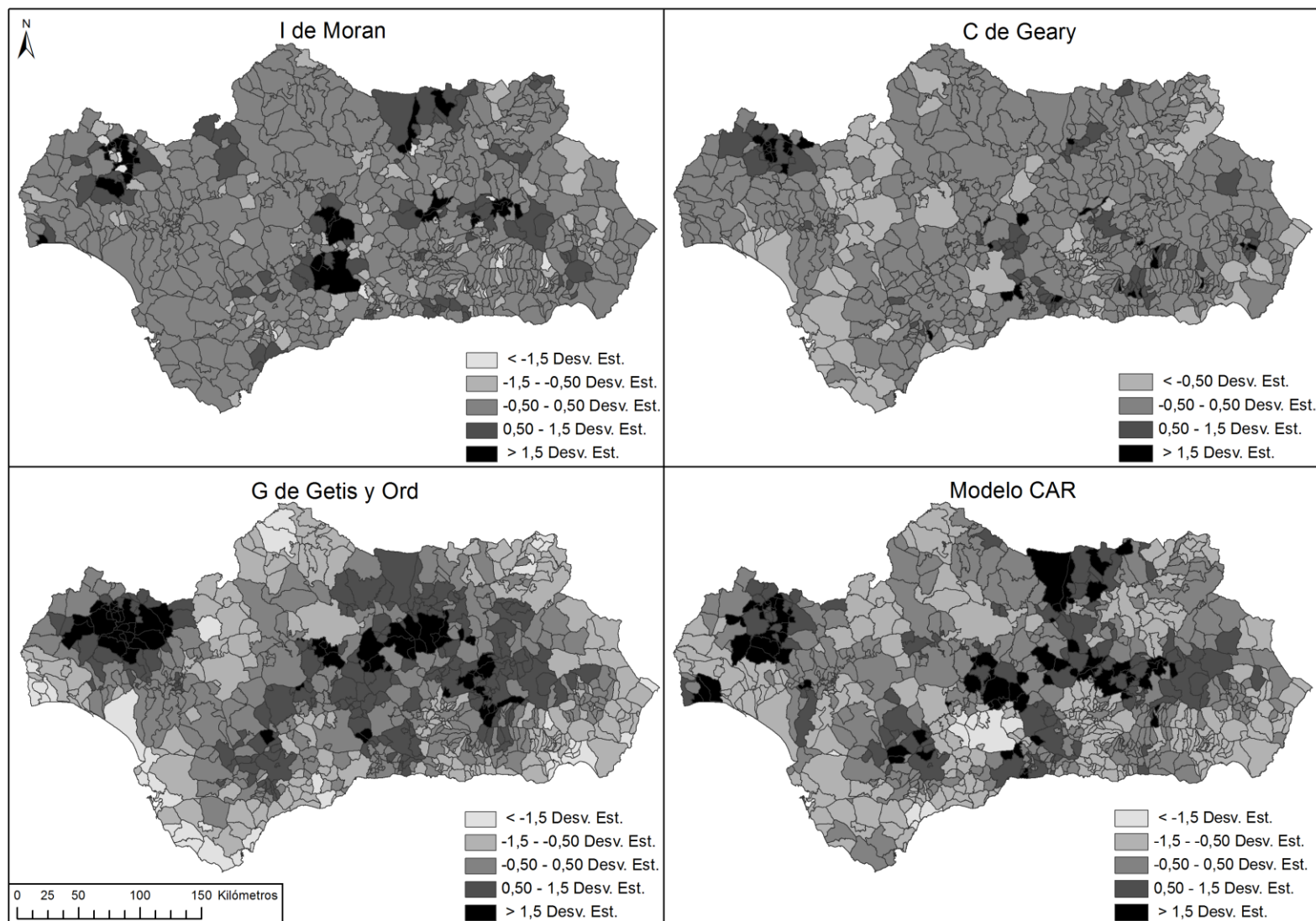
Fuente: Elaboración propia a partir del SISMA, 2004

Mapa 18: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la depresión en los municipios de Andalucía (2006)



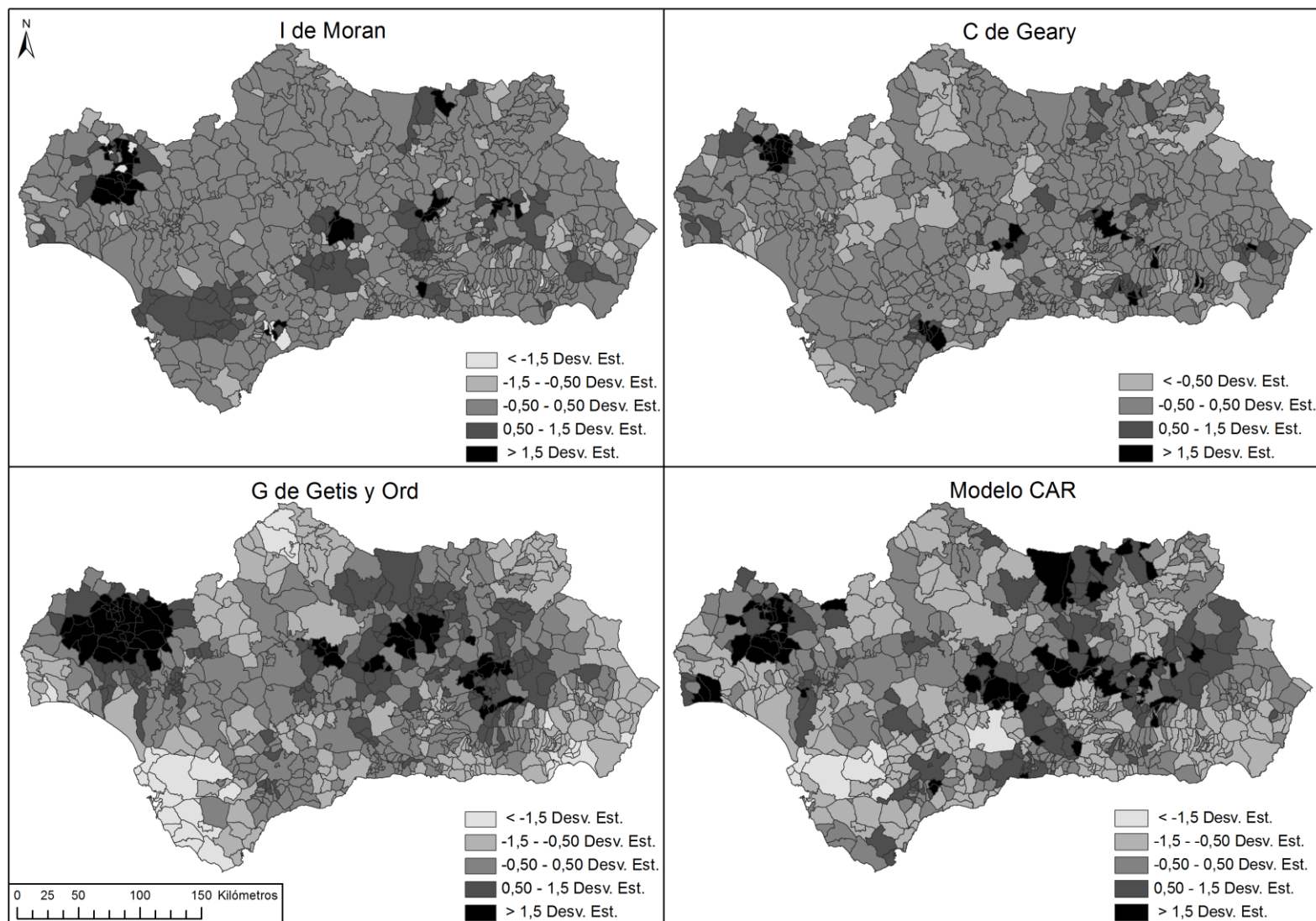
Fuente: Elaboración propia a partir del SISMA, 2006

Mapa 19: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la depresión en los municipios de Andalucía (2007)



Fuente: Elaboración propia a partir del SISMA, 2007

Mapa 20: Índices de autocorrelación local y CAR Bayesiano de la prevalencia administrativa de la depresión en los municipios de Andalucía (2008)



Fuente: Elaboración propia a partir del SISMA, 2008

### 4.2.2.3. Hot-spot de prevalencia administrativa de la depresión

#### Descripción estadística

Los *hot-spots* de prevalencia administrativa de la depresión por niveles de consenso de las *funciones de aptitud* y los métodos de selección se muestran en el Mapa 22, Mapa 24, Mapa 26 y Mapa 28. De la misma forma que con los resultados de la esquizofrenia se presentan también los grupos de municipios identificados por las cuatro funciones de aptitud del AEMO/AE en el Mapa 21, Mapa 23, Mapa 25 y Mapa 27.

Las estadísticas básicas de los municipios que componen los *hot-spots* tanto para los valores de los índices espaciales como para la prevalencia de la depresión se pueden ver en la Tabla 28, Tabla 29, Tabla 30, Tabla 31 y Tabla 32, y sus correspondientes histogramas quedan reflejados en la Figura 24.

Tabla 28: Estadísticas descriptivas de los *hot-spots* de depresión  
(prevalencia administrativa de la depresión)

	2004	2006	2007	2008
Recuento	14	20	20	17
Media	7,61	5,55	5,25	5,93
Desviación estándar	5,70	4,69	3,80	4,93
Mediana	7,43	3,30	3,45	3,05
Mínimo	0,83	0,98	1,45	1,15
Máximo	16,89	15,74	13,67	15,53

Fuente: Elaboración propia

Tabla 29: Estadísticas descriptivas de los *hot-spots* de esquizofrenia en 2004  
(LISA y CAR Bayesiano)

	I de Moran	C de Geary	G* de Getis	CAR
Media	4,65	23,85	0,87	2,01
Desviación estándar	5,03	22,53	0,03	1,42
Mediana	3,28	18,87	0,09	1,99
Mínimo	-0,20	1,47	0,04	0,48
Máximo	13,80	73,49	0,12	3,80

Fuente: Elaboración propia

Tabla 30: Estadísticas descriptivas de los *hot-spots* de esquizofrenia en 2006  
(LISA y CAR Bayesiano)

	<i>I</i> de Moran	<i>C</i> de Geary	<i>G</i> * de Getis	CAR
Media	1,55	23,89	0,07	1,39
Desviación estándar	2,27	31,28	0,02	0,99
Mediana	0,17	6,15	0,07	0,93
Mínimo	-0,06	1,91	0,04	0,44
Máximo	7,18	103,53	0,12	4,33

Fuente: Elaboración propia

Tabla 31: Estadísticas descriptivas de los *hot-spots* de esquizofrenia en 2007  
(LISA y CAR Bayesiano)

	<i>I</i> de Moran	<i>C</i> de Geary	<i>G</i> * de Getis	CAR
Media	1,10	16,22	0,06	1,40
Desviación estándar	1,62	19,37	0,02	0,96
Mediana	0,14	7,00	0,06	0,88
Mínimo	-0,14	1,59	0,04	0,63
Máximo	3,89	68,40	0,10	3,79

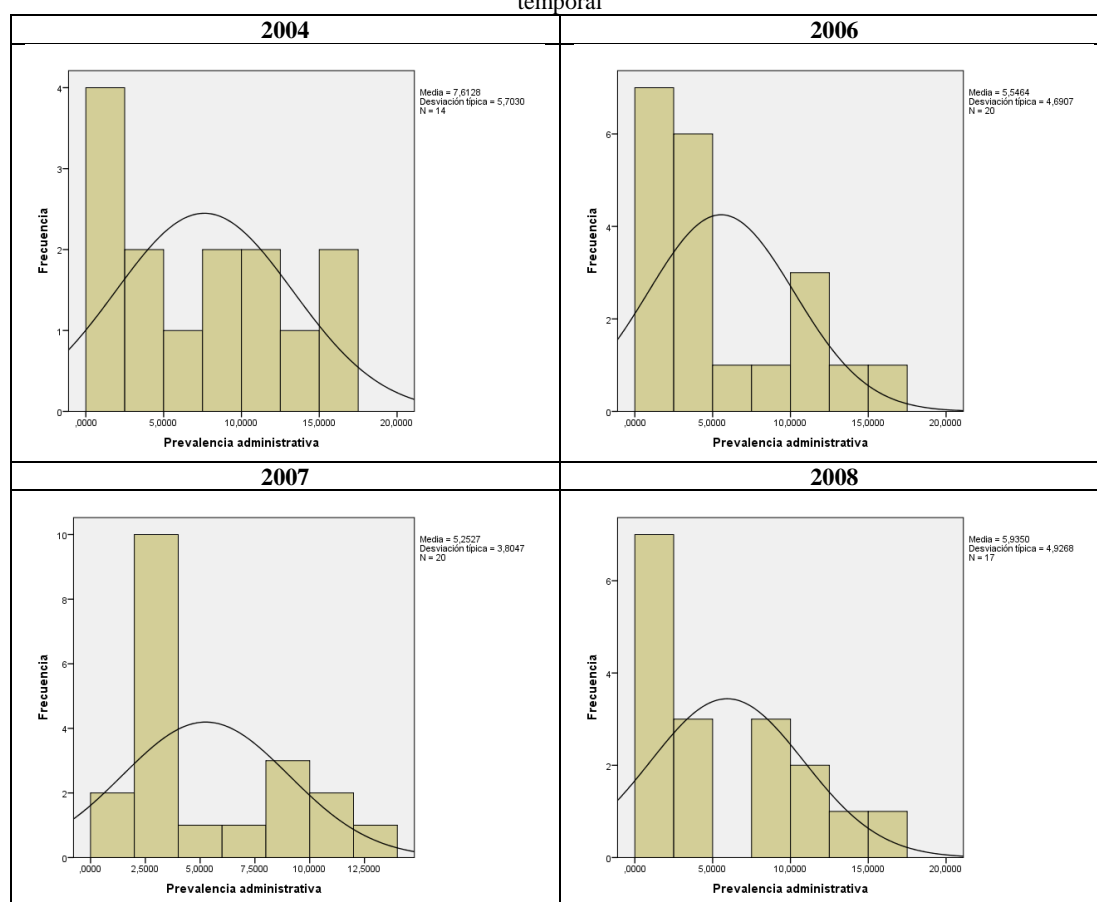
Fuente: Elaboración propia

Tabla 32: Estadísticas descriptivas de los *hot-spots* de esquizofrenia en 2008  
(LISA y CAR Bayesiano)

	<i>I</i> de Moran	<i>C</i> de Geary	<i>G</i> * de Getis	CAR
Media	2,16	21,60	0,07	1,63
Desviación estándar	2,72	30,51	0,03	1,21
Mediana	0,44	8,45	0,06	0,83
Mínimo	-0,17	1,36	0,03	0,52
Máximo	7,36	105,76	0,12	3,43

Fuente: Elaboración propia

Figura 24: Histogramas de frecuencia de la prevalencia administrativa de la depresión de los *hot-spots* en la serie temporal



Fuente: SISMA

Las unidades seleccionadas como *hot-spots* de depresión, al igual que con las de esquizofrenia, también han sido numeradas mediante la abreviatura HSd (*hot-spot* depresión) y números consecutivos en función a su proximidad (Tabla 33). También se muestra en que años aparece cada municipio como *hot-spot* y la provincia y área de salud mental a la que pertenece. Los *hot-spots* han aparecido en 10 de las 75 áreas de salud mental andaluzas.

Tabla 33: Municipios que forman parte de los *hot-spots* de depresión durante la serie temporal

HSe	Municipio	2004	2006	2007	2008	Prov.	Área SM
HSd1	Abla	x	x	x	x	AL	Almería
HSd1	Abrucena	x	x	x	x	AL	Almería
HSd1	Alboloduy		x	x	x	AL	Almería
HSd1	Canjáyar				x	AL	Almería
HSd1	Fiñana		x		x	AL	Almería
HSd1	Ohanes	x	x	x		AL	Almería
HSd1	Adra			x		AL	El Ejido
HSd1	Alcolea	x	x	x	x	AL	El Ejido



HSe	Municipio	2004	2006	2007	2008	Prov.	Área SM
HSd1	Bayárcal	x		x		AL	El Ejido
HSd1	Berja		x	x		AL	El Ejido
HSd1	Dalías			x		AL	El Ejido
HSd1	Fondón	x	x	x	x	AL	El Ejido
HSd1	Láujar de Andarax		x	x		AL	El Ejido
HSd1	Paterna del Río			x		AL	El Ejido
HSd2	Alájar	x				HU	Aracena
HSd2	Aracena	x	x	x	x	HU	Aracena
HSd2	Cortegana		x			HU	Aracena
HSd2	Galaroza	x	x	x	x	HU	Aracena
HSd2	Higuera de la Sierra		x	x		HU	Aracena
HSd2	Hiojales		x			HU	Aracena
HSd2	Jabugo	x				HU	Aracena
HSd2	Linares de la Sierra	x				HU	Aracena
HSd2	Los Marines	x	x	x	x	HU	Aracena
HSd2	Puerto Moral	x				HU	Aracena
HSd2	Santa Ana la Real				x	HU	Aracena
HSd2	Berrocal				x	HU	Valverde del Camino
HSd2	El Campillo				x	HU	Valverde del Camino
HSd2	Zalamea la Real				x	HU	Valverde del Camino
HSd3	Albánchez				x	AL	Albox
HSd3	Alcóntar			x		AL	Albox
HSd3	Bacares				x	AL	Albox
HSd3	Sierro	x		x		AL	Albox
HSd4	Lucena			x		CO	Lucena
HSd4	Moriles		x	x		CO	Lucena
HSd5	Coria del Río		x			SE	Mairena
HSd5	Las Cabezas de San Juan		x			SE	Utrera
HSd5	Bornos		x			CA	Villamartín
HSd5	Villamartín		x			CA	Villamartín
HSd6	El Guijo				x	CO	Pozoblanco

Fuente: Elaboración propia

### Comparación de las distribuciones y medianas.

La comparación entre distribuciones mediante la Prueba  $U$  muestra situaciones diferentes entre índices y años. Las distribuciones de la  $I$  son diferentes entre *hot-spots* y el resto de unidades espaciales durante toda la serie temporal, la  $G^*$  en todos los años excepto para 2007, el  $CAR$  en 2006 y 2007 y la  $C$  en 2004 ( $p < 0,05$ ) (Tabla 34). El valor  $CAR$  presenta la misma distribución en los cuatro años, la  $C$  de Geary en 2006, 2007 y 2008, la  $G$  de Getis y Ord en 2007 y 2008 y la  $I$  de Moran en 2007. Por su parte, la prueba de las medianas avala la hipótesis nula de igualdad en la mayor parte de los casos. Así, las medianas son semejantes estadísticamente en la mayor parte de los casos excepto en la  $I$  de Moran para 2004 y 2008 y la  $G^*$  de Getis y Ord en 2004.

También se aprecian diferencias en el comportamiento de los valores de prevalencia de la depresión entre unidades espaciales que son *hot-spots* y aquellas que no lo son durante la serie temporal. En este caso, la Prueba *U* rechaza la hipótesis nula en los cuatro años señalando que las distribuciones son diferentes. La Prueba de las medianas durante los cuatro años presenta una significación superior al 0,05, por lo que no se puede rechazar la hipótesis nula de igualdad de las medianas.

Tabla 34: Comparación de las distribuciones y medianas de las unidades espaciales que forman los diferentes *hot-spots* de depresión y aquellas que no forman parte de ninguno

	<b>Prueba</b>	<b>Hipótesis nula</b>	<b>2004</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>
Dependencia espacial	Prueba Kruskal-Wallis de muestras independientes	H <sub>0</sub> : la distribución de la prevalencia es la misma en las unidades espaciales identificadas como hot-spots (HSd1, HSd2, HSd3, HSd4, HSd5 Y HSd6) y las que no lo han sido.	<i>I</i> <b>Rechazar</b> (sig. 0,001 < 0,05)	<i>I</i> <b>Rechazar</b> (sig. 0,004 < 0,05)	<i>I</i> <b>Rechazar</b> (sig. 0,015 < 0,05)	<i>I</i> <b>Rechazar</b> (sig. 0,001 < 0,05)
			<i>C</i> <b>Rechazar</b> (sig. 0,022 < 0,05)	<i>C</i> Retener (sig. 0,126 > 0,05)	<i>C</i> Retener (sig. 0,275 > 0,05)	<i>C</i> Retener (sig. 0,178 > 0,05)
			<i>G*</i> <b>Rechazar</b> (sig. 0,002 < 0,05)	<i>G*</i> <b>Rechazar</b> (sig. 0,012 < 0,05)	<i>G*</i> Retener (sig. 0,125 > 0,05)	<i>G*</i> <b>Rechazar</b> (sig. 0,009 < 0,05)
			<i>CAR</i> Retener (sig. 0,097 > 0,05)	<i>CAR</i> <b>Rechazar</b> (sig. 0,036 < 0,05)	<i>CAR</i> Retener (sig. 0,153 > 0,05)	<i>CAR</i> <b>Rechazar</b> (sig. 0,031 < 0,05)
	Prueba de medianas de muestras independientes	H <sub>0</sub> : la mediana de la distribución de la prevalencia es la misma en las unidades espaciales identificadas como hot-spots (HSd1, HSd2, HSd3, HSd4, HSd5 Y HSd6) y las que no lo han sido.	<i>I</i> <b>Rechazar</b> (sig. 0,024 < 0,05)	<i>I</i> Retener (sig. 0,054 > 0,05)	<i>I</i> Retener (sig. 0,118 > 0,05)	<i>I</i> <b>Rechazar</b> (sig. 0,024 < 0,05)
			<i>C</i> Retener (sig. 0,107 > 0,05)	<i>C</i> Retener (sig. 0,659 > 0,05)	<i>C</i> Retener (sig. 0,503 > 0,05)	<i>C</i> Retener (sig. 0,421 > 0,05)
			<i>G*</i> <b>Rechazar</b> (sig. 0,024 < 0,05)	<i>G*</i> Retener (sig. 0,085 > 0,05)	<i>G*</i> Retener (sig. 0,498 > 0,05)	<i>G*</i> Retener (sig. 0,430 > 0,05)
			<i>CAR</i> Retener (sig. 0,814 > 0,05)	<i>CAR</i> Retener (sig. 0,659 > 0,05)	<i>CAR</i> Retener (sig. 0,896 > 0,05)	<i>CAR</i> Retener (sig. 0,236 > 0,05)

	<b>Prueba</b>	<b>Hipótesis nula</b>	<b>2004</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>
Prevalencia administrativa	Prueba Kruskal-Wallis de muestras independientes	H <sub>0</sub> : la distribución de la prevalencia es la misma en las unidades espaciales identificadas como hot-spots (HSd1, HSd2, HSd3, HSd4, HSd5 Y HSd6) y las que no lo han sido.	<b>Rechazar</b> (sig. 0,012 < 0,05)	<b>Rechazar</b> (sig. 0,027 < 0,05)	<b>Rechazar</b> (sig. 0,044 < 0,05)	<b>Rechazar</b> (sig. 0,015 < 0,05)
	Prueba de medianas de muestras independientes	H <sub>0</sub> : la mediana de la distribución de la prevalencia es la misma en las unidades espaciales identificadas como hot-spots (HSd1, HSd2, HSd3, HSd4, HSd5 Y HSd6) y las que no lo han sido.	Retener (sig. 0,264 > 0,05)	Retener (sig. 0,775 > 0,05)	Retener (sig. 0,891 < 0,05)	Retener (sig. 0,521 < 0,05)

Fuente: Elaboración propia

## Descripción de los *hot-spots*

### *Año 2004*

El AEMO/AE ha localizado dos *hot-spots* de primer nivel en el año 2004. El primero se sitúa en la zona occidental de Andalucía, concretamente en el área de salud mental de Aracena en Huelva (HSd2). El segundo *hot-spot* se encuentra en la zona oriental en las áreas de salud mental limítrofes de Almería y el Ejido en la provincia de Almería (HSd1). Los *hot-spot* de segundo nivel se asocian a éstos, excepto en el caso de una unidad espacial aislada en el área de Albox (HSd3).

### *Año 2006*

En 2006, los *hot-spots* se componen de un mayor número de unidades espaciales que en el año 2004. En este caso la mayor superficie la posee el *hot-spot* de primer nivel situado en las áreas de salud mental de Almería y el Ejido (HSd1) que crece hacia la costa. El *hot-spot* de la comarca de Sierra de Aracena vuelve a aparecer (HSd2). Además se ha detectado un *hot-spot* de primer nivel que marcaría la existencia de una concentración de elevada dependencia espacial en el área de salud mental de Lucena en la Subbética cordobesa (HSd4). Un último *hot-spot* de segundo nivel (HSd5) aparece en las áreas de Mairena y Utrera en la provincia de Sevilla y Villamartín en la provincia de Cádiz.

### *Año 2007*

En 2007, los tres *hot-spots* principales del año anterior se mantienen: Aracena (HSd2), Almería y el Ejido (HSd1) y Albox (HSd3). Aparecen unidades espaciales de segundo nivel que unen HSd1 a HSe3 (detectado en 2004) y también extiende su superficie hacia el área litoral de el Ejido. Junto a estos vuelve a aparecer el *hot-spot* de la Subbética Cordobesa (HSd4).

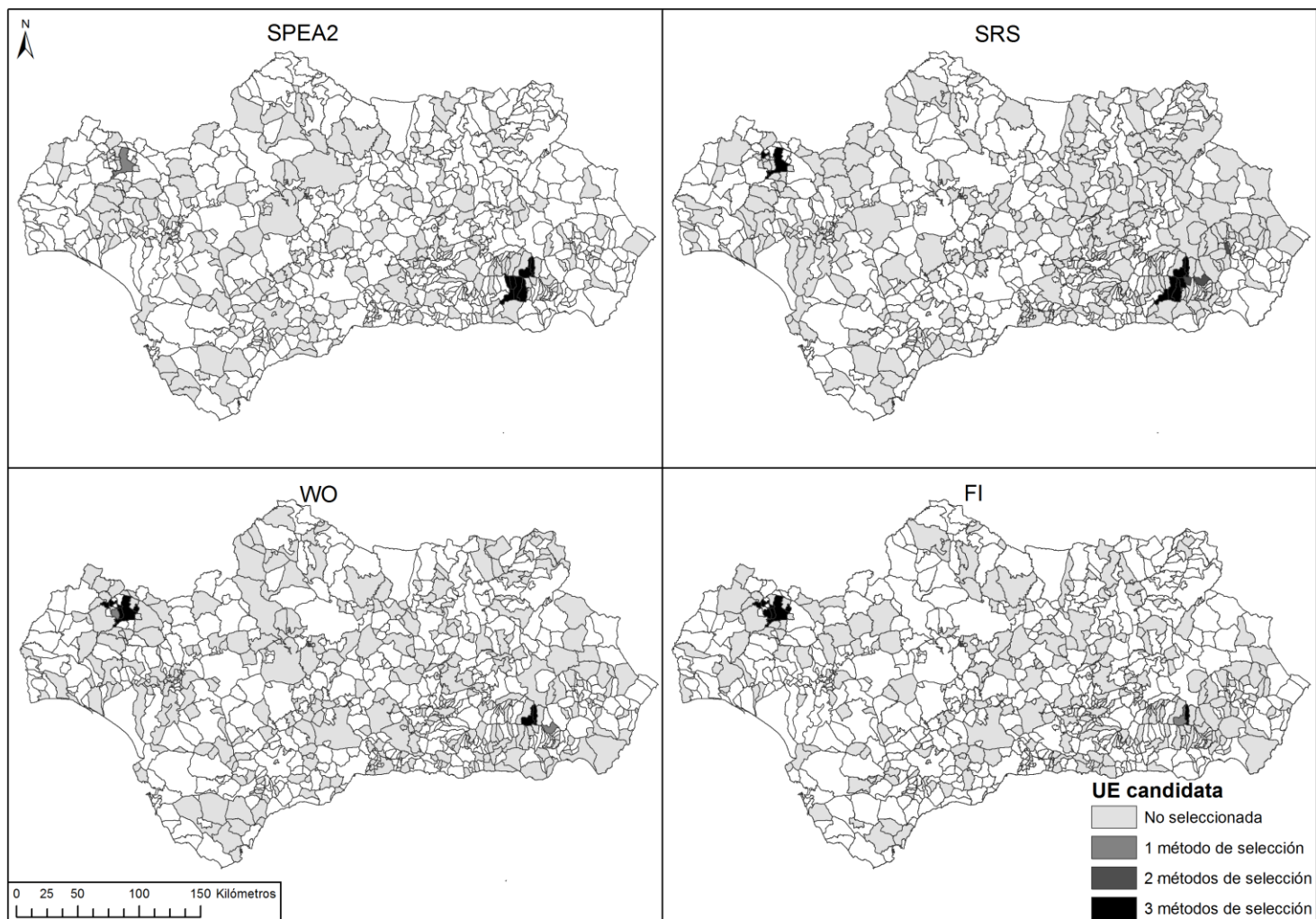
### *Año 2008*

Finalmente en 2008, se mantienen los *hot-spot* de primer nivel del área de salud mental de Aracena (HSd2) y Almería/El Ejido (HSd1) aunque desaparece el *hot-spot* de Lucena. El *hot-spot* de Aracena se expande en esta ocasión hacia el sur al área de salud mental de Valverde del Camino. En Granada, también aparece el *hot-spot* de Albox (HSd3) que apareció en 2004. El AEMO/AE ha localizado también un nuevo *hot-spot* aislado de segundo nivel en el área de Pozoblanco en la provincia de Córdoba (HSd6).

### *Comentario de la serie*

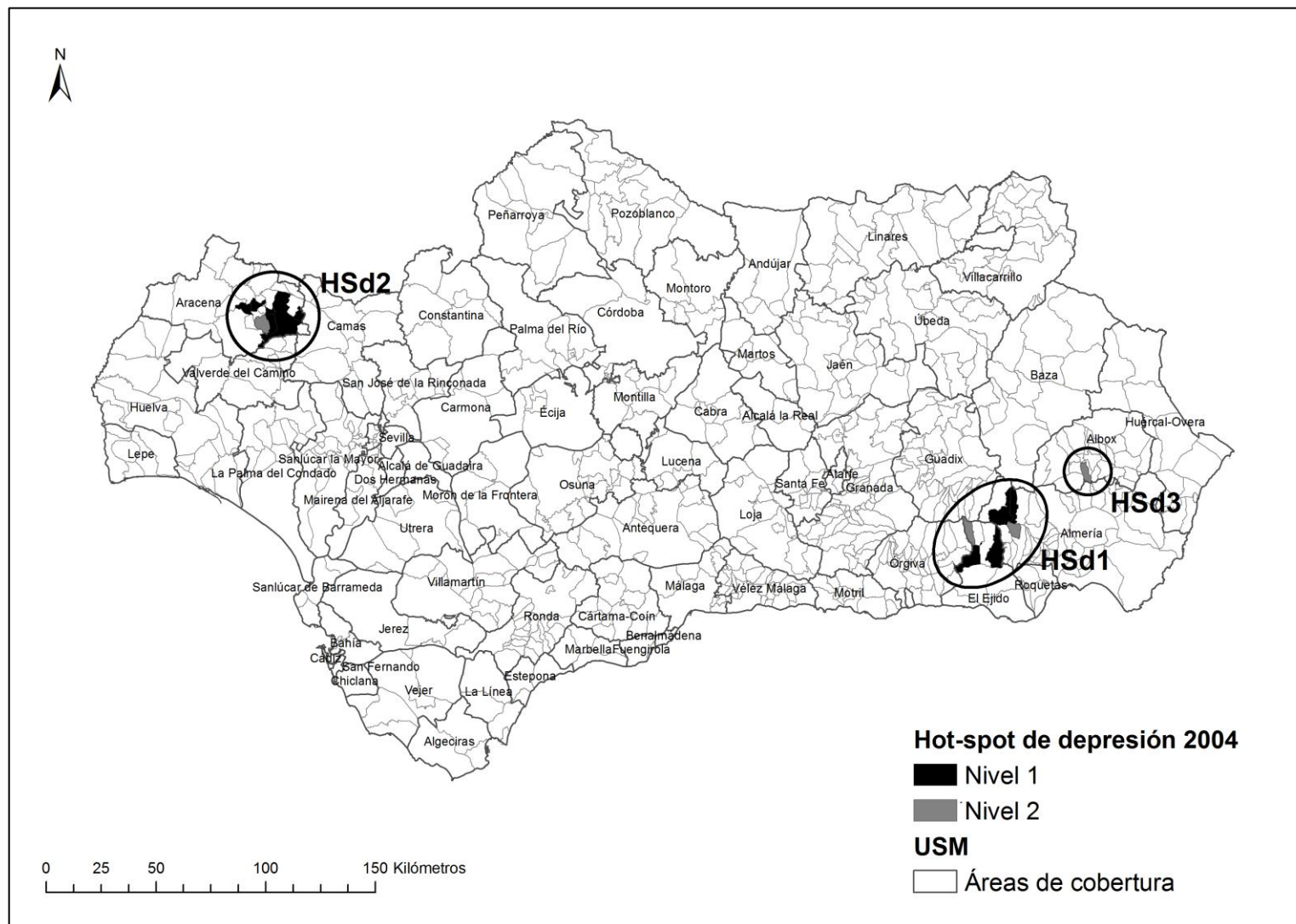
Los *hot-spot* de depresión se ubican en dos localizaciones principales que se mantienen estables a lo largo de los años. La primera se localiza en la Sierra de Aracena (HSd2) y la segunda en la zona oriental de Sierra Nevada dentro de la provincia de Almería, situada en las comarcas de Los Filabres-Tabernas y en La Alpujarra (HSd1). También aparecen otros *hot-spot* en años concretos, así en 2004, 2007 y 2008 en el Valle del Almanzora en Almería (HSd3); en 2006 en el Bajo Guadalquivir y Sierra de Cádiz (HSd5); en 2006 y 2007 en la Subbética Cordobesa (HSd4) y en 2008 en los Pedroches (HSd6). Los *hot-spots* detectados podían intuirse en alguno de los mapas de los LISA y el CAR Bayesiano vistos anteriormente, pero el AEMO permite señalarlos claramente facilitando su interpretación.

Mapa 21: Resultados del AEMO/AE por función de aptitud para los *hot-spot* de depresión en Andalucía (2004)  
 Unidades espaciales (UE) candidatas y número de métodos de selección que la han identificado como *hot-spot* potencial



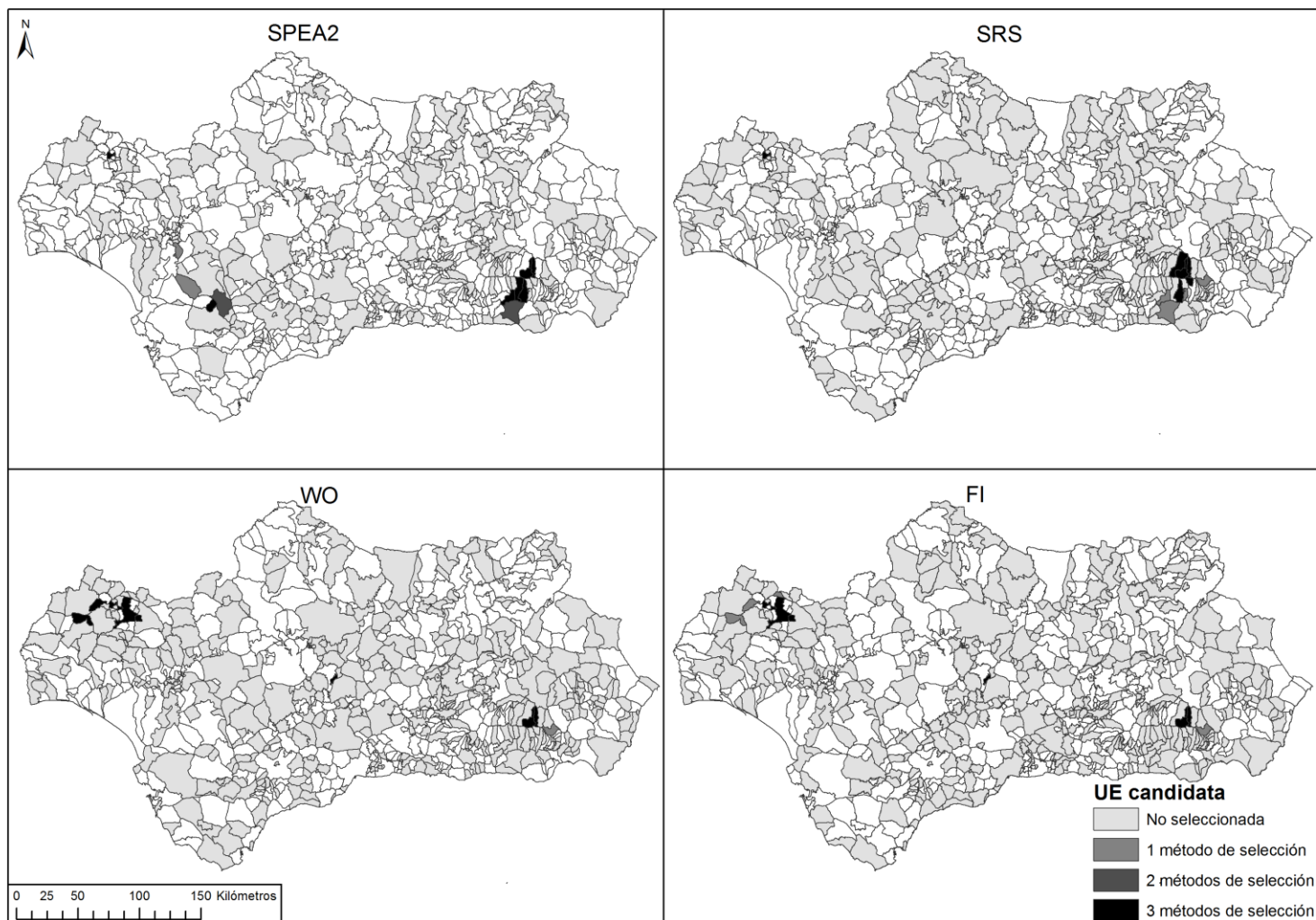
Fuente: Elaboración propia a partir del SISMA, 2004

Mapa 22: Hot-spots de dependencia espacial de la depresión en Andalucía (2004)



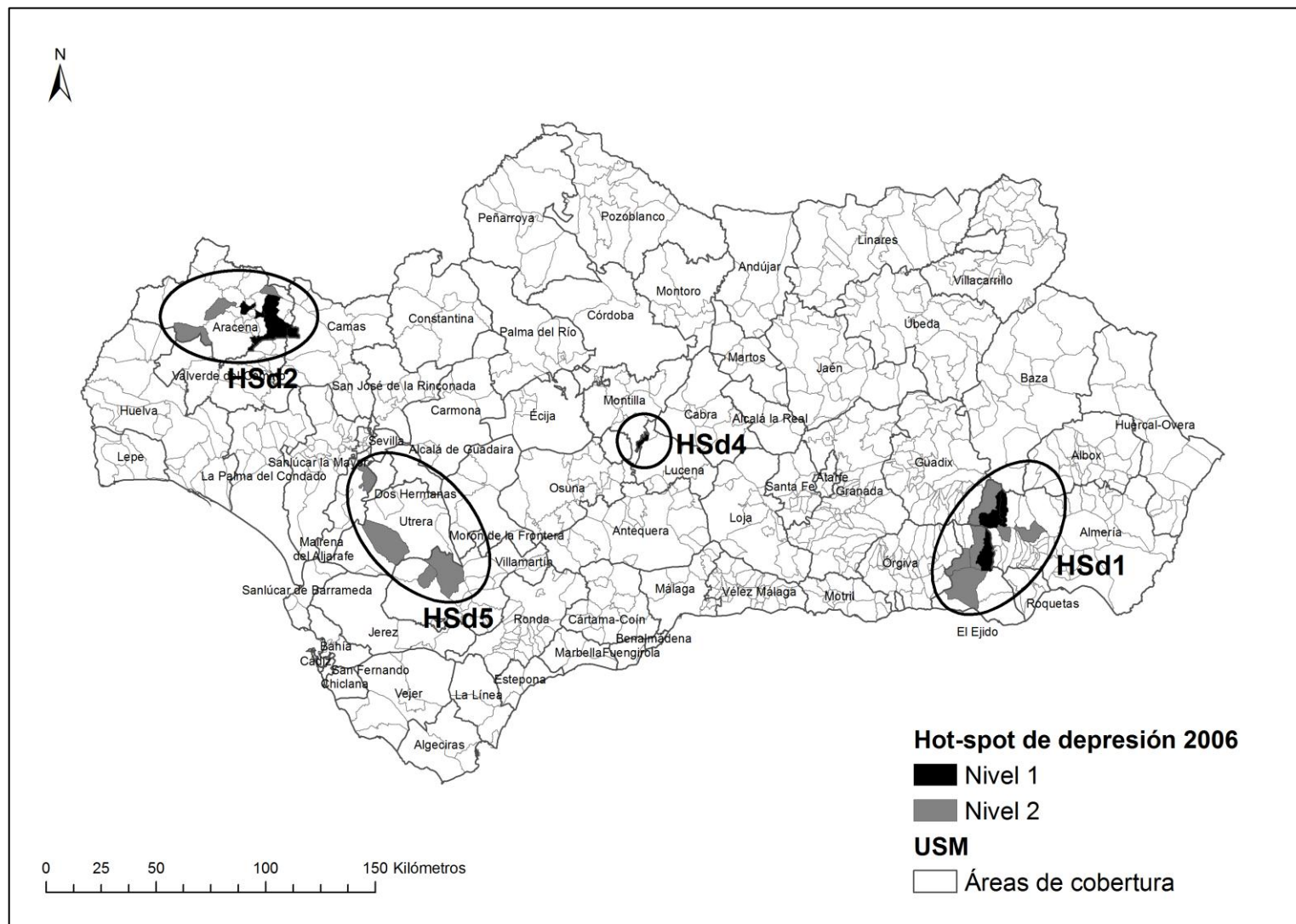
Fuente: Elaboración propia a partir del SISMA, 2004

Mapa 23: Resultados del AEMO/AE por función de aptitud para los *hot-spot* de depresión en Andalucía (2006)  
Unidades espaciales (UE) candidatas y número de métodos de selección que la han identificado como *hot-spot* potencial



Fuente: Elaboración propia a partir del SISMA, 2006

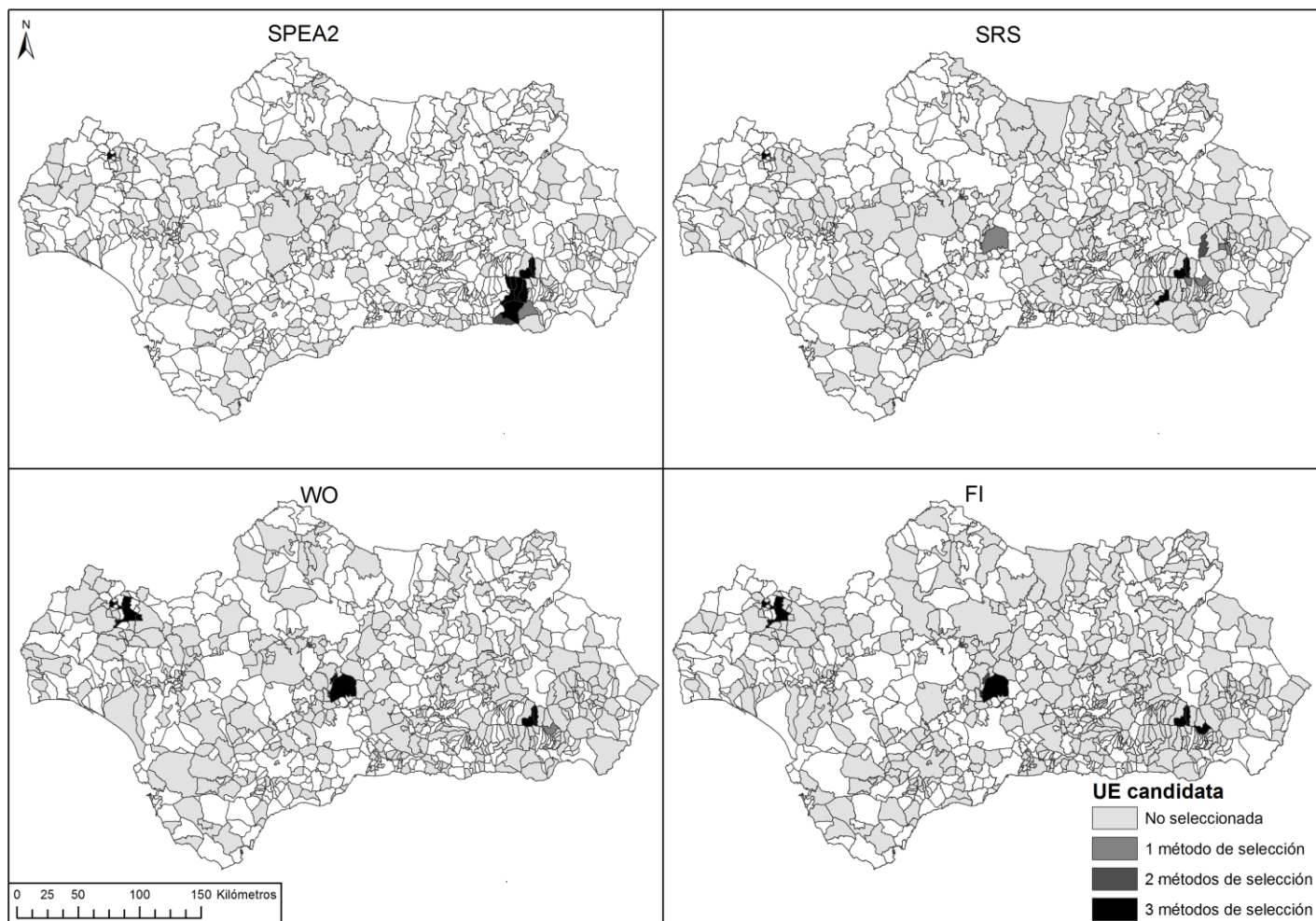
Mapa 24: Hot-spots de dependencia espacial de la depresión en Andalucía (2006)



Fuente: Elaboración propia a partir del SISMA, 2006

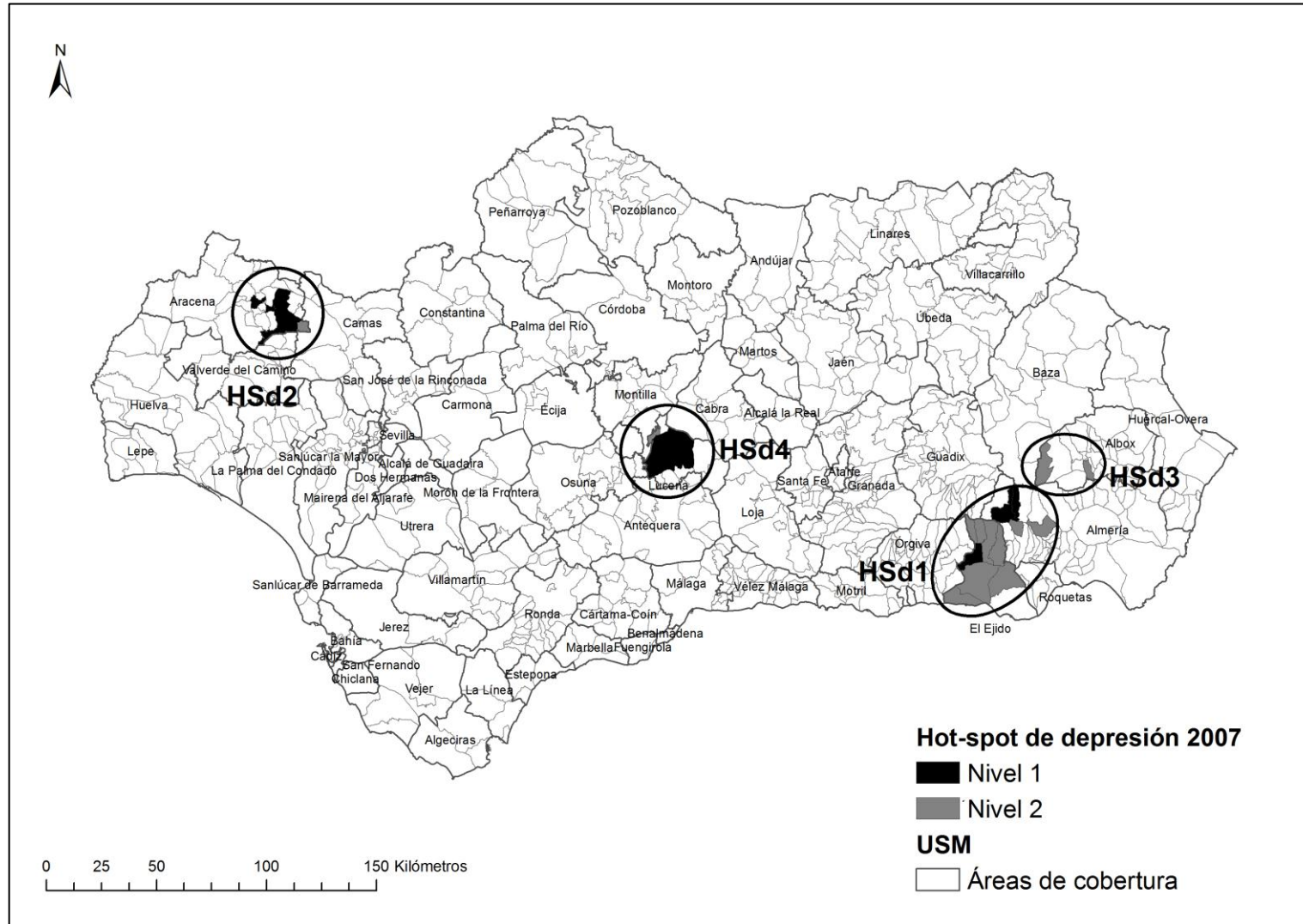


Mapa 25: Resultados del AEMO/AE por función de aptitud para los hot-spot de depresión en Andalucía (2007)  
 Unidades espaciales (UE) candidatas y número de métodos de selección que la han identificado como hot-spot potencial



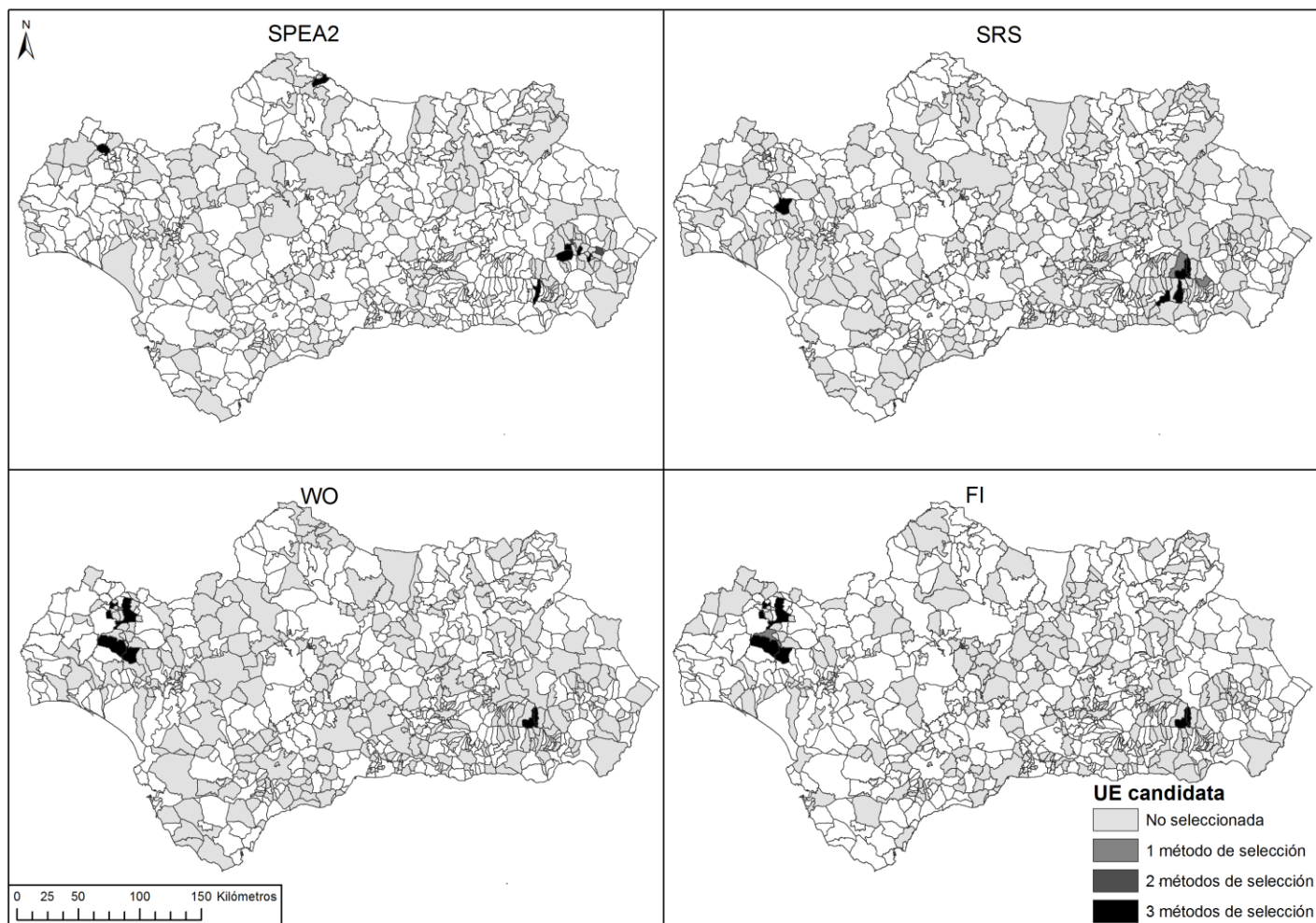
Fuente: Elaboración propia a partir del SISMA, 2007

Mapa 26: Hot-spots de dependencia espacial de la depresión en Andalucía (2007)



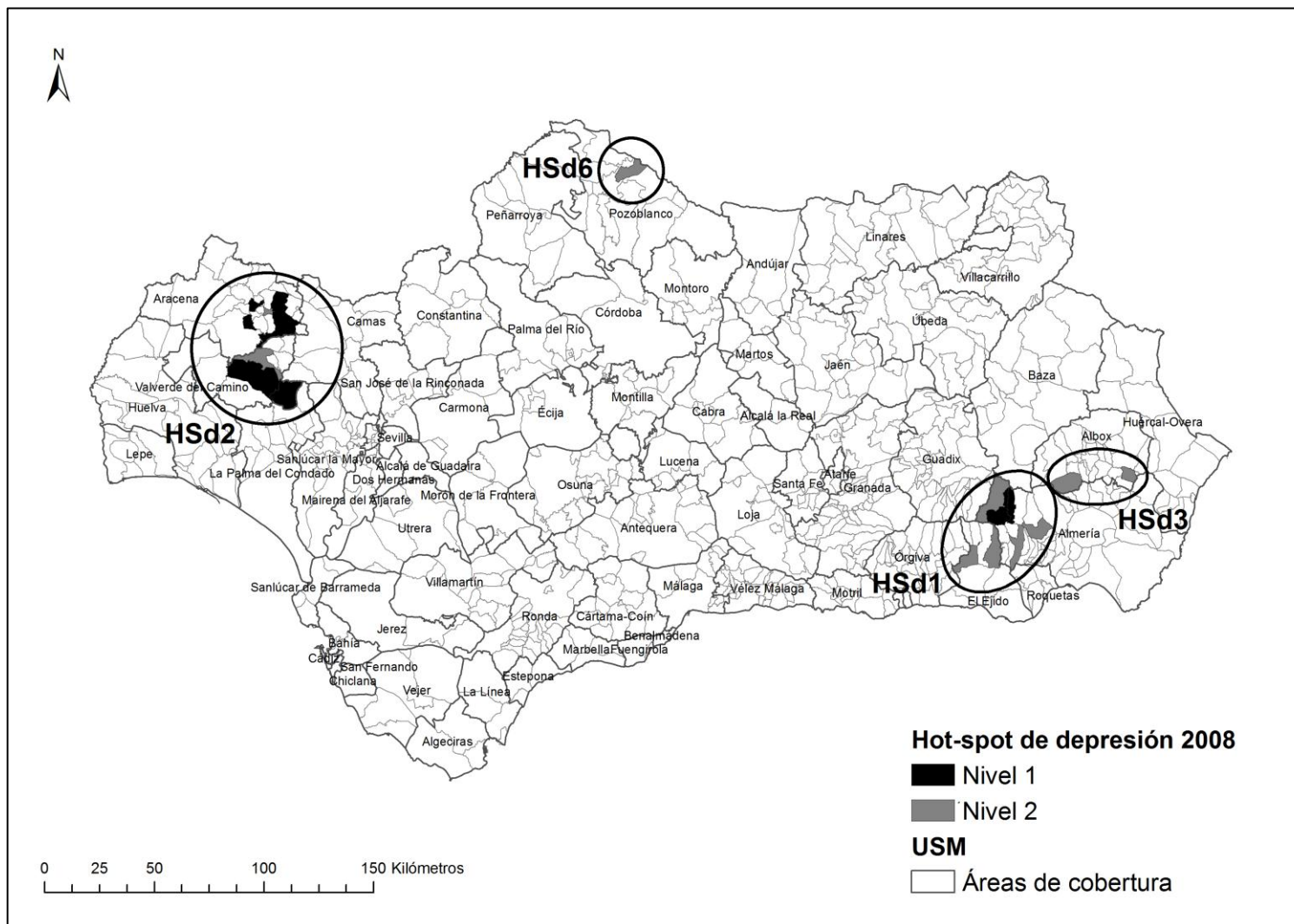
Fuente: Elaboración propia a partir del SISMA, 2007

Mapa 27: Resultados del AEMO/AE por función de aptitud para los *hot-spot* de depresión en Andalucía (2008)  
 Unidades espaciales (UE) candidatas y número de métodos de selección que la han identificado como *hot-spot* potencial



Fuente: Elaboración propia a partir del SISMA, 2008

Mapa 28: Hot-spots de dependencia espacial de la depresión en Andalucía (2008)



Fuente: Elaboración propia a partir del SISMA, 2008

### 4.3. Estudio de Demostración B: aplicación del AEMO/AE sobre datos de Enfermedades Mentales en Cataluña

#### 4.3.1. Caso 3: *Hot-spots* y *cold-spots* de prevalencia asistida de la esquizofrenia en Cataluña en 2009

##### 4.3.1.1. Distribución geográfica de la prevalencia administrativa de la esquizofrenia

La prevalencia administrativa de la esquizofrenia en los 946 municipios catalanes tuvo una media de 2,65 casos por 1.000 habitantes, una mediana de 2,27 y una desviación estándar de 2,97 en 2009 (Tabla 35). El 24,84% de los municipios no contó con casos de esquizofrenia en 2009, mientras que el valor máximo de prevalencia alcanzó el valor de 31,14 casos por 1.000 habitantes. La distribución de la prevalencia administrativa no se ajusta a una distribución normal como se puede observar en el histograma de frecuencias y confirma la prueba de Kolmogorov-Smirnov con una significación inferior a 0,05 (Figura 25).

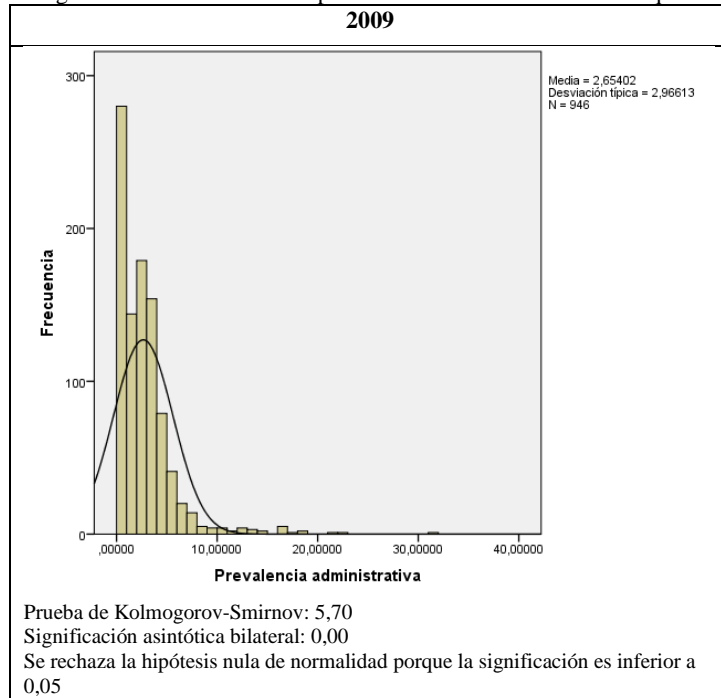
El *box-plot* de la distribución de la prevalencia de la esquizofrenia muestra la existencia de diversos *outliers* por encima de 1,5 veces (valores atípicos) y 3 veces (valores extremos) el recorrido intercuartílico entre el primer y el tercer cuartil ( $Q_3-Q_1$ ) por la derecha de la distribución (valores altos de prevalencia) (Figura 26).

Tabla 35: Estadísticas descriptivas de la prevalencia administrativa de la esquizofrenia en Cataluña

	2009
Recuento	946
Media	2,65
Desviación estándar	2,97
Mediana	2,27
Mínimo	0 (235 UE)
Máximo	31,14

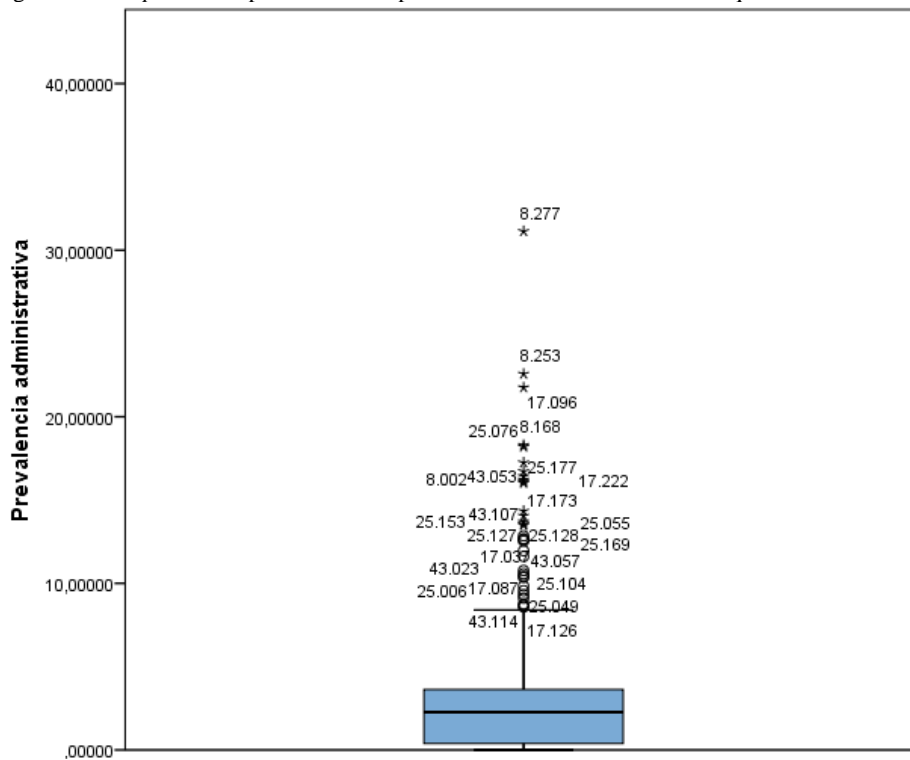
Fuente: CMBD-SMA, 2009

Figura 25: Histogramas de frecuencia de la prevalencia administrativa de la esquizofrenia en 2009



Fuente: CMBS-SMA, 2009

Figura 26: Box-plot de las poblaciones de prevalencia administrativa de la esquizofrenia en 2009



Fuente: CMBS-SMA, 2009

La dispersión de la prevalencia administrativa de la esquizofrenia es amplia como se observa en la Tabla 36. Hay 19 municipios con valores que se encuentran entre 1,5 y 2,5 veces el valor de la desviación estándar desde la media, y 26 municipios que superan la media en 2,5 veces la desviación estándar.

Tabla 36: Distribución de frecuencias de la prevalencia administrativa de la esquizofrenia en Cataluña

	2009
-1,5 a -0,5 Desv. Est.	301
-0,5 a 0,5 Desv. Est.	467
0,5 a 1,5 Desv. Est.	133
1,5 a 2,5 Desv. Est.	19
>2,5 Desv. Est.	26

Fuente: CMBS-SMA, 2009

El análisis espacial exploratorio realizado mediante la *I* de Moran global y la *G* de Getis y Ord global indica que existe una alta probabilidad de que existan patrones espaciales en la distribución geográfica de la prevalencia de la esquizofrenia con una significación de la prueba (*p*) del 0,01 (Tabla 44). La *G* indica que los clústeres de valores de prevalencia son de valores elevados con una alta significación. A pesar de esto, el AEMO/AE ha detectado concentraciones de valores bajos en Cataluña que eran enmascarados por la distribución general en los análisis globales.

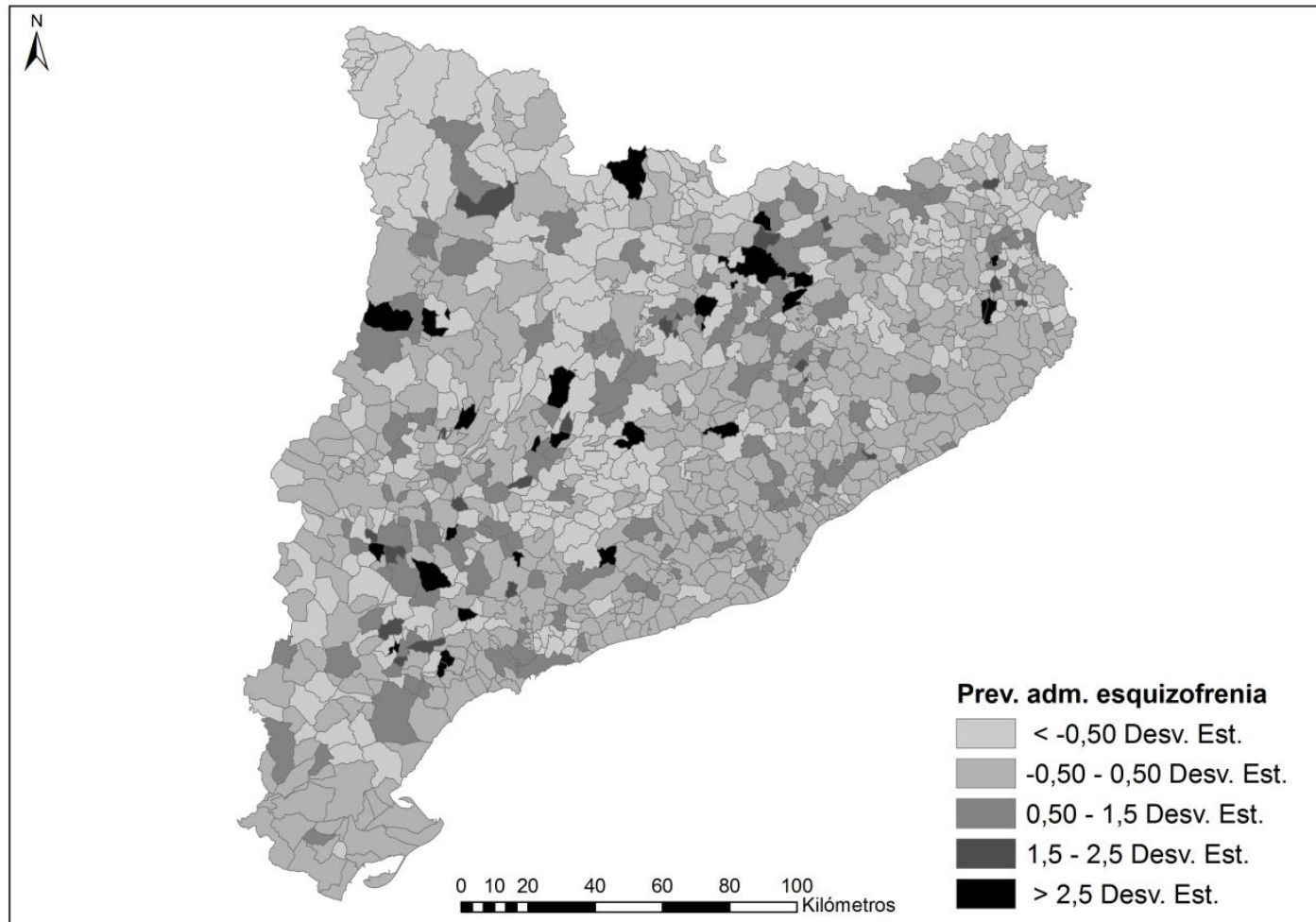
Tabla 37: Análisis de la autocorrelación espacial global de la prevalencia asistida de la depresión en la serie temporal (Índices globales de la *I* de Moran y la *G* de Getis y Ord)

	2009
<i>I</i> de Moran global	Existe agrupación espacial de la prevalencia  <i>I</i> = 0,01 Valor <i>z</i> = 5,4 d.e. Valor <i>p</i> = 0,01
<i>G</i> de Getis y Ord global	Agrupaciones de valores altos de prevalencia  <i>G</i> = 0 Valor <i>z</i> = 5,91 d.e. Valor <i>p</i> = 0,01
d.e.= desviación estándar.	

Fuente: elaboración propia

Finalmente, en el Mapa 29 se presentan los valores de prevalencia administrativa de la esquizofrenia por municipios catalanes.

Mapa 29: Prevalencia administrativa de la esquizofrenia en Centros de Salud Mental de Cataluña



Fuente: CMBD-SMA, 2009



### 4.3.1.2. *Hot-spot* y *cold-spot* de prevalencia administrativa de la esquizofrenia

#### Descripción estadística

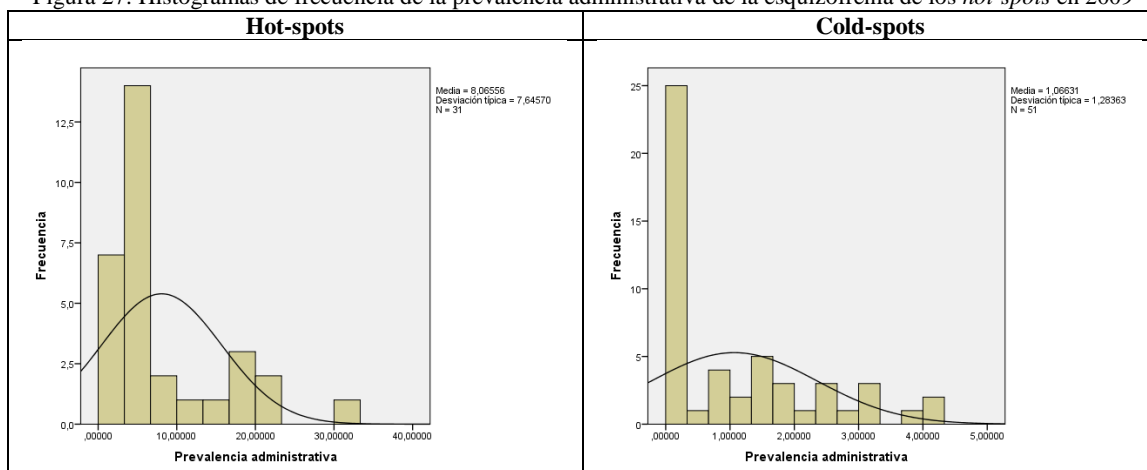
El AEMO-AE ha señalado la existencia de varios municipios que componen *hot-spots* o *cold-spots* de prevalencia administrativa de la esquizofrenia en Cataluña en el año 2009. Las unidades espaciales o municipios identificados como *hot-spots* son 31, con una media de prevalencia de 8,07 por 1.000 y una amplia desviación estándar como se observa en la Tabla 38. Los *cold-spots* se componen de 51 unidades espaciales con una media de 1,07 y una desviación estándar de 1,28. Los histogramas de ambos tipos de *spots* se muestran en la Figura 27.

Tabla 38: Estadísticas descriptivas de los *hot-spots* y *cold-spots* de esquizofrenia (prevalencia administrativa de la esquizofrenia)

	<i>Hot-spots</i>	<i>Cold-spots</i>
Recuento	31	51
Media	8,07	1,07
Desviación estándar	7,65	1,28
Mediana	4,64	0,63
Mínimo	1,19	0
Máximo	31,14	4,21

Fuente: Elaboración propia

Figura 27: Histogramas de frecuencia de la prevalencia administrativa de la esquizofrenia de los *hot-spots* en 2009



Fuente: Elaboración propia

Los municipios identificados por el algoritmo se han agrupado atendiendo a su vecindad y proximidad en diversos *hot-spots* (HSe) y *cold-spots* (CSe) numerados. En la Tabla 39 y la Tabla 40 se muestran los municipios que forman parte de cada uno y también se aportan la provincia y el área de salud mental donde se localizan.

Tabla 39: Municipios que forman parte de los *hot-spots* de esquizofrenia en 2009

HSe	Municipio	Prov.	Área SM
HSe1	Llosses (Les)	Girona	Ripollès
HSe1	Lluçà	Barcelona	Osona
HSe1	Orís	Barcelona	Osona
HSe1	Sant Boi de Lluçanès	Barcelona	Osona
HSe1	Santa Maria de Besora	Barcelona	Osona
HSe2	Aiguafreda	Barcelona	Vallès Oriental
HSe2	Artés	Barcelona	Bages
HSe2	Avinyó	Barcelona	Bages
HSe2	Balenyà	Barcelona	Osona
HSe2	Calldetenes	Barcelona	Osona
HSe2	Collsuspina	Barcelona	Osona
HSe2	Malla	Barcelona	Osona
HSe2	Seva	Barcelona	Osona
HSe2	Talamanca	Barcelona	Bages
HSe2	Taradell	Barcelona	Osona
HSe3	Caldes de Montbui	Barcelona	Vallès Oriental
HSe3	Canovelles	Barcelona	Vallès Oriental
HSe3	Franqueses del Vallès (Les)	Barcelona	Vallès Oriental
HSe3	Polinyà	Barcelona	Sabadell
HSe4	Arenys de Mar	Barcelona	Maresme Nord
HSe4	Arenys de Munt	Barcelona	Maresme Nord
HSe4	Calella	Barcelona	Maresme Nord
HSe5	Aguilar de Segarra	Barcelona	Bages
HSe6	Cabanyes (Les)	Barcelona	Alt Penedès
HSe6	Castellet i la Gornal	Barcelona	El Vendrell
HSe6	Castellví de la Marca	Barcelona	Alt Penedès
HSe6	Pontons	Barcelona	Alt Penedès
HSe6	Santa Margarida i els Monjos	Barcelona	Alt Penedès
HSe7	Cogul (El)	Lleida	Borges Blanques
HSe7	Pobla de Cérvoles (La)	Lleida	Borges Blanques
HSe8	Ascó	Tarragona	Mora d'Ebre

Fuente: Elaboración propia

Tabla 40: Municipios que forman parte de los *cold-spots* de esquizofrenia en 2009

CSe	Municipio	Prov.	Área SM
CSe1	Vilamòs	Lleida	Vielha
CSe2	Colera	Girona	Alt Empordà

<b>CSe</b>	<b>Municipio</b>	<b>Prov.</b>	<b>Àrea SM</b>
CSe2	Espolla	Girona	Alt Empordà
CSe2	Rabós	Girona	Alt Empordà
CSe3	Canet d'Adri	Girona	Gironès
CSe3	Tossa de Mar	Girona	Selva Marítima
CSe3	Vallfogona de Ripollès	Girona	Ripollès
CSe4	Alpens	Barcelona	Osona
CSe4	Castell de l'Areny	Barcelona	Berga
CSe4	Montmajor	Barcelona	Bages
CSe4	Quar (La)	Barcelona	Berga
CSe4	Vallcebre	Barcelona	Berga
CSe4	Viver i Serrateix	Barcelona	Bages
CSe5	Ametlla del Vallès (L')	Barcelona	Vallès Oriental
CSe5	Bigues i Riells	Barcelona	Vallès Oriental
CSe5	Brull (El)	Barcelona	Osona
CSe5	Cànoves i Samalús	Barcelona	Vallès Oriental
CSe5	Cardedeu	Barcelona	Vallès Oriental
CSe5	Garriga (La)	Barcelona	Vallès Oriental
CSe5	Riells i Viabrea	Girona	La Selva
CSe5	Sant Feliu de Codines	Barcelona	Vallès Oriental
CSe5	Sant Pere de Vilamajor	Barcelona	Vallès Oriental
CSe5	Sant Quirze Safaja	Barcelona	Osona
CSe5	Santa Eulàlia de Ronçana	Barcelona	Vallès Oriental
CSe6	Alella	Barcelona	Maresme Sud
CSe6	Argentona	Barcelona	Maresme Centre
CSe6	Cabrils	Barcelona	Maresme Sud
CSe6	Dosrius	Barcelona	Maresme Centre
CSe6	Martorelles	Barcelona	Mollet
CSe6	Montcada i Reixac	Barcelona	Cerdanyola
CSe6	Montornès del Vallès	Barcelona	Vallès Oriental
CSe6	Òrrius	Barcelona	Maresme Centre
CSe6	Premià de Mar	Barcelona	Maresme Sud
CSe7	Argençola	Barcelona	Anoia
CSe7	Bellprat	Barcelona	Anoia
CSe7	Cabrera d'Igualada	Barcelona	Alt Penedès
CSe7	Capellades	Barcelona	Anoia
CSe7	Carme	Barcelona	Anoia
CSe7	Copons	Barcelona	Anoia
CSe7	Jorba	Barcelona	Anoia
CSe7	Òdena	Barcelona	Anoia
CSe7	Olesa de Bonesvalls	Barcelona	Alt Penedès
CSe7	Orpí	Barcelona	Anoia
CSe7	Querol	Tarragona	Valls
CSe7	Savallà del Comtat	Tarragona	Anoia
CSe8	Alcoletge	Lleida	Lleida
CSe8	Alfés	Lleida	Lleida

CSe	Municipio	Prov.	Área SM
CSe8	Freginals	Tarragona	Amposta
CSe8	Granyena de les Garrigues	Lleida	Borges Blanques
CSe8	Maldà	Lleida	Mollerussa
CSe8	Morera de Montsant (La)	Tarragona	Reus

Fuente: Elaboración propia

Las pruebas estadísticas para la comparación de las distribuciones y medianas de los valores de prevalencia entre los diferentes *hot-spots* y *cold-spots* rechazan que éstas sean las iguales en todos los casos ( $p < 0,05$ ) (Tabla 41).

Tabla 41: Comparación de las distribuciones y medianas de las unidades espaciales que forman los diferentes *hot-spots* y *cold-spots* de esquizofrenia y aquellas que no forman parte de ninguno

	Prueba	Hipótesis nula	2009
Prevalencia administrativa	Prueba Kruskal-Wallis de muestras independientes	H <sub>0</sub> : la distribución de la prevalencia es la misma en las unidades espaciales identificadas como hot-spots (HSe1, HSe2, HSe3, HSe4, HSe5, HSe6, HSe7 y HSe8) y las que no lo han sido.	<b>Rechazar</b> (sig. 0,00 < 0,05)
		H <sub>0</sub> : la distribución de la prevalencia es la misma en las unidades espaciales identificadas como hot-spots (CSe1, CSe2, CSe3, CSe4, CSe5, CSe6, CSe7 y CSe8) y las que no lo han sido.	<b>Rechazar</b> (sig. 0,00 < 0,05)
	Prueba de medianas de muestras independientes	H <sub>0</sub> : la mediana de la distribución de la prevalencia es la misma en las unidades espaciales identificadas como hot-spots (HSe1, HSe2, HSe3, HSe4, HSe5, HSe6, HSe7 y HSe8) y las que no lo han sido.	<b>Rechazar</b> (sig. 0,00 < 0,009)
		H <sub>0</sub> : la mediana de la distribución de la prevalencia es la misma en las unidades espaciales identificadas como hot-spots (CSe1, CSe2, CSe3, CSe4, CSe5, CSe6, CSe7 y CSe8) y las que no lo han sido.	<b>Rechazar</b> (sig. 0,00 < 0,001)

Fuente: Elaboración propia

### Descripción de los *hot-spots* y *cold-spots*

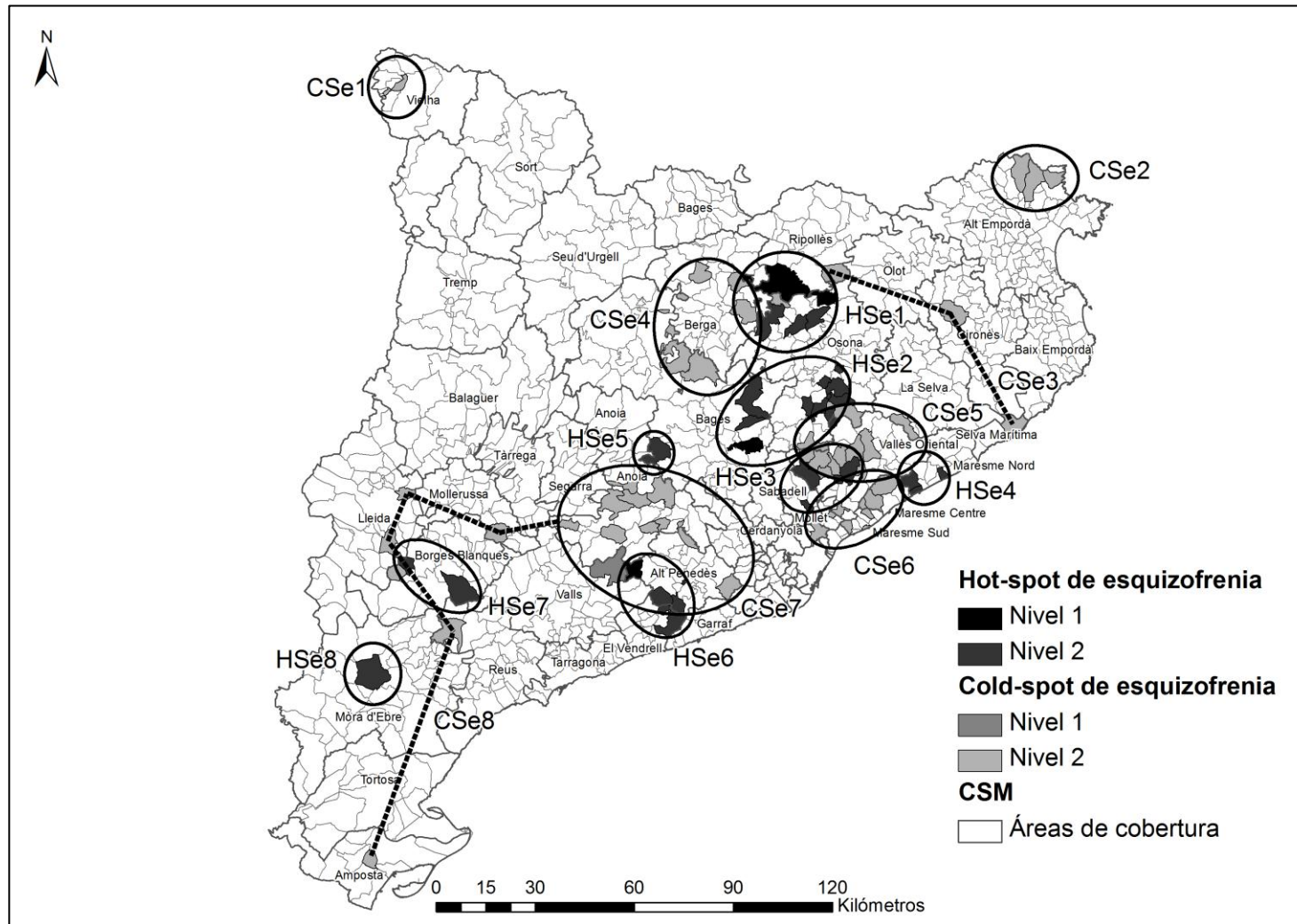
Las áreas de salud mental que acogen a municipios que forman parte de *hot-spots* de esquizofrenia son 10, mientras que para el caso de los *cold-spots* ascienden a 22. En el Mapa 30 se representan los *hot-spot* y *cold-spots* de prevalencia administrativa de la esquizofrenia en los que se pueden apreciar ocho concentraciones de diferente tamaño. Una de las concentraciones

de *hot-spots* se ubica entre las áreas de salud mental de Ripollès y Osona (HSe1) en un área limítrofe entre las provincias de Barcelona y Girona. El segundo *hot-spot* se encuentra al sur del anterior (HSe2), acoge a municipios de las áreas de Bages, Osona y Vallès Oriental. El tercer *hot-spot* se localiza en las áreas del Vallès Oriental y Sabadell 2 (HSe3). El cuarto está ubicado en el área de salud mental del Maresme Nord (HSe4). El quinto es un municipio aislado en el área de Bages (HSe5). Otro clúster se encuentra en las áreas del Alt Penedés y El Vendrell limítrofes a la provincia de Tarragona (HSe6). En la provincia de Lleida aparece otro *hot-spot* en el área de Boges Blanques (HSe7). Por último, hay un municipio que aparece de forma aislada en el área de Morà d'Ebre en Tarragona (HSe8).

Los *cold-spots* se han agrupado en 8 clústeres. En los Pirineos existen dos *cold-spots* que se sitúan cada uno en un extremo de la Comunidad. El primero en la zona occidental en el área de Vielha en el Valle de Arán (Lleida) (CSe1) y el oriental en el área de Alt Empordà de Girona (CSe2). En el área central de la Cataluña se localizan varios *cold-spots*, uno en las áreas de Bages y Berga (CSe4), otro en el Vallès Oriental con algún municipio en Osona y La Selva (CSe5); y un tercero que se localiza principalmente en la comarca del Maresme compuesto por municipios de las áreas del Maresme Sud y Centre junto a municipios de las áreas limítrofes de Cerdanyola, Mollet y Vallès Oriental. Otro *cold-spot* se encuentra formado principalmente por municipios del área de Anoia (CSe7) junto a otros próximos de las áreas de Valls y Alt Penedès. Existen varios municipios aislados identificados como *cold-spots* que se han agrupado en un clúster lineal (CSe3 y CSe8) porque podrían rodear y delimitar *hot-spots*. Así, el CSe8 parece delimitar al HSe7 y separarlo del HSe8, y el CSe3 parece constituir un *cold-spot* periférico entorno a la concentración central de clústeres.

En la localización de los *hot-spots* y *cold-spot* se aprecia la existencia de relaciones de vecindad, de hecho ciertos *cold-spots* se localizan junto a *hot-spots*. Algunos de ellos se han seguido para dividir concentraciones de *hot-spots* que podrían haber sido considerados como únicas como es el caso del CSe5 y CSe6 que dividen HSe2, HSe3 y HSe4.

Mapa 30: Hot-spots y cold-spots de prevalencia administrativa de la esquizofrenia en Cataluña (2009)



Fuente: Elaboración propia a partir del CMBD-SMA, 2009

#### 4.3.2. Caso 4: *Hot-spots* y *cold-spots* de prevalencia asistida de la depresión en Cataluña en 2009

##### 4.3.2.1. Distribución geográfica de la prevalencia administrativa de la depresión

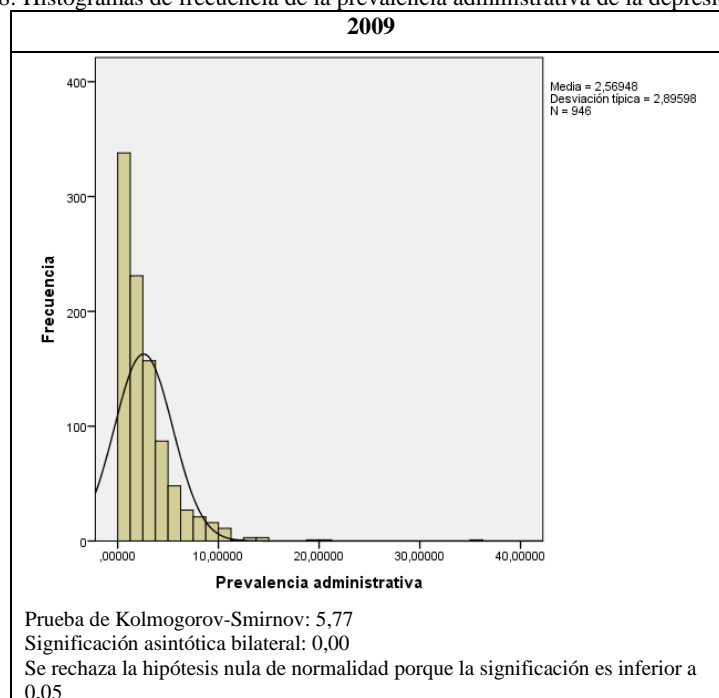
La prevalencia administrativa de la depresión en Cataluña para el año de estudio tuvo una media de 2,57 por 1.000 habitantes, una desviación típica de 2,9 y una mediana de 1,99. Existen 242 unidades espaciales con valores de prevalencia 0, mientras que el valor máximo es de 35,8 por 1.000 (Tabla 42). El histograma de frecuencias de la prevalencia de la depresión se muestra en la Figura 28. La distribución no se ajusta a la curva normal como indica la prueba de Kolmogorov-Smirnov ( $p < 0,05$ ).

Tabla 42: Estadísticas descriptivas de la prevalencia administrativa de la depresión en Cataluña

	2009
Recuento	946
Media	2,57
Desviación estándar	2,90
Mediana	1,99
Mínimo	0 (242 UE)
Máximo	35,80

Fuente: CMBD-SMA, 2009

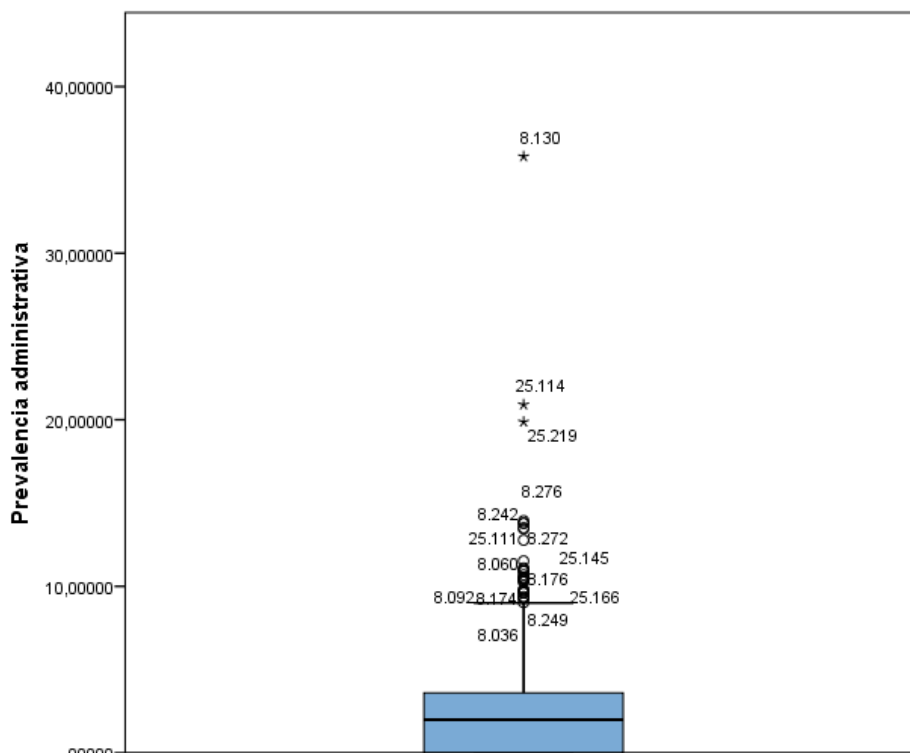
Figura 28: Histogramas de frecuencia de la prevalencia administrativa de la depresión en 2009



Fuente: CMBS-SMA, 2009

El *box-plot* de la prevalencia de la depresión (Figura 29) muestra la dispersión de los valores más elevados de prevalencia. De nuevo aparecen valores atípicos y extremos en varias unidades espaciales siempre en el extremo derecho de la distribución (valores altos).

Figura 29: *Box-plot* de las poblaciones de prevalencia administrativa de la depresión en 2009 (los números identificativos de los outliers corresponden al código INE municipal)



Fuente: CMBS-SMA, 2009

En la Tabla 43 se expone la clasificación de los valores municipales de prevalencia por intervalos de frecuencia en base a la desviación estándar. El límite por la izquierda de la distribución será siempre 0, mientras que por la derecha la existencia de valores atípicos o extremos se materializa en una mayor dispersión. En este caso, hay 49 unidades espaciales con valores entre 1,5 y 2,5 veces la desviación típica desde la media, y por encima de 2,5 veces existen 22 unidades.



Tabla 43: Distribución de frecuencias de la prevalencia administrativa de la depresión en Cataluña

	2009
-1,5 a -0,5 Desv. Est.	318
-0,5 a 0,5 Desv. Est.	435
0,5 a 1,5 Desv. Est.	122
1,5 a 2,5 Desv. Est.	49
>2,5 Desv. Est.	22

Fuente: CMBS-SMA, 2009

La existencia de autocorrelación espacial en la prevalencia administrativa de la depresión queda avalada por los índices globales de análisis espacial con un valor de significación del 0,01. Por una parte, la *I* de Moran certifica la existencia de dependencia espacial y, por su parte, la *G* de Getis y Ord señala que ésta es de valores altos de prevalencia aunque, como se verá en el próximo epígrafe, el AEMO/AE ha localizado también concentraciones de valores bajos.

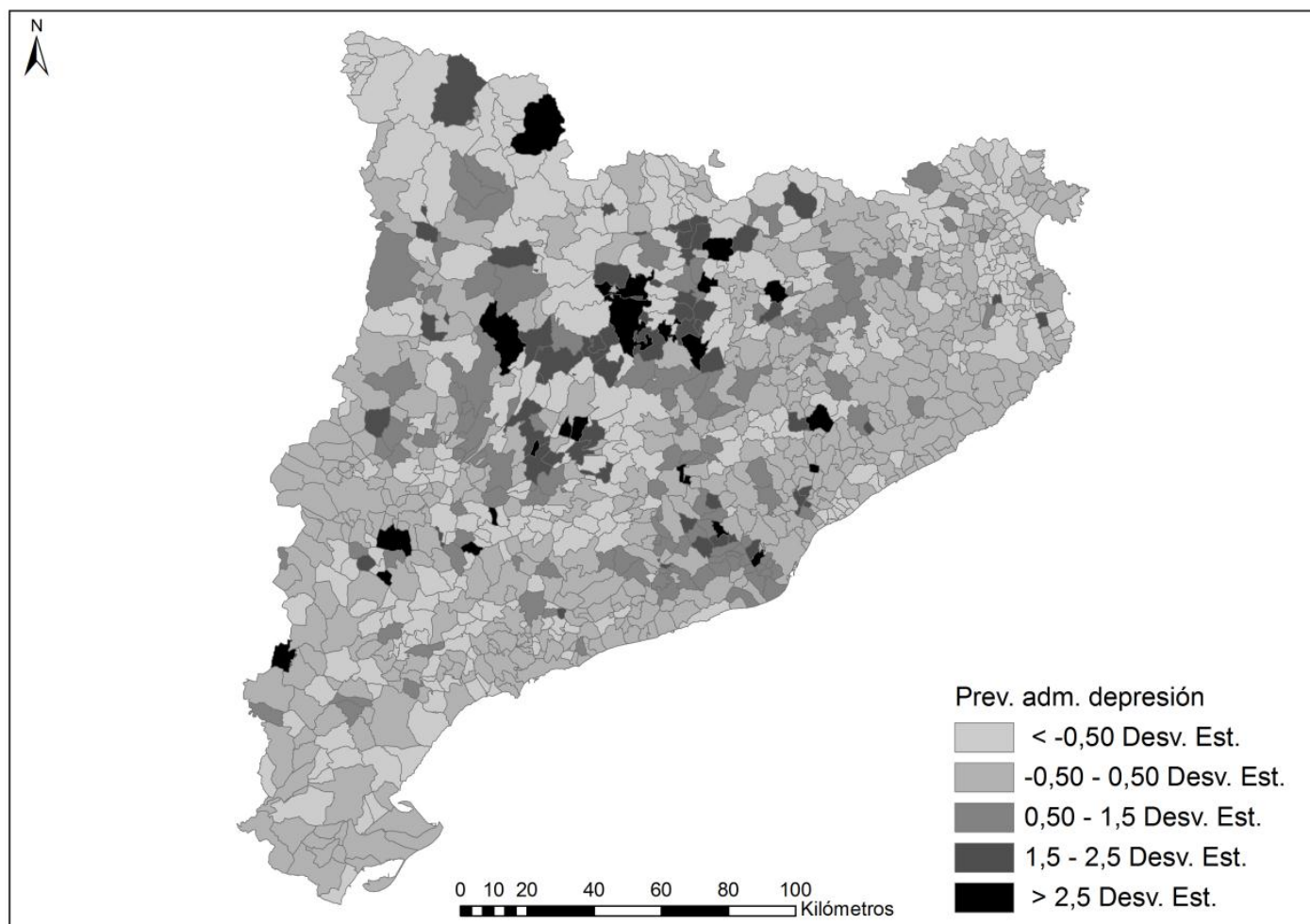
Tabla 44: Análisis de la autocorrelación espacial global de la prevalencia asistida de la depresión en la serie temporal (Índices globales de la *I* de Moran y la *G* de Getis y Ord)

	2009
<i>I</i> de Moran global	Existe agrupación espacial de la prevalencia  $I = 0,04$ Valor $z = 23,79$ d.e. Valor $p = 0,01$
<i>G</i> de Getis y Ord global	Agrupaciones de valores altos de prevalencia  $G = 0$ Valor $z = 14,32$ d.e. Valor $p = 0,01$
d.e.=desviación estándar.	

Fuente: Elaboración propia

Finalmente, la prevalencia administrativa municipal de la depresión se ha representado en el Mapa 31 donde se muestra distribuida por intervalos en base a la desviación estándar del conjunto de la población analizada.

Mapa 31: Prevalencia administrativa de la depresión en Centros de Salud Mental de Cataluña



Fuente: CMBD-SMA, 2009

### 4.3.2.2. Hot-spot y cold-spot de prevalencia administrativa de la depresión

#### Descripción estadística

Los *hot-spots* se localizan en 13 de las 74 áreas de salud mental de Cataluña. El AEMO/AE ha localizado 39 municipios que constituyen *hot-spots*. Los municipios *hot-spots* presentan una media de 8,16 casos por 1.000 habitantes, su mediana es 7,85 y su desviación típica es de 6,49. El valor más elevado es 35,80 y el más pequeño es de 1,74 por 1.000.

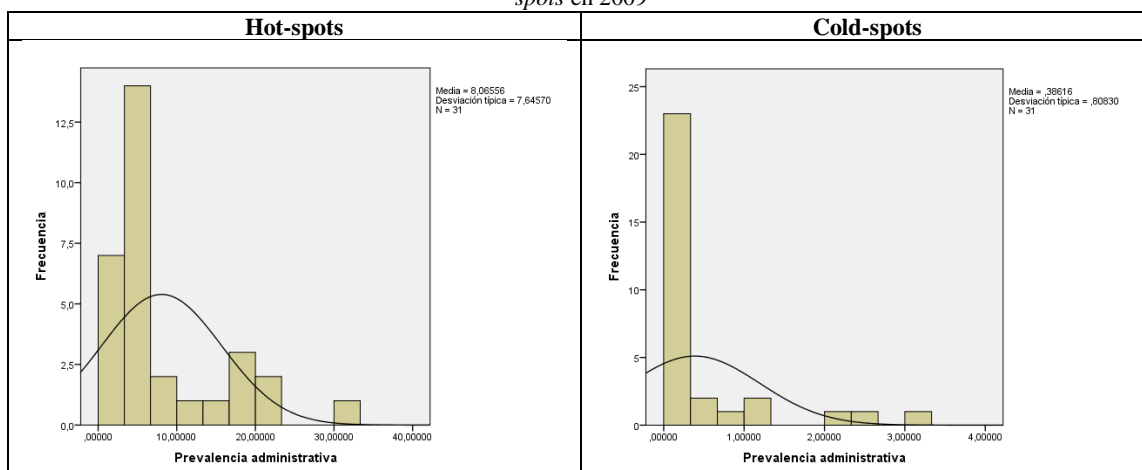
Los municipios identificados como *cold-spots* son 31 y se encuentran en 16 áreas de salud mental (Tabla 45 y Figura 30). Tienen una media de 0,39 pacientes por 1.000 habitantes, la mediana se encuentra en el valor 0 y la desviación estándar es de 0,81. El valor mínimo es de 0 casos y el máximo es 3,01 casos por 1.000 habitantes.

Tabla 45: Estadísticas descriptivas de los *hot-spots* de depresión (prevalencia administrativa de la depresión)

	Hot-spots	Cold-spots
Recuento	39	31
Media	8,16	0,39
Desviación estándar	6,49	0,81
Mediana	7,85	0
Mínimo	1,74	0
Máximo	35,80	3,01

Fuente: Elaboración propia

Figura 30: Histogramas de frecuencia de la prevalencia administrativa de la esquizofrenia de los *hot-spots* y *cold-spots* en 2009



Fuente: CMBD-SMA, 2009

Los municipios identificados por el AEMO/AE se han agrupado en 6 grupos de *hot-spots* (HSd) en base a su proximidad geográfica y de cara a facilitar su descripción (Tabla 46). Junto a éstos, han sido definidos 4 grupos de *cold-spots* de depresión (CSd) que se presentan en la Tabla 47.

Tabla 46: Municipios que forman parte de los *hot-spots* de depresión en 2009

HSd	Municipio	Prov.	Área SM
HSd1	Cabó	Lleida	Seu d'Urgell
HSd1	Guixers	Lleida	Bages
HSd1	Montmajor	Barcelona	Bages
HSd2	Avià	Barcelona	Berga
HSd2	Berga	Barcelona	Berga
HSd2	Casserres	Barcelona	Berga
HSd2	Gironella	Barcelona	Berga
HSd2	Montclar	Barcelona	Bages
HSd2	Prats de Lluçanès	Barcelona	Osona
HSd2	Riner	Lleida	Bages
HSd2	Sora	Barcelona	Osona
HSd3	Calaf	Barcelona	Anoia
HSd3	Calonge de Segarra	Barcelona	Anoia
HSd3	Castellfollit de Riubregós	Barcelona	Anoia
HSd3	Ivorra	Lleida	Segarra
HSd3	Pujalt	Barcelona	Anoia
HSd3	Tarroja de Segarra	Lleida	Segarra
HSd3	Torrefeta i Florejacs	Lleida	Segarra
HSd4	Aiguafreda	Barcelona	Vallès Oriental
HSd4	Ametlla del Vallès (L')	Barcelona	Vallès Oriental
HSd4	Bigues i Riells	Barcelona	Vallès Oriental
HSd4	Caldes de Montbui	Barcelona	Vallès Oriental
HSd4	Campins	Barcelona	Vallès Oriental
HSd4	Canovelles	Barcelona	Vallès Oriental
HSd4	Cànoves i Samalús	Barcelona	Vallès Oriental
HSd4	Franqueses del Vallès (Les)	Barcelona	Vallès Oriental
HSd4	Granollers	Barcelona	Vallès Oriental
HSd4	Lliçà de Vall	Barcelona	Vallès Oriental
HSd4	Sant Martí de Centelles	Barcelona	Osona
HSd5	Avinyonet del Penedès	Barcelona	Alt Penedès
HSd5	Begues	Barcelona	Gavà
HSd5	Canyelles	Barcelona	Garraf
HSd5	Castellví de Rosanes	Barcelona	Martorell
HSd5	Cervelló	Barcelona	Sant Feliu
HSd5	Corbera de Llobregat	Barcelona	Sant Feliu
HSd5	Gelida	Barcelona	Martorell
HSd5	Olesa de Bonesvalls	Barcelona	Alt Penedès

HSd	Municipio	Prov.	Àrea SM
HSd5	Olivella	Barcelona	Garraf
HSd6	Torns (Els)	Lleida	Borges Blanques

Fuente: Elaboración propia

Tabla 47: Municipios que forman parte de los *cold-spots* de depresión en 2009

CSd	Municipio	Prov.	Àrea SM
CSd1	Albons	Girona	Alt Empordà
CSd1	Colomers	Girona	Gironès
CSd1	Palau de Santa Eulàlia	Girona	Alt Empordà
CSd1	Riumors	Girona	Alt Empordà
CSd1	Sant Mori	Girona	Alt Empordà
CSd1	Viladasens	Girona	Gironès
CSd2	Alella	Barcelona	Maresme Sud
CSd2	Arenys de Mar	Barcelona	Maresme Nord
CSd2	Arenys de Munt	Barcelona	Maresme Nord
CSd2	Argentona	Barcelona	Maresme Centre
CSd2	Cabrils	Barcelona	Maresme Sud
CSd2	Dosrius	Barcelona	Maresme Centre
CSd2	Òrrius	Barcelona	Maresme Centre
CSd2	Santa Maria de Martorelles	Barcelona	Mollet
CSd3	Morera de Montsant (La)	Tarragona	Reus
CSd3	Perafort	Tarragona	Tarragona Nord
CSd3	Pradell de la Teixeta	Tarragona	Mora d'Ebre
CSd3	Vilanova d'Escornalbou	Tarragona	Reus
CSd4	Alòs de Balaguer	Lleida	Balaguer
CSd4	Argençola	Barcelona	Anoia
CSd4	Bellprat	Barcelona	Anoia
CSd4	Biosca	Lleida	Anoia
CSd4	Capolat	Barcelona	Berga
CSd4	Castellar del Riu	Barcelona	Berga
CSd4	Castellfollit del Boix	Barcelona	Bages
CSd4	Das	Girona	Bages
CSd4	Ivars de Noguera	Lleida	Lleida
CSd4	Olost	Barcelona	Osona
CSd4	Pinós	Lleida	Anoia
CSd4	Sanauja	Lleida	Balaguer
CSd4	Vall de Cardós	Lleida	Sort

Fuente: Elaboración propia

La comparación estadística de las distribuciones y las medianas entre los diferentes grupos de *hot-spots* y *cold-spots* y el resto de las unidades espaciales ha dado como resultado rechazar la

hipótesis nula de igualdad entre las mismas. Tanto la Prueba *U* como la prueba de las medianas coinciden en rechazar la hipótesis nula con un nivel de significación de 0,05 (Tabla 48).

Tabla 48: Comparación de las distribuciones y medianas de las unidades espaciales que forman los diferentes *hot-spots* de depresión y aquellas que no forman parte de ninguno

	<b>Prueba</b>	<b>Hipótesis nula</b>	<b>2009</b>
Prevalencia administrativa	Prueba Kruskal-Wallis de muestras independientes	H <sub>0</sub> : la distribución de la prevalencia es la misma en las <i>unidades espaciales</i> identificadas como <i>hot-spots</i> (HSd1, HSd2, HSd3, HSd4, HSd5 y HSd6) y las que no lo han sido.	<b>Rechazar</b> (sig. 0,00 < 0,05)
		H <sub>0</sub> : la distribución de la prevalencia es la misma en las <i>unidades espaciales</i> identificadas como <i>cold-spots</i> (CSd1, CSd2, CSd3 y CSd4) y las que no lo han sido.	<b>Rechazar</b> (sig. 0,00 < 0,05)
	Prueba de medianas de muestras independientes	H <sub>0</sub> : la mediana de la distribución de la prevalencia es la misma en las <i>unidades espaciales</i> identificadas como <i>hot-spots</i> (HSd1, HSd2, HSd3, HSd4, HSd5 y HSd6) y las que no lo han sido.	<b>Rechazar</b> (sig. 0,00 < 0,05)
		H <sub>0</sub> : la mediana de la distribución de la prevalencia es la misma en las <i>unidades espaciales</i> identificadas como <i>cold-spots</i> (CSd1, CSd2, CSd3 y CSd4) y las que no lo han sido.	<b>Rechazar</b> (sig. 0,00 < 0,05)

Fuente: Elaboración propia

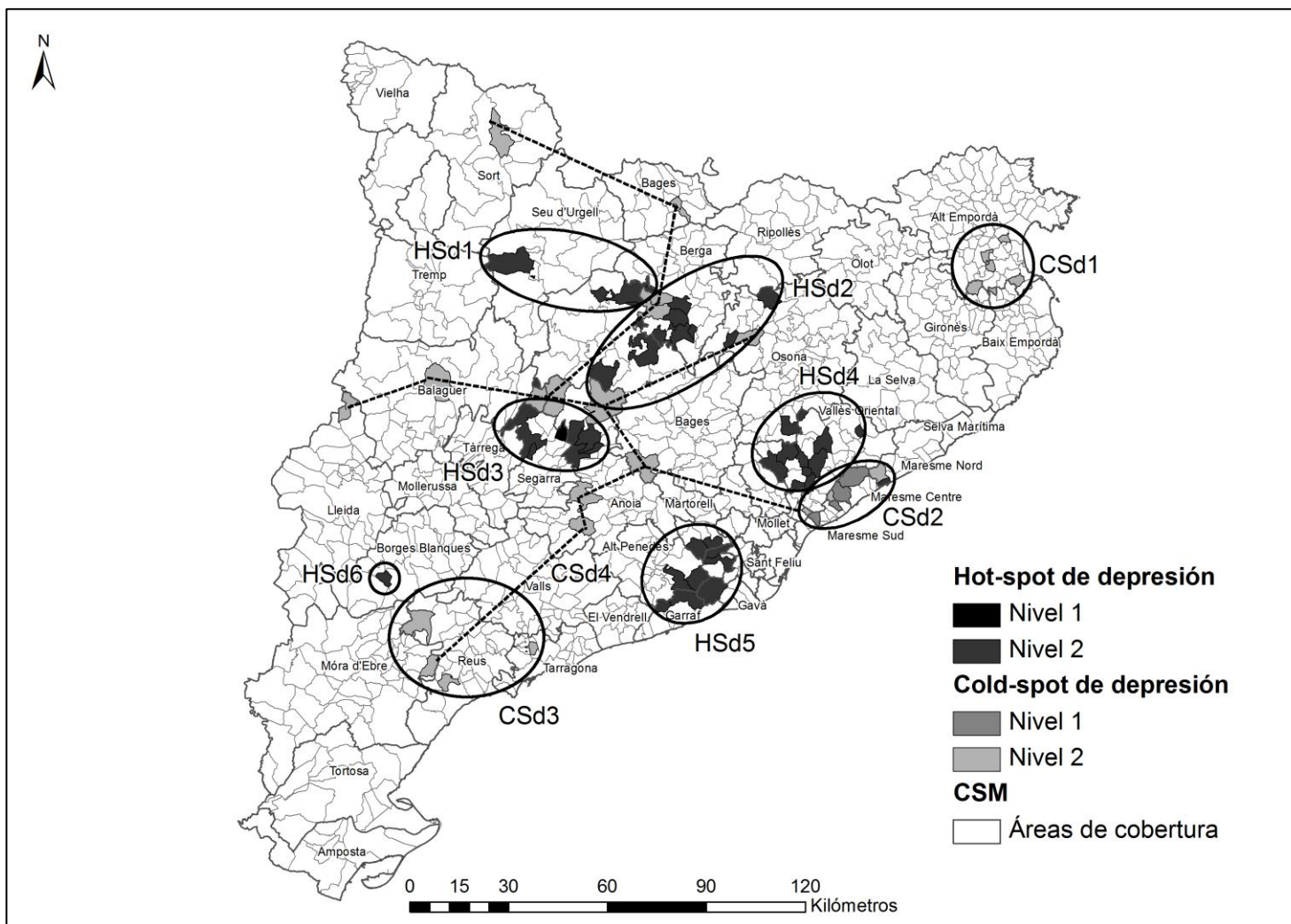
### Descripción de los *hot-spots* y *cold-spots*

El AEMO/AE ha localizado varios municipios que componen *hot-spots* en el centro de la región (Mapa 32). Siguiendo la configuración espacial de los *cold-spots* se ha dividido esta concentración en tres. El HSd1 se sitúa en las áreas de salud mental de Seu d'Urgell (Lleida) y Bages (Barcelona). El área de Bages también se relaciona con la de Berga (HSd2). Y la Segarra (Lleida) con Anoia (Barcelona) en el HSd3. Otro *hot-spot* se encuentra en el área del Vallès Oriental fundamentalmente aunque también acoge municipios de Osona. El quinto *hot-spot* se halla entre diversas áreas de salud mental, concretamente Alt Penedès, Gavà, Garraf, Martorell y Sant Feliu (HSd5). Por último, aparece un municipio aislado identificado como *hot-spot* en el área de Borges Blanques en la provincia de Lleida (HSd6).

Un *cold-spot* se ubica en la provincia de Girona entre las áreas de salud mental de Alt Empordà y Gironès (CSd1). En la comarca del Maresme existe otro *cold-spot* cuyos municipios son atendidos en tres áreas diferentes: Nord, Centre y Sud (CSd2). El tercer *cold-spot* se encuentra

en la provincia de Tarragona atendido por los CSM de Tarragona Nord, Reud y Morà d'Ebre (CSd3). Por último, el resto de *cold-spots* corresponden a municipios aislados que podrían formar un clúster lineal. Éste delimita varias concentraciones de *hot-spots* que podrían haber sido consideradas como una única como los mencionados HSd1, Hsd2 y Hsd3. La relación entre los *hot* y *cold-spots* también se aprecia entre Hsd4 y CSd2 que se encuentran en dos comarcas vecinas como son el Maresme y el Vallès Oriental.

Mapa 32: Hot-spots y cold-spots de prevalencia administrativa de la depresión en Cataluña (2009)



Fuente: Elaboración propia a partir del CMBD-SMA, 2009





## V. DISCUSIÓN

---

### 5.1. El modelo AEMO/AE

En este trabajo se ha diseñado e implementado una nueva herramienta de análisis espacial de datos basada en métodos de inteligencia artificial. El Algoritmo Evolutivo Multi-objetivo de Análisis Espacial (AEMO/AE) ha sido aplicado con éxito en la optimización de problemas multi-objetivos de tipo geográfico tal como se planteó en la primera hipótesis.

La potencia de los AEMO es tal que son capaces de encontrar soluciones óptimas a problemas de difícil resolución, pero en ocasiones pueden dar soluciones que, aunque factibles, no son adecuadas para el problema que se pretende resolver. El punto de partida de estos algoritmos es el planteamiento del problema multi-objetivo basado en la optimización de varios objetivos, para ello debe mantener un equilibrio entre los mismos en el que ningún objetivo sea perjudicado por la mejora de otro (óptimo de Pareto). Es fundamental plantear de forma correcta e inequívoca el problema para evitar resultados no deseados, identificando los objetivos y diseñando cuidadosamente los parámetros de funcionamiento del algoritmo y las *funciones de aptitud* que evalúen correctamente la calidad de la soluciones.

Las estadísticas descriptivas han mostrado que existen municipios dentro de los *hot-spots* y *cold-spots* que son, en principio, incoherentes con el razonamiento del problema multi-objetivo. La causa se encuentra en el diseño del algoritmo que restringe a 10 el número de unidades espaciales que componen cada solución. El algoritmo siempre buscará completar cada solución con unidades espaciales que no comprometan su aptitud asegurando su óptimo desde el punto de vista de Pareto. En los Estudios de Demostración esta restricción se ha manifestado en dos tipos de resultados no deseados. Por una parte, la existencia de unidades espaciales con valores no buscados en cuanto a media y/o desviación estándar y, por otra, la aparición de unidades espaciales aisladas.

En el primer caso, las unidades espaciales con valores incoherentes se integran en los clústeres espaciales porque ha prevalecido la minimización de la distancia relativa sobre la maximización/minimización de la media y/o la desviación estándar de la prevalencia o los índices espaciales. Este caso, al menos desde el punto de vista visual, no introduce grandes distorsiones en los resultados ni en la interpretación de los mismos porque estas unidades se integran en los clústeres junto a otras soluciones coherentes.

En el segundo caso, aparecen unidades espaciales aisladas con valores óptimos de los índices espaciales o de la prevalencia pero lejanos geográficamente al resto. El algoritmo ha priorizado la optimización de los valores de la media y la desviación estándar sobre la mínima distancia relativa. En ambos casos, las distorsiones son resultado del funcionamiento del algoritmo que trata de equilibrar los objetivos por lo que aparecen soluciones que dan preferencia a unos objetivos sobre otros pero siempre manteniendo la máxima factibilidad y la igualdad en la *aptitud* de las soluciones para resolver el problema multi-objetivo.

Dada la naturaleza de las variables analizadas en el Estudio de Demostración A, índices de análisis espacial, la aparición de unidades espaciales aisladas no son del todo incoherentes a pesar de que el problema multi-objetivo planteaba la búsqueda de agrupaciones próximas de municipios. Los valores en los índices espaciales de un municipio aislado responden a las relaciones espaciales con los municipios de alrededor que se recogen en la matriz de distancias. Así, el *hot-spot* no es sólo resultado del municipio en cuestión sino también de sus relaciones con los municipios de su entorno. Este hecho es común a otros métodos de AED como los índices clásicos de autocorrelación espacial en los que también aparecen unidades espaciales aisladas a pesar de buscar concentraciones. En resumen, es importante considerar que la dependencia espacial no afecta a una única unidad espacial sino a todas aquellas con las que presenta interrelaciones. En el caso del Estudio de Demostración B, las unidades aisladas son claramente un resultado no deseado aunque señalan situaciones de interés en el territorio, indicando unidades espaciales con muy alta o baja prevalencia administrativa. Además, en el caso de los *cold-spots* parecen localizarse en torno los *hot-spots* ayudando a delimitarlos.

El mismo diseño del AEMO/AE ha tratado de reducir la aparición de ambas situaciones mediante la aplicación de tres métodos de selección sobre las soluciones del algoritmo (QQ-Plots y definición estadística del punto de corte). No todas las unidades espaciales que aparecen en las soluciones del algoritmo se consideran *hot-spots* y *cold-spots*. Como se ha explicado, los resultados finales se seleccionan a partir de la frecuencia con la que aparecen entre las soluciones del algoritmo. De esta forma, muchas de las unidades espaciales introducidas para completar los grupos de 10 unidades espaciales de cada solución se descartan en la fase de selección del AEMO/AE por haber aparecido una sola vez o de forma esporádica.

Una posible solución a este problema sería ajustar el número de unidades espaciales por solución en el diseño del AEMO/AE. Este número se ha fijado en 10 unidades espaciales buscando que no fuera un número tan bajo que dejase fuera posibles soluciones óptimas, pero tampoco tan alto que recogiese unidades espaciales con valores no deseados para completar una solución. No es indispensable predeterminar previamente el número de unidades espaciales que

compone cada solución pero se diseñó así para mejorar el rendimiento del algoritmo. Sería posible flexibilizar el diseño para que el algoritmo buscase soluciones factibles sin tener que completar este número o, si fuese necesario, pudiese añadir más unidades espaciales a cada solución. El problema de flexibilizar este número es que demandaría una gran potencia de computación y tiempo porque el algoritmo identificaría soluciones dentro de soluciones (de más pequeñas a más grandes) multiplicando el número de soluciones posibles.

La aparición de soluciones no deseadas no se debe a un problema del método porque estas soluciones efectivamente resuelven el problema multi-objetivo sino al diseño de los parámetros del AEMO/AE. En el futuro se podría indagar en restringir aún más la función objetivo de la distancia relativa o buscar algún sustituto. En todo caso, los resultados del AEMO/AE han marcado zonas con valores atípicos de prevalencia o dependencia espacial con un potencial interés para la planificación y gestión de los servicios de salud mental.

Otro aspecto en el diseño del AEMO/AE que puede introducir distorsiones en los resultados es la asignación de pesos a las *variables de decisión*. Las variables que han sido utilizadas en los Estudios de Demostración poseen distintas escalas. La media y la desviación estándar de la prevalencia se miden en casos por 1.000 habitantes; la media y la desviación estándar de los índices de dependencia espacial han sido re-escalados entre [0,1-0,9]; y la distancia relativa tiene un rango de [1-10]. El uso de diferentes escalas dificulta la asignación de pesos en el diseño de las *funciones de aptitud* y hace necesario fijar los mismos cuidadosamente. La asignación de pesos es uno de los elementos subjetivos y cuestionables en el diseño de los Algoritmos Evolutivos al depender del criterio del diseñador y/o los expertos. En este sentido, los pesos del AEMO/AE han sido ajustados buscando el equilibrio entre las funciones objetivo considerando sus diferentes escalas de medida.

La selección de las *funciones de aptitud* que evalúan las soluciones también es una decisión crítica. El AEMO/AE utiliza cuatro *funciones de aptitud* para evaluar la factibilidad de las soluciones. Los resultados de cada función son diferentes debido a sus características aunque sus resultados no suelen diferir demasiado. En principio, las soluciones proporcionadas por una *función de aptitud* no son mejores que las de las demás porque cada una aborda de forma diferente, pero con similar eficacia, la valoración de la calidad o eficiencia de la solución.

Los municipios seleccionados como *hot-spots* o *cold-spots* han sido agrupados por proximidad para su descripción. Esta agrupación se ha realizado de forma subjetiva dada la dificultad de delimitar las fronteras de los clústeres espaciales. Ésta es una de las líneas de investigación actual en el campo del AED (Jacquez y Goovaerts, 2010; Jacquez, Kaufmann, et al., 2008).

Entre estas investigaciones destaca la utilización de algoritmos multi-objetivos para optimizar funciones que penalizan la irregularidad geométrica y los resultados de una *scan spatial statistic* para obtener y delimitar clústeres espaciales (Cançado, Duarte, et al., 2010; Duczmal, Cançado, et al., 2007, 2008). La función de penalización de irregularidades geométricas sustituye a la mínima distancia relativa empleada en este trabajo, mientras que la proximidad de valores elevados y significativos se consigue mediante un método de análisis espacial. Podría ser interesante estudiar las funciones de penalización geométrica e incorporarlas a los dos tipos de análisis con los que se ha probado el AEMO/AE.

## 5.2. Estudios de Demostración

La segunda y tercera hipótesis sobre la utilidad del modelo AEMO/AE para la resolución de dos problemas específicos de análisis espacial también han sido demostradas. El modelo ha encontrado efectivamente una solución de consenso entre índices estadísticos de dependencia espacial identificando áreas con elevada dependencia espacial de la prevalencia administrativa de la esquizofrenia y la depresión (*hot-spots*) como se ha probado en el Estudio de Demostración A. De igual forma, la herramienta también ha tenido éxito en localizar concentraciones de valores significativamente elevados y bajos de prevalencia administrativa de la esquizofrenia y la depresión (*hot-spots* y *cold-spots*).

El análisis espacial exploratorio realizado sobre las series de prevalencia administrativa de esquizofrenia y depresión en Andalucía (Estudio de Demostración A) ha indicado la existencia de dependencia espacial en su distribución geográfica, y que ésta es principalmente de valores altos con un nivel de significación muy estricto ( $p=0.01$ ). Este análisis previo ha sido confirmado por los tres índices locales de asociación espacial y por el índice Bayesiano, que han mostrado las áreas geográficas concretas que presentan dependencia espacial de la prevalencia frente a otras en las que su distribución es aleatoria. Tal como se esperaba los resultados de los cuatro índices muestran claros desacuerdos en las áreas indicadas lo que dificulta la interpretación de los mismos (Moreno, García-Alonso, et al., 2008). Existen numerosos trabajos de investigación dedicados a comparar diferentes métodos y establecen una clasificación de su potencia o robustez como se ha señalado en la Introducción. La conclusión es siempre la misma, la mayor parte de los métodos son adecuados, eficaces y presentan sus ventajas e inconvenientes (Lawson, Biggeri, et al., 2000). Estos análisis previos justifican la utilidad del modelo híbrido que se propone en este trabajo como herramienta de consenso entre diferentes métodos geoestadísticos. Las soluciones finales aportadas por el modelo AEMO/AE han ofrecido efectivamente un resultado único fruto del consenso entre los cuatro métodos.

El análisis del Estudio de Demostración A se ha efectuado sobre nueve objetivos basados en cinco variables: cuatro índices de análisis espacial de datos, de los que se ha optimizado la media y la desviación estándar, y un índice de distancia relativa entre *unidades espaciales*. El modelo AEMO/AE es flexible y permite introducir nuevos objetivos y variables por lo que permitiría consensuar otros índices de análisis espacial de forma sencilla aunque con el coste de incrementar el tiempo de procesamiento y los requisitos computacionales.

El problema planteado en Andalucía consistía en encontrar concentraciones de municipios próximos en el espacio geográfico con una fuerte dependencia espacial de la esquizofrenia y la depresión asistida en los servicios ambulatorios de salud mental. El algoritmo ha señalado con certeza unidades espaciales próximas con un comportamiento homogéneo e interdependiente de la prevalencia que no es debido al azar. La existencia de *hot-spots* es una situación relevante porque marca situaciones atípicas en la localización geográfica de los datos, que se presume aleatoria, más aún si la variable se refiere a casos de enfermedades.

Se han realizado diversos análisis no paramétricos para la comparación de las distribuciones (Prueba *U*) y medianas de los índices espaciales y la prevalencia entre los municipios que constituyen *hot-spots* y el resto. Los resultados muestran que los *hot-spots* de esquizofrenia presentan un comportamiento de los índices espaciales y la prevalencia muy diferente entre sí excepto en las medianas de la *I* de Moran. En cambio, el análisis de los *hot-spots* de depresión ha dado un resultado diferente. En este caso, algunos índices espaciales diferencian a los *hot-spots* del resto y otros no. En cuanto a los valores de prevalencia, la prueba *U* distingue claramente a los *hot-spots* aunque la prueba de las medianas, más estricta, dice lo contrario. En conclusión, el algoritmo ha identificado grupos de unidades espaciales con un comportamiento diferente y atípico en sus valores de los índices espaciales y de la prevalencia ratificado por la mayor parte de las pruebas estadísticas. Aun así, no existe un acuerdo total entre los métodos de comparación y los *hot-spots* de esquizofrenia han mostrado diferencias más claras entre sí y el resto de unidades espaciales que los de depresión.

En cuanto a los resultados en sí, cabe señalar que el *hot-spot* de esquizofrenia localizado en Sierra Nevada ha sido confirmado en estudios previos por otras metodologías de análisis espacial (Moreno, García-Alonso, et al., 2008). También es interesante señalar el caso de la Sierra de Aracena que no ha sido identificado como *hot-spot* de esquizofrenia por el Algoritmo Evolutivo aunque albergaba a varios municipios con prevalencia elevada y para los que algunos indicadores de análisis espacial señalaban la existencia de autocorrelación espacial. Se podría explicar por la existencia de municipios con prevalencia de esquizofrenia significativamente baja en Aracena y el efecto frontera con Portugal y Extremadura que ha limitado el número de

municipios vecinos. Otra situación de interés es que se ha constatado que los *hot-spots* de diferentes diagnósticos no son excluyentes y coinciden territorialmente. Así, en el área oriental de Sierra Nevada en Granada parece que existe relación espacial entre la esquizofrenia y la depresión por la aparición de clústeres de ambos diagnósticos. Se ha constatado que con cierta frecuencia existe comorbilidad de ambos trastornos con la aparición de síntomas depresivos en pacientes con esquizofrenia (Campos y Martínez-Larrea, 2002; Zisook, McAdams, et al., 1999). Otra posible explicación sería un funcionamiento particular de los servicios de salud mental que atienden a estas áreas (variabilidad de la práctica médica).

El planteamiento del estudio de Cataluña es diferente aunque se hayan estudiado los mismos diagnósticos psiquiátricos. En este caso, el modelo se ejecuta directamente como técnica de análisis espacial sin que necesite los resultados previos de otros métodos. El problema se reduce a la optimización de la prevalencia (media) y la distancia relativa. Este análisis permite introducir como objetivo no sólo la maximización de la prevalencia sino también su minimización buscando *cold-spots* junto a los *hot-spots*. En todo caso, se trata de dos problemas diferentes a resolver a diferencia del Estudio de Demostración A, es decir, el AEMO/AE se ejecutó dos veces para cada diagnóstico que se diferenciaban en la optimización llevada a cabo sobre la media de la prevalencia (maximizar o minimizar). Es importante para los gestores y planificadores sanitarios conocer donde se localizan concentraciones anormales de casos de una determinada enfermedad o zonas con “mala salud”, pero también es importante saber dónde se encuentran concentraciones de pocos casos o relativa “buena salud”.

Previamente a la puesta en marcha del algoritmo se realizaron dos análisis espaciales exploratorios sobre las distribuciones de esquizofrenia y depresión en Cataluña mediante índices globales. En ambos casos se confirmó la existencia de dependencia espacial y clústeres de valores elevados. El AEMO/AE fue capaz de encontrar *hot-spots* tal como había previsto el índice global  $G$  pero también *cold-spots*. Este resultado no es contradictorio porque el análisis global capta tendencias generales en la distribución de una variable pero son los análisis locales los que identifican fenómenos concretos en el espacio.

A diferencia del Estudio de Demostración A, las diferencias estadísticas de la prevalencia entre los *hot-spots* y los *cold-spots* y el resto de unidades espaciales son claras para la Prueba  $U$  y la prueba de las medianas. El algoritmo ha encontrado efectivamente grupos de unidades espaciales con un comportamiento diferente de la prevalencia.

A partir de los resultados obtenidos se puede observar una estrecha relación entre los *hot-spots* y *cold-spots* de una determinada enfermedad ya que aparecen a menudo próximos o limítrofes.

Además, se aprecia que muchos de los municipios que son identificados como *cold-spots* aislados parecen disponerse alrededor de los *hot-spots* adquiriendo una forma lineal o radial. Como se ha mencionado, las agrupaciones lineales de unidades espaciales se han trazado en los mapas de forma subjetiva atendiendo a su proximidad y forma. Los clústeres radiales de *cold-spots* no han sido descritos previamente y ayudan a la delimitación de ciertos *hot-spots* que podrían haber sido considerados de forma conjunta debido a su proximidad. Los *hot-spots* y los *cold-spots* han sido estudiados juntos en varios artículos pero sus relaciones y significado no han sido analizados en profundidad (Cheng, Chen, et al., 2011; Sridharan, Koschinsky, et al., 2011). Ambos tipos de concentraciones deberían ser estudiadas juntas para mejorar la comprensión de la distribución espacial de enfermedades en el territorio.

Existen áreas de salud mental en la que coexisten *hot* y *cold-spots* de prevalencia de esquizofrenia y depresión. Anteriormente se ha comentado la posible existencia de comorbilidad entre ambas enfermedades o variaciones en la práctica médica en los servicios ubicados en dichas áreas como posibles explicaciones a esta situación.

### **5.3. Limitaciones del estudio**

En España, los datos de atención en servicios ambulatorios de salud mental no se recopilan en un único sistema de información centralizado como ocurre con la hospitalización breve (Conjunto Mínimo de Datos al Alta Hospitalaria-CMBD) sino que cada Comunidad posee sus propios sistemas. Además, en muchos casos las bases de datos de atención psiquiátrica ambulatoria no son integrables con otras bases sanitarias como hospitalización, atención primaria, etc., por lo que es complicado realizar seguimiento de casos y definir patrones de atención. Andalucía y Cataluña poseen registros de casos extra-hospitalarios de salud mental que recogen la actividad realizada en los servicios ambulatorios. Este tipo de dispositivos son la principal puerta de entrada de pacientes a la red de salud mental por lo que, en principio, deberían recoger la mayor parte de los casos atendidos en la atención especializada de salud mental del sistema público de salud (Moreno Küstner, 2008).

En Andalucía, el SISMA ha sido sustituido recientemente por el sistema Diraya que es un sistema integral de historias clínicas en el que se incluye también la atención especializada en salud mental (González Cocina y Pérez Torres, 2007). Desgraciadamente será difícil que en el futuro se puedan relacionar ambas bases de datos por sus diferentes características lo que limitará la realización de estudios epidemiológicos longitudinales.



Las bases de datos SISMA y CMBD-SMA han introducido algunas limitaciones a los Estudios de Demostración del modelo AEMO/AE. La principal limitación ha sido la ausencia de información sobre determinados centros de salud mental lo que ha supuesto la reducción del número de casos en sus respectivas áreas de cobertura.

En el caso de Andalucía, el área de Antequera no ha informado en los cuatro años de la serie temporal y además en 2008 la transición al sistema de información Diraya ha hecho perder los datos de cinco áreas más (Santa Fe, Sanlúcar, Villamartín, Jerez y Cádiz) con especial incidencia en la provincia de Cádiz. Esta situación ha podido producir que los *hot-spots* de esquizofrenia y depresión que han aparecido en la Sierra de Cádiz en ciertos años no lo hayan hecho en el año 2008.

En Cataluña, la ausencia de información afecta a nueve áreas de salud mental. Así, no hay datos de población adulta en las áreas de l'Hospitalet 2, Maragall y Vielha; de población infanto-juvenil en Ciutat Vella, Cerdanyola y Sant Martí Sud; y de ambos grupos de población en Ciutat Badia, la Mina y Sant Martí Nord.

La ausencia de datos para determinadas áreas de salud mental en ambas Comunidades Autónomas podría haber enmascarado la existencia de *hot-spots/cold-spots* en dichas áreas. Los municipios incluidos en las mismas presentan algunos casos porque no todas las personas acuden a su centro de referencia por diversos motivos, como la posibilidad de libre elección de facultativo, la accesibilidad al centro, etc. En Cataluña, la ausencia de datos es minimizada porque muchas de las áreas sin datos son inframunicipales mientras que la escala del estudio es municipal. Así, los municipios a los que pertenecen estas áreas siempre presentan casos aunque sean inferiores de los reales.

Otra pérdida en el número de casos es debida a la falta de cumplimentación de las variables municipio de residencia, sexo, edad y diagnóstico. Los pacientes diagnosticados con esquizofrenia y depresión sin municipio de residencia nunca han superado el 10% de la base de datos en Andalucía, mientras que en Cataluña la ausencia de estos datos ha sido mínima (inferior al 2%). En este sentido, la recopilación de información sanitaria en el SISMA y el CMBD-SMA responde a la voluntad del personal sanitario por lo que la saturación de los servicios en ciertos momentos puede influir en el registro de casos y visitas y, por tanto, en la calidad y exhaustividad de la información registrada en la base de datos.

El AEMO/AE permite plantear modelos de causalidad para intentar explicar el exceso o déficit de morbilidad en áreas concretas para lo cual es indispensable contar con series temporales que

permitan identificar patrones espaciales con estabilidad temporal en el territorio. El estudio de Demostración A se ha realizado sobre una serie temporal de cuatro años en el que se han observado clústeres espaciales que aparecen todos los años junto a otros que son inestables. Sería interesante estudiar la relación entre la inestabilidad de determinadas áreas con cambios en los indicadores sociodemográficos y ambientales. En cambio, el Estudio de Demostración B es una “foto fija” del año 2009. En este caso es claro el interés de estudiar nuevos años para observar la persistencia de los *hot-spots* y *cold-spots* que han sido identificados.

Otro aspecto importante es que los Estudios de Demostración no han analizado la prevalencia real de ambos diagnósticos psiquiátricos sino solamente aquellos que han sido atendidos en el Sistema Público de atención a la Salud Mental. Han quedado fuera de los estudios aquellos casos que son resueltos en los servicios de atención primaria y no precisan ser derivados a los servicios especializados. Esta situación también puede haber dificultado la identificación de otras concentraciones espaciales. De la misma forma, cierto número de casos psiquiátricos son atendidos en centros privados y tampoco quedan contabilizados en las bases de datos públicas.

Sería muy interesante poder integrar los casos atendidos en atención primaria para tener la *fotografía* completa de la prevalencia administrativa pero con los sistemas de información actuales esto es muy complejo (Moreno Kustner, 2011). En Cataluña, se ha realizado un intento de integrar la atención primaria con la atención especializada psiquiátrica (Aragonès, Salvador-Carulla, et al., en prensa) pero desafortunadamente las bases de datos de atención primaria en Cataluña no cubren todo el territorio y solo está disponible parte de la información sobre trastornos psiquiátricos atendidos en los mismos.

Junto a los problemas derivados de la disponibilidad y la calidad de las bases de datos se pueden encontrar una serie de problemas o limitaciones inherentes a los estudios de análisis espacial como la definición de la unidad espacial de estudio, el efecto frontera entre el ámbito de estudio y su entorno, la *falacia ecológica*, el problema de la *unidad espacial modificable* y el efecto de los municipios poco poblados sobre las tasas por habitante.

La unidad espacial de estudio (municipio) ha sido representada mediante polígonos en base a la delimitación geográfica de los términos municipales. A cada polígono se le ha asignado su valor de prevalencia correspondiente pero esto es una simplificación de la realidad porque no toda el área del polígono presenta la misma prevalencia sino que ésta se concentra en los núcleos de población (puntos). Se trata de un problema de escala y sería necesario aumentar la escala a niveles territoriales inferiores para que los análisis se aproximasen más a la realidad. Es complicado realizar estudios a escala inframunicipal debido a las normativas legales de secreto

estadístico que buscan evitar la posible identificación de individuos en escalas muy grandes. En las bases de datos analizadas existen variables destinadas a recoger los distritos postales y zonas/áreas básicas de salud donde residen los pacientes pero poseen una calidad baja en su cumplimentación en comparación a la escala municipal. El municipio es la división administrativa más adecuada para este tipo de estudios a fin de evitar problemas de secreto estadístico y encontrar información con la suficiente calidad. Los términos municipales españoles son más pequeños que los existentes en otros países europeos lo que permite un análisis a mayor detalle aunque dificultaría las comparaciones transnacionales.

Las unidades espaciales localizadas en las fronteras de la región suponen una importante restricción para el análisis espacial porque los valores de sus áreas vecinas localizadas en otras regiones o países no son conocidos. Si estos valores estuviesen disponibles posiblemente aparecerían *hot/cold-spots* adicionales en las fronteras administrativas porque la distribución de las enfermedades no se ajusta a las mismas (prevalencia real). El problema se reduce en ambas Comunidades debido a sus características geográficas con un amplio litoral y a la organización de la atención por áreas de salud definidas que se ajustan a las divisiones administrativas provinciales e internacionales (prevalencia administrativa).

Otro problema frecuente de este tipo de estudios es la *falacia ecológica*. Según la misma, no es posible inferir las características de un individuo a partir de las características de la población a la que pertenece (Macintyre, Ellaway, et al., 2002; Ocaña-Riola, 2010). Por esta razón es complejo relacionar la aparición de *hot-spots* y *cold-spots* con factores de riesgo que expliquen su existencia. No obstante, esta es una línea de investigación clara para continuar profundizando en esta metodología de AED en el futuro.

El problema de la *unidad espacial modificable* también es una dificultad adicional en los análisis espaciales (Elliott y Wartenberg, 2004; Rezaeian, Dunn, et al., 2007). Este problema se manifiesta en la interpretación de las estadísticas y sus resultados debido al tamaño del área geográfica donde los individuos han sido agregados. El impacto de los dos últimos problemas ha sido reducido en este trabajo al utilizar la escala municipal. Como se ha mencionado, este nivel de agregación es el que cuenta con mejor disponibilidad de información sin comprometer la confidencialidad de los pacientes.

También es habitual que las áreas con baja densidad de población y/o número de habitantes pueden presentar tasas con valores atípicos o extremos. La aparición de pocos casos dispara las tasas debido a que su denominador es pequeño (número de habitantes). Esta situación provoca

altas variaciones longitudinales debido a la posible existencia de variaciones aleatorias en el tiempo en el número de casos (Bithell, 2000).

El AEMO/AE busca grupos de municipios cercanos por lo que en teoría la aparición de tasas excepcionalmente altas serán identificadas como *hot-spot* si están acompañadas de municipios próximos con valores significativamente altos formando un clúster espacial. Como se ha visto anteriormente, el diseño del AEMO/AE puede hacer que se seleccionen municipios poco poblados con valores atípicos como *hot-spots* aislados porque el algoritmo ha primado otras *variables de decisión* sobre la distancia. Aun así, es también interesante para los clínicos y decisores conocer aquellas áreas poco pobladas donde se ha dado una proporción de casos extremos que pudieran ser debidos a factores excepcionales. En el caso de los *cold-spots* es más sencillo que aparezcan municipios aislados porque no existen valores extremos por debajo de cero casos, por lo que en muchas ocasiones el algoritmo completa las soluciones de 10 municipios con unidades espaciales aisladas con valor cero.

La aparición de pequeños municipios con tasas extremas entre los resultados puede dar lugar a conclusiones erróneas ya que se puede tratar de casos puntuales en un determinado año. El análisis longitudinal de varias series temporales permite detectar aquellos municipios pequeños en los que ha aparecido algún caso de forma aleatoria. Los resultados del AEMO/AE en Andalucía son estables en el tiempo en el caso de los *hot-spots* más extensos y en algunos casos de municipios aislados como Albox para ambos diagnósticos y la Subbética para la depresión. La mayor parte de los *hot-spots* aislados aparecen sólo en alguno de los años por lo que pueden depender circunstancias aleatorias. En el caso de Cataluña sería necesario analizar otros años para detectar si aparece este problema.

#### **5.4. *Hot-spots/cold-spots* y posibles factores explicativos**

La aparición de clústeres espaciales de esquizofrenia y depresión en el territorio podría ser debida a las características de las áreas de salud mental donde se localizan. Como se ha comprobado, la mayor parte de las concentraciones espaciales encontradas por el AEMO/AE comprenden municipios de varias áreas de salud mental que presentan en muchas ocasiones características diferentes en la práctica médica, en la provisión de servicios especializados de salud mental o en la accesibilidad a los centros de salud mental. En los dos últimos casos se podría descartar la existencia de anomalías o situaciones especiales en el funcionamiento de determinados servicios de salud mental que pudiesen provocar una acumulación de casos en sus ámbitos geográficos.

Las diferencias en el funcionamiento de los servicios de salud se asocian en muchas ocasiones a la variabilidad de la práctica médica que supone una línea propia en la investigación de servicios de salud. En España existe un grupo de investigación<sup>6</sup> especializado en el estudio territorial de la variabilidad de la práctica médica. Este grupo ha publicado un trabajo sobre la hospitalización psiquiátrica que señala que, aun considerando los efectos de la demografía y el azar, existe una variabilidad más alta de la esperada que puede deberse a diferencias territoriales en práctica médica (Aizpuru, Latorre, et al., 2008). Otros estudios sobre la variación de la morbilidad psiquiátrica hospitalaria han señalado la misma causa (Bas Sarmiento, Almenara Barrios, et al., 2009). La relación entre los clústeres espaciales y la práctica médica constituye una línea de investigación futura de este trabajo.

Otra característica de las áreas de salud mental que podrían relacionarse con los *hot-spots* y los *cold-spots* es la accesibilidad temporal a los centros de salud mental ambulatorios. La accesibilidad a los servicios de salud mental está vinculada a su localización que procede de una decisión política. Parece que una mejor accesibilidad a los centros de salud mental podría generar un mayor uso de los servicios y un incremento en los casos atendidos en centros de salud mental (Gittelsohn y Powe, 1995). A pesar de esto, la accesibilidad a los servicios de salud mental ambulatorios dentro de cada área presenta amplias diferencias en ambas regiones (Garrido, Salinas, et al., 2006; Olivet, Aloy, et al., 2008). En Andalucía, las distancias temporales pueden llegar hasta 45 minutos en Sierra Nevada pero no son tan altas en Aracena o la Sierra de Cádiz. En Cataluña, existen áreas con una accesibilidad que puede llegar a los 45 minutos como Anoia, Bages Borges Blanques, Berga, Osona, Seu d'Urgell y Vallès Oriental, pero en otras no llegan a los 15 minutos como por ejemplo aquellas que se encuentran en el área metropolitana de Barcelona o el litoral.

Las diferencias por áreas en la provisión de servicios de la cartera de salud mental (atención hospitalaria de corta, media y larga estancia, atención de día, ambulatoria, etc.) podrían repercutir en la prevalencia administrativa de servicios ambulatorios por varios motivos (Marín y Briones, 2007). La carencia de servicios intermedios puede provocar una sobrecarga de casos en servicios ambulatorios. También una disponibilidad baja de servicios especializados podría generar una demanda menor y por tanto un menor número de casos. La provisión de servicios de las áreas de salud mental andaluzas y catalanas es elevada y casi todas las áreas de salud mental tienen asignados todos los tipos de servicios disponibles. Las diferencias entre áreas no se encuentran tanto en la asignación de dispositivos sino en que éstos se encuentren localizados dentro o fuera del área, es decir, sean más o menos accesibles. En este sentido, se aprecian

---

<sup>6</sup> El grupo Atlas VPM estudia los beneficios y/o riesgos en los cuidados de salud a los que está sometida la población por vivir en un determinado lugar en España (<http://www.atlasvpm.org>).

claras diferencias en la provisión de servicios de salud mental entre las áreas de salud mental que presentan clústeres espaciales en ambas Comunidades por lo que no se intuye un patrón común (Garrido, Salinas, et al., 2006; GEOSCAT Group, en prensa).

Otros factores de riesgo que pueden influir en la salud mental son los ambientales. Por ejemplo, las diferencias en la morbilidad psiquiátrica que se producen entre áreas urbanas y rurales. La esquizofrenia parece que es más frecuente en la población urbana (Kelly, O'Callaghan, et al., 2010; Kirkbride, Fearon, et al., 2007), aunque un estudio en Finlandia no encontró relación con el grado de urbanización quizás por la carencia de grandes ciudades en este país (Perälä, Saarni, et al., 2008).

La clasificación de las áreas de salud mental siguiendo los criterios de ruralidad de la Organización para la Cooperación y el Desarrollo Económico (OCDE)<sup>7</sup> muestra que las áreas donde se localizan los *hot-spots* y *cold-spots* no presentan un patrón claro en este sentido en ninguna de las dos Comunidades Autónomas (Garrido, Salinas, et al., 2006; Salinas-Pérez, García-Alonso, et al., 2012). En Andalucía, las áreas de Albox, Aracena, Guadix y Órgiva son rurales; la del Poniente y Roquetas son urbanas y el áreas de Almería presenta una gran superficie que mezcla espacios rurales y urbanos. Las demás áreas donde han aparecido *hot-spots* en alguno de los años son predominantemente rurales como la Sierra de Cádiz o la Subbética excepto los pequeños *hot-spots* en torno a Sevilla. Y en Cataluña, ciertos *hot-spots* y *cold-spots* de esquizofrenia y depresión se localizan en áreas urbanas como el Vallès Oriental, el Maresme Nord, Garraf, Gavà, Sant Feliu de Llobregat, Martorell o el Vendrell; pero también en áreas rurales como Alt Empordà, Anoia, Bages, Berga, Borges Blanques, La Segarra, Mollerussa, Mora d'Ebre o Seu d'Urgell.

Se han encontrado más áreas de salud mental con concentraciones espaciales con características urbanas en Cataluña que en Andalucía. Aunque cabe destacar que muchos de los municipios identificados en áreas urbanas en Cataluña presentan una densidad de población baja. La aparición de más casos de los esperados en estos municipios podría haber provocado la aparición de *hot-spots*. Sería necesario un estudio longitudinal para comprobar si estos *hot-spots* se mantienen en el tiempo.

También existen diversas variables demográficas y socioeconómicas que parecen influir sobre la aparición, difusión y comportamiento de las enfermedades como es la estructura de la

---

<sup>7</sup> Un área geográfica es “predominantemente urbana” cuando el 85% de los habitantes residen en municipios con una densidad mayor que 150 habitantes/km<sup>2</sup>, “significativamente rural” cuando el porcentaje se encuentra entre el 50% y el 84% y “predominantemente rural” cuando el porcentaje es inferior al 50% (OECD, 1994).

población por edad y sexo, estado civil, nivel educativo, situación laboral, etc. (Alonso, Angermeyer, et al., 2004; Bones Rocha, Pérez, et al., 2010). También puede relacionarse con factores sociales como los efectos sobre la salud mental que provocan las diferencias territoriales en la colaboración social entre individuos y grupos (capital social) (Almedom, 2005; Whitley y McKenzie, 2005).

La esquizofrenia ha sido asociada a diversos factores de riesgo. Por ejemplo, numerosos estudios han señalado la existencia de factores genéticos junto a factores de riesgo exógenos (Owen, 2012). Las variables socioeconómicas que se han relacionado con la esquizofrenia son el nivel educativo, el aislamiento social, migración, nutrición y la alta exposición a infecciones (Eaton, Mortensen, et al., 2000). Un reciente estudio ha relacionado una alta utilización de servicios de salud mental por parte de pacientes de esquizofrenia con el desempleo, un bajo nivel educativo y la recepción de ayudas sociales (Moreno-Küstner, Mayoral, et al., 2011).

En cuanto a la depresión, también parece existir una componente genética como señala una reciente revisión bibliográfica (Wilde, Mitchell, et al., 2012). El proyecto internacional ESEMeD muestra relaciones entre la prevalencia de la depresión y el sexo, el desempleo, el estado civil y la discapacidad (Gabilondo, Rojas-Farreras, et al., 2010). También se han encontrado relaciones con un bajo nivel educativo y vivir solo (Sabes-Figuera, Knapp, et al., 2012). El desempleo parece un factor de riesgo claro en el caso de la depresión (Jefferis, Nazareth, et al., 2011).

Los estudios de análisis espacial de enfermedades son de gran utilidad a la hora de establecer hipótesis que deben ser verificadas mediante estudios de casos para evitar la *falacia ecológica*. El siguiente paso de esta investigación es relacionar la existencia de las concentraciones espaciales detectadas por el AEMO/AE con variables socioeconómicas y/o ambientales que pudieran ser posibles factores de riesgo, ya sea mediante análisis multivariantes o índices sintéticos (Garrido-Cumbrera, Almenara-Barrios, et al., 2008). En este sentido, existe en marcha un proyecto del Instituto de Salud Carlos III (ISCIII) que relacionará los resultados del AEMO/AE con factores de riesgo mediante métodos multi-nivel, regresión y clasificación ordinal.

## **5.5. Difusión del modelo AEMO/AE**

Como se ha comentado tanto el diseño del modelo AEMO/AE como los Estudios de Demostración han sido realizados en el marco de dos proyectos de investigación financiados por

el Instituto de Salud Carlos III y en un proyecto financiado por el Departamento de Salud de Cataluña. La producción científica derivada de los mismos se ha difundido en cuatro congresos:

- García-Alonso C, Salinas-Pérez JA, Salvador-Carulla L. *Spatial Identification of Hot and Cold Spots for the Prevalence of Schizophrenia and Depression*. International Conference on Operations Research, OR2010. Munich, 1 – 3 de septiembre de 2010.
- García-Alonso C, Salinas-Pérez JA, Salvador-Carulla L. *Spatial Analysis of Schizophrenia Prevalence Using a Multiobjective Evolutionary Algorithm*. 5th Workshop on Data Mining and Health Informatics. Institute for Operations Research and the Management Sciences INFORMS Annual Meeting 2010. Austin, 5 – 11 de noviembre de 2010.
- Pérez-Ortiz M, García-Alonso C, Gutiérrez PA, Salvador-Carulla L, Salinas-Pérez JA, Hervás-Martínez C. *Ordinal classification of depression spatial hotspots of prevalence*. ISDA 2011: Eleventh International Conference on Intelligent Systems Design and Applications. Córdoba, 22 – 24 de noviembre de 2011.
- García Alonso CR, Salinas Pérez JA, Salvador Carulla L, Almenara Barrios J, Peralta Sáez JL, Poole Quintana, ML, Rodero Cosano ML, De los Reyes Vázquez SM. *Identificación y caracterización socioeconómica de hot-spots de depresión en Andalucía*. Póster presentado a la XXX Reunión científica de la Sociedad Española de Epidemiología: Epidemiología en tiempos de crisis: Haciendo sostenible el sistema de salud. Santander, 17-19 octubre de 2012.

Y publicado en dos revista científicas:

- García-Alonso CR, Salvador-Carulla L, Negrín-Hernández MA, Moreno-Küstner B. *Development of a new spatial analysis tool in mental health: identification of highly autocorrelated areas (hot-spots) of schizophrenia using a Multiobjective Evolutionary Algorithm model (MOEA/HS)*. *Epidemiologia e Psichiatria Sociale* 2010 dic;19(4):302–313.
- Salinas-Perez JA, Garcia-Alonso CR, Molina-Parrilla C, Jorda-Sampietro E, Salvador-Carulla L. *Identification and location of hot and cold spots of treated prevalence of depression in Catalonia (Spain)*. *International Journal of Health Geographics* 2012 ago;11(1):36. (Anexo 2)



La herramienta también ha mostrado su validez con variables económicas y se ha empleado para explorar la existencia de *hot-spot* de valores altamente auto-correlacionados de financiación pública comprometida al cultivo del olivar también en Andalucía (García-Alonso, Pérez-Naranjo, et al., 2011). Una línea de trabajo futuro es aplicar el modelo AEMO/AE a otras series temporales y territorios. Así como su aplicación sobre otras patologías como ansiedad, trastornos obsesivos, etc.

Las técnicas espaciales de representación de variables socioeconómicas y sanitarias han sido empleadas anteriormente en diversos Atlas de Salud Mental desarrollados por PSICOST bajo la dirección del Dr. Luis Salvador Carulla:

- 2005. *Atlas de la Salud Mental de Andalucía*. Desarrollado por la Asociación Científica PSICOST para el Servicio Andaluz de Salud.
- 2008. *Atlas de la Salud Mental de Catalunya*. Desarrollado por la Asociación Científica PSICOST para el Servei Català de la Salut (CatSalut).
- 2009. *Atlas de la Salud Mental de Cantabria*. Desarrollado por la Asociación Científica PSICOST para la Consejería de Sanidad del Gobierno de Cantabria.
- 2009. *Atlas de la Salud Mental de España*. Desarrollado por la Asociación Científica PSICOST para el Ministerio de Sanidad y Política Social.
- 2011. *Atlas de la Salud Mental de Madrid*. Desarrollado por la Asociación Científica PSICOST para la Consejería de Sanidad de la Comunidad de Madrid.
- 2012. *Atlas de Salud Mental de Gipuzkoa*. Desarrollado por la Asociación Científica PSICOST para el Instituto Vasco de Innovación Sanitaria / o+berri (Fundación Vasca de Innovación e Investigación Sanitarias).
- 2012. *Atlas de Salud Mental de Bizkaia*. Contrato Desarrollado por la Asociación Científica PSICOST para la Red de Salud Mental de Bizkaia.

Dos de los mismos han sido publicados:

- Garrido M, Salinas JA, Almenara J, Salvador L. Atlas de Salud Mental de Andalucía 2005. Sevilla: Servicio Andaluz de Salud; 2006. ISBN: 978-84-690-6666-9.
- Vázquez-Barquero JL, Gaité L, Salvador L, Salinas JA. Atlas de Salud Mental de Cantabria. Santander: Consejería de Sanidad; 2010. ISBN: 978-84-693-6768-1.

La visualización espacial de información sanitaria también es parte importante de un *workpackage* del proyecto internacional *Research on Financing systems' Effect on the Quality of Mental health care (REFINEMENT)* perteneciente al VII Programa Marco de la Unión Europea (*Project number 261459*) coordinado desde la Universidad de Verona y donde PSICOST es el socio español (2011-2013) bajo la dirección del Dr. Luis Salvador Carulla y el Dr. Carlos Ramón García Alonso.

Una posibilidad interesante para la difusión futura del modelo AEMO/AE sería su conversión en un paquete informático. Existen diversos ejemplos de metodologías de análisis espacial que han sido convertidas en *software* de libre utilización y que han permitido su amplia difusión científica<sup>8</sup>. Es el caso paradigmático de las *spatial scan statistics* desarrolladas por Martin Kulldorff e implementadas en SaTScan<sup>9</sup>. Este *software* ha tenido una amplia difusión en la literatura científica incrementando el número de citas bibliográficas a los trabajos de Kulldorff en diversos campos de investigación sanitaria y social. También ha permitido su colaboración directa en investigaciones sobre mejoras del método.

Aprovechando esta experiencia, sería interesante convertir la herramienta creada en un *software* libre con el que permitir a otros usuarios realizar análisis sobre sus propios datos. Esto permitiría probar el Algoritmo Evolutivo en otros tipos de datos y obtener mayor difusión en el ámbito científico. Sería necesario diseñar un entorno de usuario que facilitara su utilización a analistas no familiarizados con la programación informática. La principal limitación en este sentido es que la herramienta precisa de una potencia de computación importante que no se puede desarrollar en computadoras convencionales.

---

<sup>8</sup> Existen numerosos ejemplos de software de análisis espacial de datos, mencionaremos los open source: OpenGeoDa del GeoDa Center for Geospatial Analysis and Computation, SaTScan de M. Kulldorff y Information Management Services Inc., PASSaGE desarrollado por M. Rosenberg y C.D. Anderson de la Universidad del Estado de Arizona.

<sup>9</sup> Más información sobre el *software* y los trabajos científicos derivados del mismo se puede encontrar en <http://www.satscan.org/>

## 5.6. AED como apoyo a la decisión en planificación de la salud mental

Los Estudios de Demostración han analizado la prevalencia administrativas que es un indicador de resultado que puede ayudar a evaluar la actividad y la equidad en la atención de los servicios de salud mental. La localización de los *hot-spots* y *cold-spots* puede ser útil dentro de los Sistemas de Apoyo a la Decisión en salud, ya que identifica áreas geográficas con un número de casos atípicos donde podrían requerirse la puesta en marcha de acciones específicas de prevención y atención desde los servicios de salud mental. Los resultados de los análisis espaciales y sus posibles relaciones con variaciones en las condiciones locales permitiría también implementar programas, planes y políticas de salud flexibles y dirigidos a espacios geográficos concretos (Koschinsky, en prensa).

Los modelos de apoyo a la decisión en política sanitaria han tenido un gran avance en los últimos años a través de la sustitución de la aproximación tradicional de atención basada en la evidencia (*Evidence-Based Care*) por la evidencia informada (*Evidence-Informed*) (Lavis, Oxman, et al., 2009), junto al desarrollo de nuevos marcos conceptuales y directrices para el apoyo a la toma de decisiones (Lewin, Oxman, et al., 2009).

El análisis de la distribución espacial de enfermedades y la búsqueda de posibles patrones en la misma poseen un gran interés para el conocimiento epidemiológico de la enfermedad y como fuente de la evidencia informada de cara al diseño de la planificación, programas y políticas sanitarias. Tradicionalmente se ha considerado la *evidencia informada global* a escala mundial para conocer los efectos que determinadas actuaciones podían generar y los factores que los modificaban, pero se ha comprobado la necesidad de considerar las escalas nacionales, regionales o locales (Salvador-Carulla, Costa-Font, et al., 2010). La evidencia más próxima a los decisores en salud es aquella que informa sobre las condiciones locales (*evidencia informada local*) de su entorno (factores, necesidades, valores, costes y recursos) y que es necesaria para juzgar las decisiones y acciones que deben tomarse en el desarrollo de la política sanitaria (Lewin, Oxman, et al., 2009).

Los *hot-spot* y *cold-spot* identificados por el algoritmo precisan ser interpretados por expertos con conocimientos sobre la salud mental en el ámbito geográfico de estudio. La interpretación correcta de los resultados de estos análisis requieren la utilización de nuevos sistemas de análisis como el *Knowledge Discovery from Data* (KDD) en el que la visualización de la información junto al conocimiento experto a priori es fundamental (Gibert y Nonell, 2008). Su aplicación a la investigación de Sistemas Sanitarios facilita la obtención de conocimiento experto a priori y mejora la comprensión de información compleja a través, por ejemplo, del

Análisis Cooperativo basado en Expertos (Expert-based Cooperative Analysis (EbCA) desarrollado por Gibert y colaboradores (Gibert, García-Alonso, et al., 2010)

Una vez aplicado el modelo AEMO/AE sobre datos epidemiológicos de salud mental y difundido sus resultados se ha considerado interesante conocer la opinión de expertos en investigación, planificación y gestión de servicios de salud mental. Se ha entrevistado a 17 expertos mediante un cuestionario sobre herramientas de análisis y representación espacial (Anexo 1) que recogía preguntas sobre su utilidad y relevancia y posibles variables de interés para futuras investigaciones. Los resultados señalan que la utilidad y la relevancia del modelo AEMO/AE tienen una puntuación de 8,1 y 8,4 respectivamente en la escala de 0 a 10, siendo el valor mínimo 6 y el máximo 10. En cuanto a la escala ordinal, la utilidad ha sido denominada 11 veces como “Alta” (64,7%) y 6 veces como “Media” (35,3%). Y la relevancia ha sido calificada 12 veces como “Alta” (70,6%) y 5 veces como “Media” (29,4%).

La valoración del AEMO/AE es muy positiva aunque inferior a otras herramientas geográficas más consolidadas como la representación cartográfica de variables sanitarias y los atlas de salud mental. El motivo es la necesidad de dar un paso más en el análisis espacial. La utilidad y relevancia del modelo AEMO/AE debe acrecentarse en gran medida cuando se relacionen los *hot-spots* y *cold-spots* con posibles factores de riesgo sociales y ambientales. También cuestiones propias de la planificación y gestión de los servicios como la provisión de servicios, incluyendo todas las tipologías de la red de atención a la salud mental, la accesibilidad a los mismos, o la propia variabilidad de la práctica médica deben ser estudiadas en profundidad junto a los resultados del análisis espacial de datos. Como es sabido, es complejo encontrar factores causales que expliquen totalmente las concentraciones espaciales de casos pero su estudio puede ayudar a generar hipótesis sobre los posibles factores que intervienen en su origen. Las perspectivas en el futuro son amplias tanto para la mejora del modelo AEMO/AE como para las aplicaciones del mismo en el campo de la salud mental o en el análisis de la distribución geográfica de otros tipos de fenómenos.

### **5.7. Líneas futuras de desarrollo de la investigación**

1. Estudiar la posibilidad de no limitar el número de las unidades espaciales que componen las soluciones del Algoritmo Evolutivo Multi-Objetivo.
2. Estudiar la posibilidad de considerar la proximidad entre unidades espaciales mediante otros métodos ya sea por distancia o por geometría.

3. Estudiar posibles relaciones de los clústeres espaciales de casos de esquizofrenia y depresión con factores de riesgo sociales y ambientales mediante técnicas de clasificación ordinal.
4. Estudiar posibles relaciones de los clústeres espaciales de casos de esquizofrenia y depresión con factores derivados de la planificación y gestión de la salud mental (provisión de servicios, accesibilidad temporal, disponibilidad por habitante, práctica médica, etc.).
5. Incorporar otros índices espaciales como las *spatial scan statistics* que permita obtener un resultado híbrido que englobe las principales metodologías de AED actuales.
6. Estudiar los *hot-spots* y *cold-spots* de prevalencia administrativa de trastornos mentales en Andalucía al igual que en el estudio de Cataluña.
7. Estudiar la distribución espacial de la esquizofrenia y la depresión en una serie temporal en Cataluña para comparar sus resultados con los obtenidos en 2009.
8. Aplicar el modelo AEMO/AE en otros territorios (existe interés desde las diputaciones forales del País Vasco) o en otros diagnósticos psiquiátricos o de otro tipo.
9. Estudiar bases de datos integrales, como el Diraya en Andalucía, que permitan aproximarse a la prevalencia administrativa real de salud mental tanto en servicios especializados como en atención primaria.
10. Estudiar otras variables sanitarias de interés como la incidencia, frecuentación, altas hospitalarias, mortalidad, etc.
11. Aplicar el modelo AEMO/AE en otros campos científicos como la Economía o la Cooperación al Desarrollo.
12. Estudiar la posibilidad de convertir el modelo AEMO/AE en un paquete informático *open source* para facilitar su difusión y desarrollo futuro.

## VI. CONCLUSIONES

---

1. Los Algoritmos Evolutivos Multi-objetivo han sido aplicados de forma exitosa al análisis de concentraciones espaciales estadísticamente significativas de casos de enfermedades.

Como se proponía, esta investigación ha diseñado, implementado y probado un modelo de análisis espacial basado en los Algoritmos Evolutivos Multi-objetivo (AEMO/AE). El modelo ha identificado de manera eficaz ciertas áreas geográficas con concentraciones significativas de las variables objeto de estudio como se ha podido comprobar en dos Estudios de Demostración. La descripción estadística y la cartografía de los resultados avalan esta afirmación. Por tanto, los AEMO pueden ser destinados a la resolución eficaz de problemas espaciales tal como se planteó en la primera hipótesis de partida de esta investigación.

2. El modelo AEMO/AE ha identificado eficazmente *hot-spots* de alta dependencia espacial de la esquizofrenia y la depresión en Andalucía.

Tal como exponía la segunda hipótesis de partida, el modelo AEMO/AE ha funcionado de forma correcta como método híbrido entre varias técnicas de análisis espacial consolidadas. Los resultados del Estudio de Demostración A señalan áreas geográficas con una dependencia espacial significativa en la distribución geográfica de la prevalencia administrativa de enfermedades mentales (*hot-spots*) que están avaladas por los cuatro métodos de análisis espacial estudiados. La solución de consenso muestra de forma inequívoca donde existe una fuerte interrelación de los datos debida a su posición geográfica.

3. El modelo AEMO/AE ha identificado eficazmente *hot-spots* y *cold-spots* de esquizofrenia y depresión en Cataluña.

El modelo también ha funcionado como herramienta de análisis espacial por sí mismo sin necesidad de ser aplicado como un método híbrido. Los resultados del Estudio de Demostración B señalan con efectividad áreas geográficas en las que la prevalencia administrativa de enfermedades mentales es significativamente elevada o baja (*hot-spots* y *cold-spots*). Por tanto, la tercera hipótesis de partida queda demostrada también.

4. Los resultados del modelo AEMO/AE son soluciones óptimas a los problemas multi-objetivo de tipo territorial que se han planteado. Los resultados están influidos por problemas intrínsecos a los estudios ecológicos y geográficos pero que son externos al modelo.

Los resultados obtenidos en los Estudios de Demostración responden de forma óptima a los problemas multi-objetivos propuestos. Aun así, se han detectado algunas soluciones no esperadas o incoherentes que recomiendan seguir profundizando en esta investigación con el fin de mejorar el modelo AEMO/AE. También se han encontrado otras limitaciones externas al diseño del modelo AEMO/AE que afectan a los resultados como los problemas de disponibilidad, calidad e integración de las bases de datos sanitarias y una serie de limitaciones inherentes a los estudios geográficos y ecológicos que tienen que ser solventadas o reducidas con anterioridad al funcionamiento del modelo.

5. Los resultados del AEMO/AE son útiles y relevantes para la planificación en salud mental.

Los resultados del AEMO/AE han identificado zonas geográficas de relativa “*mala salud*” y “*buena salud*” que son de interés para la planificación y la gestión de los servicios de salud mental. En todo caso, los resultados deben ser interpretados por expertos en planificación, gestión e investigación de servicios de salud mental con conocimientos sobre el área de estudio. El modelo AEMO/AE ha sido valorado por un grupo de expertos como útil y relevante. Aun así, su valoración será más elevada cuando se relacionen los *hot-spots* y *cold-spots* con posibles factores de riesgo social y/o ambiental y con las características de las áreas de atención relacionadas con la planificación territorial y práctica médica de los servicios de salud mental.

## VII. BIBLIOGRAFÍA

---

1. Aamodt G, Samuelsen SO, Skrondal A. A simulation study of three methods for detecting disease clusters. *Int J Health Geogr.* 2006;5:15.
2. Abellana R, Ascaso C. Métodos estadísticos para analizar el riesgo con patrones de distribución espacial. *Med Clin (Barc)* 2004;122(Supl 1):68–72.
3. Aizpuru F, Latorre K, Ibáñez B, Pérez de Arriba J, Mosquera F, Bernal-Delgado E, et al. Variabilidad en la tasa de hospitalizaciones por problemas de Salud Mental en Centros hospitalarios de agudos [Internet]. *Atlas de Variaciones de la Práctica Médica* 2008; [citado 2012 sep 24] Available from: <http://www.atlasvpm.org/avpm/nodo.datos.do?tipo=FICH&index=0&ID=43>
4. Almedom AM. Social capital and mental health: an interdisciplinary review of primary evidence. *Soc Sci Med* 2005 sep;61(5):943–964.
5. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, et al. Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatrica Scandinavica* 2004;109:21–27.
6. Anselin L, Bera AK. Spatial dependence in linear regression models with an introduction to spatial econometrics. En: Ullah A, Giles DE., editores. *Handbook of applied economic statistics*. New York: Marcel Dekker; 1998
7. Anselin L, Hudak S. Spatial econometrics in practice: A review of software options. *Regional Science and Urban Economics* 1992 sep;22(3):509–536.
8. Anselin L. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer; 1988.
9. Anselin L. Local indicators of spatial association-LISA. *Geographical Analysis* 1995;27:93–115.
10. Anselin L. Under the hood: Issues in the specification and interpretation of spatial regression. *Agricultural Economics* 2002;27(3):247–267.
11. Aragonès E, Salvador-Carulla L, López-Muntaner J, Ferrer M, Piñol JL. Registered prevalence of borderline personality disorder in primary care databases. *Gac Sanit.* en prensa;
12. Aragonés N, Pérez-Gómez B, Pollán M, Ramis R, Vidal E, Lope V, et al. The striking geographical pattern of gastric cancer mortality in Spain: environmental hypotheses revisited. *BMC Cancer* 2009;9:316.
13. Arbor A. *The international classification of diseases, 9th revision, clinical modification, ICD-9-CM*. Michigan: Commission on Professional and Hospital Activities; 1978.



14. Auchincloss AH, Gebreab SY, Mair C, Diez Roux AV. A Review of Spatial Methods in Epidemiology, 2000–2010. *Annu Rev Public Health*. 2012;33(1):107–122. [citado 2012 mar 29 ]
15. Bäck T. *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. New York: Oxford University Press; 1996.
16. Bailey TC, Gatrell AC. *Interactive spatial data analysis*. Harlow: Longman Scientific & Technical; 1995.
17. Bas Sarmiento P, Almenara Barrios J, Salvador Carulla L, Salinas Pérez JA, Fernández Gutiérrez M, García Gutiérrez JC, et al. Epidemiología de gestión de la morbilidad psiquiátrica atendida en los hospitales españoles. *Anales de Psiquiatria* 2009;25(4):167–175. [citado 2011 feb 17 ]
18. Beale L, Hodgson S, Abellan JJ, Lefevre S, Jarup L. Evaluation of spatial relationships between health and the environment: the rapid inquiry facility. *Environ. Health Perspect* 2010 sep;118(9):1306–1312.
19. Beirlant J, Goegebeur Y, Teugels J, Segers J. *Statistics of extremes: theory and applications*. Chichester, West Sussex: Wiley; 2004.
20. Bentley PJ, Wakefield JP. Finding acceptable solutions in the Pareto-optimal range using multiobjective genetic algorithms. En: Chawdhry PK, Roy R, Pant K R, editores. *Soft Computing in Engineering Design and Manufacturing*. London: Springer Verlag; 1997 p. 231–240.
21. Besag J, Green PJ. Spatial statistics and bayesian computation. *J. R. Stat. Soc. Ser. B-Methodol*. 1993;55(1):25–37.
22. Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991;43(1):1–20.
23. Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. *Stat Methods Med Res* 2005;14(1):35–39.
24. Bilancia M, Fedespina A. Geographical clustering of lung cancer in the province of Lecce, Italy: 1992-2001. *Int J Health Geogr*. 2009;8(1):40.
25. Bithell JF. A classification of disease mapping methods. *Stat Med*. 2000 sep;19(17-18):2203–2215.
26. Bodson P, Peeters D. 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* 1975;7:455–472.
27. Bones Rocha K, Pérez K, Rodríguez-Sanz M, Borrell C, Obiols JE. Prevalencia de problemas de salud mental y su asociación con variables socioeconómicas, de trabajo y salud: resultados de la Encuesta Nacional de Salud de España. *Psicothema* 2010 ago;22(3):389–395.

28. Borja-Aburto VH. Estudios Ecológicos. *Salud Publica Mex* 2000 dic;42(6):533–538.
29. Brooks SP, Gelman A. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics* 1998 dic;7(4):434–455.
30. Campos MS, Martínez-Larrea JA. [Affective disorders: analysis of their comorbidity in the more frequent psychiatric disorders]. *An Sist Sanit Navar* 2002;25 Suppl 3:117–136.
31. Cançado AL, Duarte AR, Duczmal LH, Ferreira SJ, Fonseca CM, Gontijo EC. Penalized likelihood and multi-objective spatial scans for the detection and inference of irregular clusters. *Int J Health Geogr.* 2010 oct;9(1):55.
32. Case AC, Rosen HS, Hines JR. Budget spillovers and fiscal policy interdependence: evidence from the states. *Journal of public economics* 1993;52:285–307.
33. Chaix B, Leyland AH, Sabel CE, Chauvin P, Råstam L, Kristersson H, et al. Spatial clustering of mental disorders and associated characteristics of the neighbourhood context in Malmö, Sweden, in 2001. *J Epidemiol Community Health.* 2006;60(5):427–435.
34. Chang S-S, Sterne JAC, Wheeler BW, Lu T-H, Lin J-J, Gunnell D. Geography of suicide in Taiwan: Spatial patterning and socioeconomic correlates. *Health & Place* 2011 mar;17(Issue 2):641–650.
35. Cheng C-L, Chen Y-C, Liu T-M, Kao-Yang Y-H. Using Spatial Analysis to Demonstrate the Heterogeneity of the Cardiovascular Drug-Prescribing Pattern in Taiwan. *BMC Public Health.* 2011 may;11(1):380.
36. Cliff AD, Ord JK. The problem of spatial autocorrelation. En: Scott AJ, editor. *London Papers in Regional Science. Studies in Regional Science.* London: Pion; 1969 p. 25–55.
37. Cliff AD, Ord JK. Spatial autocorrelation: a review of existing and new measures with applications. *Economic Geography* 1970;46:269–292.
38. Cliff AD, Ord JK. Testing for spatial autocorrelation among regression residuals. *Geographical Analysis* 1972;4(3):267–284.
39. Coello CAC, Lamont GB. *Applications of multi-objective evolutionary algorithms.* Singapore: World Scientific Publishing; 2004.
40. Coello-Coello C, Lamont G, Van Veldhuizen D. *Evolutionary algorithms for solving multi-objective problems.* New York: Springer; 2007.
41. Cohon JL, Marks DH. A review and evaluation of multiobjective programming techniques. *Water Resour. Res.* 1975;11(2):208–220.
42. Consejería de Obras Públicas y Vivienda de Andalucía. Instituto de Estadística y Cartografía de Andalucía [Internet]. 2012; [citado 2012 sep 24] Available from: <http://www.juntadeandalucia.es/institutodeestadisticaycartografia>
43. Consejería de Salud. *Plan Integral de Salud Mental de Andalucía (2003-2007).* Sevilla: Consejería de Salud; 2003.

44. Consejería de Salud. II Plan Integral de Salud Mental de Andalucía (2008-2012). Sevilla: Consejería de Salud. Junta de Andalucía; 2008a.
45. Consejería de Salud. Decreto 77/2008, de 4 de marzo, de ordenación administrativa y funcional de los servicios de Salud Mental en el ámbito del Servicio Andaluz de Salud. BOJA nº 53; 2008b.
46. Cressie NAC. *Statistics for spatial data*. New York: John Wiley & Sons; 1991.
47. Cromley EK, McLafferty S. *GIS and public health*. New York: Guilford Press; 2002.
48. Curtis S, Copeland A, Fagg J, Congdon P, Almog M, Fitzpatrick J. The ecological relationship between deprivation, social isolation and rates of hospital admission for acute psychiatric care: a comparison of London and New York City. *Health Place*. 2006;12(1):19–37.
49. Das I, Dennis J. A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Structural optimization* 1997;14(1):63–69.
50. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *Evolutionary Computation, IEEE Transactions on* 2002;6(2):182–197.
51. Deb K. *Multi-objective optimization using evolutionary algorithms*. West Sussex: Wiley; 2004.
52. Departament de Sanitat i Seguretat Social. *Manual de notificació Registre del Conjunt Mínim Bàsic de Dades: Centres de Salut Mental Ambulatoria*. Barcelona: Servei Català de la Salut; 2003.
53. Departament de Territori i Sostenibilitat. Institut Cartogràfic de Catalunya [Internet]. 2012; [citado 2012 sep 24] Available from: <http://www.icc.cat/>
54. Doi Y, Yokoyama T, Sakai M, Nakamura Y, Tango T, Takahashi K. Spatial Clusters of Creutzfeldt-Jakob Disease Mortality in Japan between 1995 and 2004. *Neuroepidemiology* 2008;30(4):222–228.
55. Doreian P. Linear Models with Spatially Distributed Data. *Spatial Disturbances or Spatial Effects? Sociological Methods & Research* 1980;9(1):29 –60.
56. Duczmal L, Caçado ALF, Takahashi RHC, Bessegato LF. A genetic algorithm for irregularly shaped spatial scan statistics. *Computational Statistics & Data Analysis* 2007 sep;52(1):43–52.
57. Duczmal L, Caçado ALF, Takahashi RHC. Geographic Delineation of Disease Clusters through Multi-Objective Optimization. *Journal of Computational & Graphical Statistics* 2008;17:243–262.

58. Earnest A, Morgan G, Mengersen K, Ryan L, Summerhayes R, Beard J. Evaluating the effect of neighbourhood weight matrices on smoothing properties of Conditional Autoregressive (CAR) models. *Int J Health Geogr.* 2007;6(1):54.
59. Eaton WW, Mortensen PB, Frydenberg M. Obstetric factors, urbanization and psychosis. *Schizophr. Res.* 2000 jun;43(2-3):117–123.
60. Ehrgott M. *Multicriteria optimization*. Second edition. Berlin: Springer Verlag; 2005.
61. Eitan O, Yuval ., Barchana M, Dubnov J, Linn S, Carmel Y, et al. Spatial analysis of air pollution and cancer incidence rates in Haifa Bay, Israel. *Sci Total Environ* 2010;408(20):4429–4439.
62. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environ Health Perspect.* 2004;112(9):998–1006.
63. Escaramis G, Carrasco JL, Aponte JJ, Nhalungo D, Nhacolo A, Alonso P, et al. Spatio-temporal analysis of mortality among children under the age of five in Manhica (Mozambique) during the period 1997-2005. *Int J Health Geogr.* 2011;10(1):14.
64. ESRI. ESRI [Internet]. ESRI 2011; [citado 2012 sep 24] Available from: <http://www.esri.com/>
65. Fonseca CM, Fleming PJ. Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. En: *Proceedings of the fifth international conference on genetic algorithms*. San Mateo, California: University of Illinois at Urbana-Champaign, Morgan Kaufman Publishers.; 1993 p. 416–423.
66. Fortney JC, Rushton G, Wood S, Zhang L, Xu S, Dong F, et al. Community-Level Risk Factors for Depression Hospitalizations. *Adm Policy Ment Health.* 2007 jul;34(4):343–352–352.
67. Fortney JC, Xu S, Dong F. Community-Level Correlates of Hospitalizations for Persons With Schizophrenia. *Psychiatr Serv.* 2009;60(6):772–778.
68. Gabilondo A, Rojas-Farreras S, Vilagut G, Haro JM, Fernández A, Pinto-Meza A, et al. Epidemiology of major depressive episode in a southern European country: results from the ESEMeD-Spain project. *J Affect Disord.* 2010 ene;120(1-3):76–85.
69. García-Alonso CR, Guardiola J, Hérvez-Martínez C. Logistic evolutionary product-unit neural networks: innovation capacity of poor Guatemalan households. *European J Operational Res* 2009;195(2):543–551.
70. García-Alonso CR, Pérez-Naranjo LM, Fernández-Caballero JC. Multiobjective evolutionary algorithms to identify highly autocorrelated areas: the case of spatial distribution in financially compromised farms. *Ann Oper Res.* 2011;1–16.
71. García-Alonso CR, Salvador-Carulla L, Negrín-Hernández MA, Moreno-Küstner B. Development of a new spatial analysis tool in mental health: identification of highly

- autocorrelated areas (hot-spots) of schizophrenia using a Multiobjective Evolutionary Algorithm model (MOEA/HS). *Epidemiol Psichiatr Soc.* 2010 dic;19(4):302–313.
72. Garrido M, Salinas JA, Almenara J, Salvador L. Atlas de Salud Mental de Andalucía 2005. Sevilla: Servicio Andaluz de Salud; 2006.
  73. Garrido-Cumbrera M, Almenara-Barrios J, López-Lara E, Peralta-Sáez JL, García-Gutierrez JC, Salvador-Carulla L. Development and spatial representation of synthetic indexes of outpatient mental health care in Andalusia (Spain). *Epidemiol Psichiatr Soc* 2008;17(3):192–200.
  74. Geary RC. The contiguity ratio and statistical mapping. *The Incorporated Statistician* 1954;5:115–145.
  75. GEOSCAT Group. Mapa integral de recursos de salud mental de Cataluña. Barcelona: Department of Health of Catalonia; en prensa.
  76. Getis A, Mur J, Zoller HG, editores. *Spatial econometrics and spatial statistics*. Hampshire: Palgrave Macmillan; 2004.
  77. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* 1992;24(3):189–206.
  78. Getis A. Reflections on spatial autocorrelation. *Regional Science and Urban Economics* 2007;37:491–496.
  79. Gibert K, García-Alonso CR, Salvador-Carulla L. Integrating clinicians, knowledge and data: expert-based cooperative analysis in healthcare decision support. *Health Res Policy Syst.* 2010;8(1):28.
  80. Gibert K, Nonell R. Pre and post-processing in KLASS. En: *Proc. of the DM-TEs'08 Workshop at iEMSs 2008*. 2008 p. 1965–1966.
  81. Gilks WR, Richardson S, Spiegelhalter DJ, editores. *Markov chain Monte Carlo in practice*. New York: Chapman & Hall/CRC; 1996.
  82. Gittelsohn A, Powe NR. Small area variations in health care delivery in Maryland. *Health Serv Res.* 1995 jun;30(2):295–317.
  83. Goldberg DE. *Genetic algorithms in search, optimization and machine learning*. Reading: Addison Wesley; 1989.
  84. Gómez-Barroso D, Nogareda F, Cano R, Pina MF, Luis Del Barrio J, Simon F. Patrón espacial de la legionelosis en España, 2003-2007. *Gac Sanit* 2011;25(4):290–5.
  85. González Cocina E, Pérez Torres F. La historia clínica electrónica. Revisión y análisis de la actualidad. *Diraya: la historia de salud electrónica de Andalucía. Revista Española de Cardiología Suplementos* 2007;7(3):37C–46C.
  86. Goodchild MF, Kemp KK, editores. *NCGIA Core Curriculum in GIS*. Santa Barbara CA: National Center for Geographic Information and Analysis, University of California; 1990.

87. Griffith E. Geographic Information Systems (GIS) and Spatial Analysis. En: Williams M, Vogt WP, editores. *The SAGE Handbook of Innovation in Social Research Methods*. London: SAGE Publications; 2011
88. Gustavsson A, Svensson M, Jacobi F, Allgulander C, Alonso J, Beghi E, et al. Cost of disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol*. 2011 oct;21(10):718–779.
89. Haase I, Olson S, Behr MA, Wanyeki I, Thibert L, Scott A, et al. Use of geographic and genotyping tools to characterise tuberculosis transmission in Montreal. *Int J Tuberc Lung Dis* 2007;11(6):632–638.
90. Haining RP. *Spatial data analysis in the social and environmental sciences*. Cambridge: Cambridge University Press; 1993.
91. Haining RP. Spatial Autocorrelation. En: Smelser NJ, Baltes PB, editores. *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Pergamon; 2001 p. 14763–14768.
92. Haro JM, Palacín C, Vilagut G, Martínez M, Bernal M, Luque I, et al. [Prevalence of mental disorders and associated factors: results from the ESEMeD-Spain study]. *Medicina Clínica* 2006 abr;126(12):445–451.
93. Holland JH. *Adaptation in Natural and Artificial Systems*. Ann Arbor: The University of Michigan Press; 1975.
94. Holowaty E, Norwood T, Wanigaratne S, Abellan J, Beale L. Feasibility and utility of mapping disease risk at the neighbourhood level within a Canadian public health unit: an ecological study. *Int J Health Geogr*. 2010;9(1):21.
95. Horn J, Nafpliotis N, Goldberg DE. A niched Pareto genetic algorithm for multiobjective optimization. En: *Proceedings of the First IEEE Conference on Evolutionary Computation*. Orlando: IEEE; 1994 p. 82–87.
96. Huang L, Pickle LW, Das B. Evaluating spatial methods for investigating global clustering and cluster detection of cancer cases. *Stat Med* 2008;27(25):5111–5142.
97. Instituto Nacional de Estadística. Padrón municipal de habitantes [Internet]. INE 2011; [citado 2011 jul 16] Available from: <http://www.ine.es/>
98. Jackson M, Huang L, Luo J, Hachey M, Feuer E. Comparison of tests for spatial heterogeneity on data with global clustering patterns and outliers. *Int J Health Geogr*. 2009;8(1):55.
99. Jacquez GM, Goovaerts P. The emerging role and benefits of boundary analysis in spatio-temporal epidemiology and public health. *Spat Spatiotemporal Epidemiol* 2010 dic;1(4):197–200.
100. Jacquez GM, Kaufmann A, Goovaerts P. Boundaries, links and clusters: a new paradigm in spatial analysis? *Environ Ecol Stat*. 2008;15(4):403–419.

101. Jarup L. Health and environment information systems for exposure and disease mapping, and risk assessment. *Environ. Health Perspect* 2004 jun;112(9):995–997.
102. Jefferis BJ, Nazareth I, Marston L, Moreno-Kustner B, Bellón JÁ, Svab I, et al. Associations between unemployment and major depressive disorder: evidence from an international, prospective study (the predict cohort). *Soc Sci Med* 2011 dic;73(11):1627–1634.
103. Kelly BD, O’Callaghan E, Waddington JL, Feeney L, Browne S, Scully PJ, et al. Schizophrenia and the city: A review of literature and prospective study of psychosis and urbanicity in Ireland. *Schizophrenia Research* 2010 ene;116(1):75–89.
104. Khamis FG. Investigating the spatial pattern of mortality in Jordan-2008. *Am J Sci Ind Res* 2010;1(3):585–594.
105. Kirkbride JB, Fearon P, Morgan C, Dazzan P, Morgan K, Murray RM, et al. Neighbourhood variation in the incidence of psychotic disorders in Southeast London. *Soc Psychiatry Psychiatr Epidemiol.* 2007 jun;42(6):438–445.
106. Kistemann T, Dangendorf F, Schweikart J. New perspectives on the use of Geographical Information Systems (GIS) in environmental health sciences. *Int J Hyg Environ Health* 2002 abr;205(3):169–181.
107. Knowles J, Corne D. The Pareto archived evolution strategy: a new baseline algorithm for Pareto multiobjective optimisation. En: *Proceedings of the 1999 Congress on Evolutionary Computation, 1999. CEC 99. Washington: 1999*
108. Koschinsky J. The case for spatial analysis in evaluation to reduce health inequities. *Eval Program Plann.* en prensa;
109. Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic spatial scan statistic. *Stat Med* 2006;25(22):3929–3943.
110. Kulldorff M, Nagarwalla N. Spatial disease clusters: Detection and inference. *Stat Med* 1995;14(8):799–810.
111. Kulldorff M, Song C, Gregorio D, Samociuk H, DeChello L. Cancer map patterns - Are they random or not? *Am J Prev Med* 2006;30(2):S37–S49.
112. Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods* 1997;26(6):1481–1496.
113. Lavis JN, Oxman AD, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP). *Health Res Policy Syst.* 2009;7 Suppl 1:I1.
114. Lawson AB, Biggeri AB, Boehning D, Lesaffre E, Viel JF, Clark A, et al. Disease mapping models: an empirical evaluation. *Disease Mapping Collaborative Group. Statistics in medicine* 2000;19(17-18):2217–41.

115. Lee MA, Esbensen H. Fuzzy/Multiobjective genetic systems for intelligent systems design tools and components. En: Predrycz W, editor. Fuzzy evolutionary computation. Boston: Kluwer Academic Publishers; 1997
116. LeSage J, Pace RK. Introduction to Spatial Econometrics. Boca Raton (USA): CRC Press/Taylor & Francis Group; 2009.
117. Lewin S, Oxman AD, Lavis JN, Fretheim A, Garcia Marti S, Munabi-Babigumira S. SUPPORT tools for evidence-informed policymaking in health 11: Finding and using evidence about local conditions. *Health Res Policy Syst.* 2009;7 Suppl 1:S11. [citado 2012 ene 23 ]
118. Leyland AH, Davies CA. Empirical Bayes methods for disease mapping. *Stat Methods Med Res* 2005 feb;14(1):17–34.
119. Lope V, Pollán M, Pérez-Gómez B, Aragonés N, Vidal E, Gómez-Barroso D, et al. Municipal distribution of ovarian cancer mortality in Spain. *BMC Cancer* 2008;8:258.
120. López-Abente G, Aragonés N, Pérez-Gómez B, Ramis R, Vidal E, García-Pérez J, et al. Kidney cancer mortality in Spain: geographic patterns and possible hypotheses. *BMC Cancer* 2008;8:293.
121. Macintyre S, Ellaway A, Cummins S. Place effects on health: how can we conceptualise, operationalise and measure them? *Soc Sci Med* 2002 jul;55(1):125–139.
122. Marín I, Briones E. Variabilidad y gestión clínica. De cómo le puede servir el atlas a los Ulises-clínicos para superar la visión del cíclope. *Atlas de variaciones en la práctica médica* 2007;2(2):139–41. [citado 2012 may 18 ]
123. Ministerio de Sanidad y Política Social. Evaluación de la Estrategia en Salud Mental del Sistema Nacional de Salud [Internet]. Madrid: Ministerio de Sanidad y Política Social; 2009. [citado 2011 Jul 27] Disponible en: [http://www.msps.es/organizacion/sns/planCalidadSNS/pdf/excelencia/Resumen\\_Evaluacion.pdf](http://www.msps.es/organizacion/sns/planCalidadSNS/pdf/excelencia/Resumen_Evaluacion.pdf)
124. Moran PAP. The interpretation of statistical maps. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* 1948;10(2):243–251.
125. Moran PAP. Notes on continuous stochastic phenomena. *Biometrika* 1950;37(1-2):17–23.
126. Moreno B, García-Alonso CR, Negrín Hernández M, Torres-González F, Salvador-Carulla L. Spatial analysis to identify hotspots of prevalence of schizophrenia. *Soc Psychiatry Psychiatr Epidemiol.* 2008;43(10):782–791.
127. Moreno Kustner B. La información sanitaria se enreda en la informática. *Gac Sanit.* 2011;25(4):343–344.
128. Moreno Küstner B. Situación de los sistemas de información sobre salud mental en España. *Monografías de psiquiatría* 2008;20(1):4–14.



129. Moreno R, Vayá E. Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial. Barcelona: Edicions de la Universitat de Barcelona; 2000.
130. Moreno-Küstner B, Mayoral F, Rivas F, Angona P, Requena J, García-Herrera J, et al. Factors associated with use of community mental health services by schizophrenia patients using multilevel analysis. *BMC Health Services Research* 2011 oct;11(1):257.
131. Moscone F, Knapp M, Tosetti E. Mental health expenditure in England: A spatial panel approach. *J Health Econ.* 2007;26(4):842–864.
132. MRC Biostatistics Unit. The BUGS project Geobugs [Internet]. The BUGS (Bayesian inference Using Gibbs Sampling) project 2008; [citado 2012 sep 24] Available from: <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/geobugs.shtml>
133. National Institute for Occupational Safety and Health. Research and demonstration grants relating to occupational safety and health [Internet]. National Institute of Health Guide, Volume 22, Number 23; 1993. [citado 2012 sep 24] Available from: <http://grants.nih.gov/grants/guide/pa-files/PA-93-097.html>
134. Nelson KE, Williams CM, editores. *Infectious Disease Epidemiology: Theory and Practice*. Second Edition. Sudbury: Jones & Bartlett Learning; 2007.
135. Ocaña-Riola R, Mayoral-Cortés JM. Spatio-temporal trends of mortality in small areas of Southern Spain. *BMC Public Health* 2010;10:26.
136. Ocaña-Riola R. Common errors in disease mapping. *Geospat Health.* 2010 may;4(2):139–154.
137. OECD. *Creating rural indicators for shaping territorial policy*. Paris: Organisation for Economic Co-operation and Development; 1994.
138. Oliva-Moreno J, López-Bastida J, Montejo-González AL, Osuna-Guerrero R, Duque-González B. The socioeconomic costs of mental illness in Spain. *European Journal of Health Economics* 2009;10(461-369)
139. Olivet M, Aloy J, Prat E, Pons X. Oferta de servicios de salud y accesibilidad geográfica. *Medicina Clínica* 2008 dic;131(Supplement 4):16–22. [citado 2011 feb 18 ]
140. Ord JK, Getis A. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis* 1995;27(4):186–306.
141. Ostfeld RS, Glass GE, Keesing F. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology and Evolution* 2005;20(6):328–336.
142. Osyczka A. *Multicriterion optimization in engineering with FORTRAN programs*. New York: John Wiley & Sons; 1984.
143. Osyczka A. *Evolutionary Algorithms for Single and Multicriteria Design Optimization*. Berlin: Physica-Verlag; 2002.
144. Owen MJ. Implications of genetic findings for understanding schizophrenia. *Schizophr Bull* 2012 sep;38(5):904–907.

145. Paelinck J, Klaassen L. *Spatial Econometrics*. Farnborough: Saxon House; 1979.
146. Pareto V. *Cours d'Economie Politique*. Laussane: F. Rouge; 1896.
147. Patel A, Knapp M. Costs of mental illness in England. *Ment Health Res Rev* 1998;5:4–10.
148. Patil GP, Taillie C. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environ Ecol Stat* 2004;11(2):183–197.
149. Pedro-Cuesta J de, Rodríguez-Farré E, Lopez-Abente G. Spatial distribution of Parkinson's disease mortality in Spain, 1989-1998, as a guide for focused aetiological research or health-care intervention. *BMC Public Health* 2009;9:445.
150. Perälä J, Saarni SI, Ostamo A, Pirkola S, Haukka J, Härkänen T, et al. Geographic variation and sociodemographic characteristics of psychotic disorders in Finland. *Schizophr. Res* 2008 dic;106(2-3):337–347.
151. Pollán M, Ramis R, Aragones N, Perez-Gomez B, Gomez D, Lope V, et al. Municipal distribution of breast cancer mortality among women in Spain. *BMC Cancer* 2007;7(1):78.
152. Qi X, Tong S, Hu W. Spatial distribution of suicide in Queensland, Australia. *BMC Psychiatry* 2010;10:106.
153. Rezaeian M, Dunn G, St Leger S, Appleby L. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *J Epidemiol Community Health*. 2007 feb;61(2):98–102.
154. Rosenberg RS. Simulation of genetic populations with biochemical properties: II. selection of crossover probabilities. *Mathematical Biosciences* 1970;8(1-2):1–37.
155. Ross ME, Wakefield J, Davis S, De Roos AJ. Spatial clustering of myelodysplastic syndromes (MDS) in the Seattle-Puget Sound region of Washington State. *Cancer Causes Control* 2010 jun;21(6):829–838.
156. Sabes-Figuera R, Knapp M, Bendeck M, Mompert-Penina A, Salvador-Carulla L. The local burden of emotional disorders. An analysis based on a large health survey in Catalonia (Spain). *Gac Sanit*. 2012;26(1):24–29.
157. Salinas-Pérez JA, García-Alonso CR, Molina-Parrilla C, Jordà-Sampietro E, Salvador-Carulla L. Identification and location of hot and cold spots of treated prevalence of depression in Catalonia (Spain). *Int J Health Geogr*. 2012;11:36.
158. Salvador-Carulla L, Costa-Font J, Cabases J, McDaid D, Alonso J. Evaluating mental health care and policy in Spain. *J Ment Health Policy Econ*. 2010;13(2):73–86.
159. Shekhar S, Evans MR, Kang JM, Mohan P. Identifying patterns in spatial information: A survey of methods. *WIREs Data Mining Knowl Discov* 2011;1(3):193–214.
160. Smith MJD, Goodchild MF, Longley P. *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*. Leicester: Troubador Publishing Ltd; 2007.

161. Snow J. On the mode of communication of cholera. New York: The Commonwealth Fund; 1855.
162. Sobocki P, Jönsson B, Angst J, Rehnberg C. Cost of depression in Europe. *J Ment Health Policy Econ* 2006;9(2):87–98.
163. Song C, Kulldorff M. Power evaluation of disease clustering tests. *Int J Health Geogr.* 2003;2(1):9.
164. Sridharan S, Koschinsky J, Walker JJ. Does context matter for the relationship between deprivation and all-cause mortality? The West vs. the rest of Scotland. *Int J Health Geogr.* 2011 may;10(1):33.
165. Sridharan S, Tunstall H, Lawder R, Mitchell R. An exploratory spatial data analysis approach to understanding the relationship between deprivation and mortality in Scotland. *Soc Sci Med.* 2007;65(9):1942–1952.
166. Srinivas N, Deb K. Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms. *Evolutionary Computation* 2011 sep;2(3):221–248.
167. Staines A, Järup L. Health event data. En: Elliot P, Wakefield J, Cressie N, Briggs D, editores. *Spatial Epidemiology*. New York: Oxford University Press; 2000
168. Student. The elimination of spurious correlation due to position in time or space. *Biometrika* 1914;10:179–180.
169. Sugumaran R, Larson S, DeGroot J. Spatio-temporal cluster analysis of county-based human West Nile virus incidence in the continental United States. *Int J Health Geogr.* 2009;8(1):43.
170. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geogr.* 2005;4:11.
171. Tango T. A class of tests for detecting general and focused clustering of rare diseases. *Stat Med* 1995;14(21-22):2323–2334.
172. Tobler WR. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 1970 jun;46:234–240.
173. Tsai P-J, Lin M-L, Chu C-M, Perng C-H. Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. *BMC Public Health* 2009;9(1):464.
174. Wakefield J, Best NG, Waller L. Bayesian approaches to disease mapping. En: Elliot P, Wakefield J, Best N, Briggs D, editores. *Spatial Epidemiology*. New York: Oxford University Press; 2000
175. Wakefield J, Kelsall JE, Morris SE. Clustering, cluster detection, and spatial variations in risk. En: Elliot P, Wakefield J, Best N, Briggs D, editores. *Spatial Epidemiology*. New York: Oxford University Press; 2000
176. Wakefield J. Disease mapping and spatial regression with count data. *Biostatistics* 2007 abr;8(2):158 –183.

177. Walter SD. Disease mapping: a historical perspective. En: Elliot P, Wakefield J, Cressie N, Briggs D, editores. *Spatial Epidemiology*. New York: Oxford University Press; 2000
178. Wang J, Terpenney JP. Interactive preference incorporation in evolutionary engineering design. En: Jin Y, editor. *Knowledge incorporation in evolutionary computation*. Berlin: Springer; 2005
179. Ward M. Spatial Epidemiology: Where Have We Come in 150 Years? En: Sui DZ, editor. *Geospatial Technologies and Homeland Security*. Dordrecht: Springer; 2008 p. 257–282.
180. Weiss RA, McMichael AJ. Social and environmental risk factors in the emergence of infectious diseases. *Nat Med* 2004;10:S70–S76.
181. Westercamp N, Moses S, Agot K, Ndinya-Achola J, Parker C, Amolloh K, et al. Spatial distribution and cluster analysis of sexual risk behaviors reported by young men in Kisumu, Kenya. *Int J Health Geogr*. 2010;9(1):24.
182. Whitley R, McKenzie K. Social capital and psychiatry: review of the literature. *Harv Rev Psychiatry* 2005 abr;13(2):71–84.
183. WHO. *The ICD-10 Classification of Mental and Behavioural Disorders. Diagnostic criteria for research*. Geneva: World Health Organization; 1993.
184. Wilde A, Mitchell PB, Meiser B, Schofield PR. Implications of the use of genetic tests in psychiatry, with a focus on major depressive disorder: A review. *Depress Anxiety* 2012 sep;
185. Wolpert DH, Macready WG. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1997;1(1):67–82.
186. Yeshiwondim A, Gopal S, Hailemariam A, Dengela D, Patel H. Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia. *Int J Health Geogr*. 2009;8(1):5.
187. Zhen H, McDermott S, Lawson A, Aelion M. Are clusters of mental retardation correlated with clusters of developmental delay? *Geospat Health*. 2009;4(1):17–26.
188. Zisook S, McAdams LA, Kuck J, Harris MJ, Bailey A, Patterson TL, et al. Depressive symptoms in schizophrenia. *Am J Psychiatry* 1999 nov;156(11):1736–1743. [citado 2012 jul 8 ]
189. Zitzler E, Laumanns M, Thiele L. SPEA2: Improving the strength of Pareto evolutionary algorithm. En: Giannakoglou K, Tsahalis D, Periaux J, et al, editores. *Evolutionary methods for design, optimization and control with applications to industrial problems*. Athens: EUROGEN 2001; 2001
190. Zitzler E, Thiele L. *An evolutionary algorithm for multiobjective optimization: The strength Pareto approach*. Zurich: Swiss Federal Institute of Technology; 1998.

191. Zitzler E, Thiele L. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation* 1999 nov;3(4):257–271.
192. Zydallis JB, Veldhuizen DA, Lamont GB. A Statistical Comparison of Multiobjective Evolutionary Algorithms Including the MOMGA-II. En: Zitzler E, Thiele L, Deb K, Coello Coello CA, Corne D, editores. *Evolutionary Multi-Criterion Optimization*. Berlin: Springer; 2001 p. 226–240.

## **Anexo 1**



**CUESTIONARIO A EXPERTOS SOBRE LAS HERRAMIENTAS DE REPRESENTACIÓN ESPACIAL, ATLAS DE SALUD Y EL MODELO DE ANÁLISIS ESPACIAL AEMO/AE PARA LA TOMA DE DECISIONES EN POLÍTICA Y PLANIFICACIÓN SANITARIA**

---

**Nombre:**

**Profesión:**

**Grupo experto:**

*Marcar con una X*

Experto en investigación de servicios sanitarios

Experto en política y gestión de servicios sanitarios

Ambos grupos

<input type="checkbox"/>
<input type="checkbox"/>
<input type="checkbox"/>

**Años de experiencia:**

**Conceptos previos al cuestionario:**

*Utilidad:* Provecho, conveniencia, interés o fruto que se saca de algo (RAE).

*Relevancia:* Calidad o condición de relevante, importancia, significación (RAE).

*Mapas y representaciones espaciales:* representación cartográfica de datos de interés para la toma de decisiones en política y planificación sanitaria. Pueden ser datos sociodemográficos, morbilidad, actividad asistencial, disponibilidad y accesibilidad a los servicios sanitarios, etc.

*Modelo AEMO/AE:* Modelo del Algoritmo Evolutivo Multi-objetivo para el análisis espacial aplicado a datos de morbilidad asistida (prevalencia administrativa) con el fin de detectar y localizar concentraciones espaciales de valores significativamente elevados (*hot-spots*) y/o bajos (*cold-spots*).



## CUESTIONES

---

1. **¿Cómo valora la utilidad práctica del uso de mapas y representaciones espaciales para la toma de decisiones en política y planificación sanitaria?**

*Marque con una X*

Muy baja	Baja	Media	Alta

*Puntúe también de 0 (muy baja) a 10 (alta)*

2. **¿Cómo valora la relevancia del uso de mapas y representaciones espaciales para la toma de decisiones en política y planificación sanitaria?**

*Marque con una X*

Muy baja	Baja	Media	Alta

*Puntúe también de 0 (muy baja) a 10 (alta)*

3. **¿Cómo valora la utilidad práctica de los Atlas de Salud para la toma de decisiones en política y planificación sanitaria?**

*Marque con una X*

Muy baja	Baja	Media	Alta

*Puntúe también de 0 (muy baja) a 10 (alta)*

4. **¿Cómo valora la relevancia de los Atlas de Salud para la toma de decisiones en política y planificación sanitaria?**

*Marque con una X*

Muy baja	Baja	Media	Alta

*Puntúe también de 0 (muy baja) a 10 (alta)*

5. **¿Cómo valora la utilidad práctica del modelo de análisis espacial AEMO/AE para la toma de decisiones en política y planificación sanitaria?**

*Marque con una X*

Muy baja	Baja	Media	Alta

*Puntúe también de 0 (muy baja) a 10 (alta)*

6. **¿Cómo valora la relevancia de la información aportada por el modelo de análisis espacial AEMO/AE respecto a la información previa disponible para la toma de decisiones en política y planificación sanitaria?**

*Marque con una X*

Muy baja	Baja	Media	Alta

*Puntúe también de 0 (muy baja) a 10 (alta)*

7. **¿Qué otras variables sanitarias sería interesante analizar mediante el modelo AEMO/AE en su opinión?**

Mortalidad	
Incidencia	
Frecuentación	
Altas	
Ingresos	
Reingresos	
Estancia media	
Otras ( <i>especificar</i> )	

## Listado de expertos en gestión planificación e investigación en servicios de salud mental

### ANDALUCÍA

Berta Moreno  
Universidad de Málaga

Cristina Romero  
Universidad de Cádiz

Federico Alonso  
Instituto Andaluz de Servicios Sociales  
(Sevilla)

José Almenara  
Universidad de Cádiz

Juan Luis González  
Universidad de Cádiz

Mencía Ruiz  
Asociación Científica PSICOST  
(Cádiz)

### CANTABRIA

Luis Gaité  
Hospital Universitario Marqués de Valdecilla  
(Santander, Cantabria)

### CATALUÑA

Ana Fernández  
Fundació Sant Joan de Déu  
(San Boi de Llobregat, Barcelona)

Antonio Bulbena  
Institut de Neuropsiquiatria i Addiccions  
Parc de Salut Mar (Barcelona)

Antonio Serrano  
Fundació Sant Joan de Dèu  
(San Boi de Llobregat, Barcelona)

Enrique Aragonés  
Departament de Salut de Catalunya  
(Barcelona)

Esther Jordà  
Departament de Salut de Catalunya  
(Barcelona)

Jordi Cid  
Institut Assistencial Sociosanitari  
(Girona)

Josep Fustè  
Departament de Salut de Catalunya  
(Barcelona)

### MADRID

Miriam Poole  
Asociación Científica PSICOST  
(Madrid)

### PAÍS VASCO

Álvaro Iruin  
Red de Salud Mental de Gipuzkoa  
(San Sebastián, Guipúzcoa)

Carlos José Pereira  
Red de Salud Mental de Bizkaia  
(Bilbao, Vizcaya)

## **Anexo 2**



RESEARCH

Open Access

# Identification and location of hot and cold spots of treated prevalence of depression in Catalonia (Spain)

José A Salinas-Pérez<sup>1\*</sup>, Carlos R García-Alonso<sup>1</sup>, Cristina Molina-Parrilla<sup>2</sup>, Esther Jordà-Sampietro<sup>2</sup> and Luis Salvador-Carulla<sup>3</sup> for the GEOSCAT Group

## Abstract

**Background:** Spatial analysis is a relevant set of tools for studying the geographical distribution of diseases, although its methods and techniques for analysis may yield very different results. A new hybrid approach has been applied to the spatial analysis of treated prevalence of depression in Catalonia (Spain) according to the following descriptive hypotheses: 1) spatial clusters of treated prevalence of depression (hot and cold spots) exist and, 2) these clusters are related to the administrative divisions of mental health care (catchment areas) in this region.

**Methods:** In this ecological study, morbidity data per municipality have been extracted from the regional outpatient mental health database (CMBD-SMA) for the year 2009. The second level of analysis mapped small mental health catchment areas or groups of municipalities covered by a single mental health community centre. Spatial analysis has been performed using a Multi-Objective Evolutionary Algorithm (MOEA) which identified geographical clusters (hot spots and cold spots) of depression through the optimization of its treated prevalence. Catchment areas, where hot and cold spots are located, have been described by four domains: urbanicity, availability, accessibility and adequacy of provision of mental health care.

**Results:** MOEA has identified 6 hot spots and 4 cold spots of depression in Catalonia. Our results show a clear spatial pattern where one cold spot contributed to define the exact location, shape and borders of three hot spots. Analysing the corresponding domain values for the identified hot and cold spots no common pattern has been detected.

**Conclusions:** MOEA has effectively identified hot/cold spots of depression in Catalonia. However these hot/cold spots comprised municipalities from different catchment areas and we could not relate them to the administrative distribution of mental care in the region. By combining the analysis of hot/cold spots, a better statistical and operational-based visual representation of the geographical distribution is obtained. This technology may be incorporated into Decision Support Systems to enhance local evidence-informed policy in health system research.

**Keywords:** Spatial analysis, Hot spots, Cold spots, Mental health, Depression, Catalonia

\* Correspondence: jsalinas@uloyola.es

<sup>1</sup>Universidad Loyola Andalucía, Business Administration Faculty, Sevilla, Córdoba, Spain

Full list of author information is available at the end of the article

## Background

Spatial epidemiology is aimed at identifying patterns in the geographical distribution of health data. It may detect irregularities such as spatial clusters of a particular disease [1,2], for example, where a specific disease has significant high or low prevalence [3]. Methods for the study of spatial clusters include global spatial autocorrelation, Local Indicators of Spatial Association (LISA), spatial regression, spatial scan statistics and Bayesian inference [4].

There are numerous examples of spatial data analysis performed on health variables, such as prevalence, incidence and mortality [5]. In mental health, for example, Bayesian models have been used to study the relationship between poverty and social isolation, and psychiatric admission rates in acute hospitals in small urban areas of London and New York [6]; the variation in the incidence of psychotic disorders in urban areas in South-east London [7]; the relationship between depression and schizophrenia admission rates and socioeconomic characteristics in the counties of 14 States in the USA [8,9]; and the study of the correlation between mental retardation and clusters of developmental delay [10]. Spatial scan statistics have been used to detect clusters of mental disorders due to psychoactive substance use, and neurotic, stress-related, and somatoform disorders, and their relationship to poverty and neighbourhood social disorganization in Malmö (Sweden) [11]. LISA were applied to analyze spatial patterns of mental health in the slums of Dhaka (Bangladesh) [12]. In addition, a spatial regression model has been used to analyze spatial allocation in mental health expenditure in England [13].

However, the studies on spatial analysis show significant problems with respect to comparability, reproducibility and generalization since different methods and techniques produce different results [14,15]. We previously developed and tested [3] a Multi-Objective Evolutionary Algorithm (MOEA) that hybridised three LISA methods (Moran's *I*, Geary's *C* and Getis and Ord's *G*) and Bayesian inference to detect schizophrenia hot spots (geographical clusters of spatial units –municipalities– with significantly high rates of selected indicators of a given disease) in Andalusia (Spain). Although this hybrid technique proved to be highly effective for this aim, there were problems when trying to precisely identify the location, shapes and boundaries of the spots, as also commonly occurs with other methods of spatial analysis [16,17].

This study has incorporated the identification of cold spots (geographical clusters of spatial units –municipalities– with significantly low rates of treated prevalence of a given disease) into the spatial analysis of the regional mental health system in Spain. The presence of both spatial clusters were analysed using the outpatient mental health database in Catalonia (Spain).

This paper aims to obtain a precise identification and geographical location of hot and cold spots of treated prevalence of depression and check if they have any spatial relationship with the administrative (catchment areas) divisions of mental health care in Catalonia in order to facilitate evidence to enable well-informed policy decisions. The related descriptive hypotheses are: 1) spatial clusters of treated prevalence of depression (hot and cold spots) exist and, 2) these clusters are related to the administrative divisions of mental health care (catchment areas) in Catalonia.

## Methods

### Design

This ecological study explores the geographical distribution of depression in 946 Catalonia municipalities (considered as our spatial units in this analysis) in 2009. Catalonia is a broad region in North Eastern Spain with 7.5 million inhabitants. It is one of the most developed Spanish regions with GDP 123.75 Purchasing Power Parity (European Union one = 100) [18]. Its public health system is universal with separate planning and provision and includes both public and private organisations under contract agreements with the public health system [19]. Mental health care in Catalonia is organized territorially in 74 small catchment areas coordinated by a reference Mental Health Community Centre (MHCC). These outpatient mental health centres follow a community care model, and are coordinated with primary care, specialized hospital care and intermediate care services.

We selected the municipalities as spatial units for precise geographical identification and location of hot and cold spots of treated prevalence of depression. These are the smallest areas where reliable statistical information can be found. Mental health catchment areas were selected as secondary units of analysis. Seven urban municipalities comprise more than one mental health area. All other mental health catchment areas include several municipalities.

### Database

Psychiatric cases assisted in the 74 Adult MHCC -catchment areas- in Catalonia are registered in the Minimum Data Set for Outpatient Mental Health Centres (CMBD-SMA) [20]. This study has used the 2009 database provided by the Catalanian Department of Health, safeguarding the privacy of the patients by using anonymous registers analysed at a municipality level to prevent geographical identification of individual cases. The database collects data from anonymous patients: gender, age, residence municipality, diagnosis, activity types, date of admission and discharge, etc. The variables used to calculate de treated prevalence were: sex, age, municipality of residence and main diagnosis (single episode, depressive

disorder (F32) and recurrent-episode depressive disorder (F33) (ICD-10) [21]).

The CMBD-SMA 2009 database comprises information about 214,000 patients in total. A preliminary analysis removed 0.8% of them because the selected variables were incomplete or erroneous. Furthermore, 7 catchment areas did not provide complete information for that year. The final number of depressive patients analysed in this study was 24,580. The number of inhabitants (year 2009) in each municipality was obtained from the municipal census. Patient sex and age provided information to calculate the standard rates of treated prevalence of depression (per 1,000 population) through the direct method [22] that took into consideration the population of Catalonia.

Catchment areas in Catalonia have been described using four domains: urbanicity, service availability, accessibility to care and adequacy or appropriateness. These domains have been used in previous studies about the spatial distribution of mental illnesses prevalence. If hot/cold spots are spatially associated with specific catchment areas, it could be relevant to analyse if they are mainly rural or urban, if their accessibility is high or not and so on. The urbanicity level can be 'predominantly urban' when 85% of the inhabitants reside in municipalities whose density is greater than 150 inhabitants/km<sup>2</sup>, 'significantly rural' when this percentage is between 50% and 84%, and 'predominantly rural' when it is lower than 50% [23]. The accessibility to the MHCC of each catchment area was assessed using a standard Geographical Information System (GIS) in Catalonia (minutes by car to the corresponding MHCC from the less accessible zone of the catchment area) [24]. MHCC availability was measured by the rate of outpatient MHCC per 100,000 inhabitants. The adequacy of the provision of services in the mental health catchment areas was assessed by a group of PSICOST experts using information from the Mental Health Atlas of Catalonia [25]. Experts rated every catchment area by 7 levels of provision (very high, high, medium high, medium, medium low, low and very low). This rating was represented in semaphore scale and agreed with official ones from the Department of Health of Catalonia.

An exploratory spatial analysis on treated prevalence of depression in Catalonia was carried out to check whether its geographical distribution is distributed at random or not. Both global Moran's *I* and Getis & Ord's *G* were used [26,27].

#### **A Multi-Objective Evolutionary Algorithm (MOEA) applied to spatial data analysis**

The full technical aspects of the MOEA model are described elsewhere [3,28]. MOEA are tools used to solve complex and usually non-linear multi-objective

problems through optimization to achieve feasible and non-dominated efficient solutions [29]. The processes of optimization in MOEA are based on artificial intelligence techniques (evolutionary algorithms) and solutions (in our case, potential hot and cold spots) that are evaluated by means of different types of equations called fitness functions. These fitness functions assess the corresponding fitness degree of the solutions found in each run of the algorithm and they are designed by the objectives selected in the specific study (for example, the mean of treated prevalence of depression in a set of municipalities has to be maximized to identify spatial hot spots). The fitness value obtained represents the degree of agreement among the objectives selected to design the fitness function (improving one specific objective can lead to the worsening of another). MOEA improves solutions iteratively; in each run new and better solutions are obtained through classical genetic operators based on Nature: selection, mutation and crossover. Thus, the solution of the multi-objective problem is not unique, as there are many efficient solutions in response to the problem.

Our MOEA was designed to search for efficient solutions (potential hot and cold spots) by means of the optimization of three objectives that defined the fitness functions. MOEA analyses 100 sets of '*n*' municipalities (*n* = 10) identified by their standard codes (ie. [14004, 28097, 7009]). The initial group of 100 sets of 10 municipalities is selected at random and the improvement process starts. The standard genetic operators (selection, mutation, replication and elitism) are systematically used to improve the values of the fitness functions for each of the mentioned 100 sets of 10 municipalities (each set has their own values for the fitness functions). For example, mutation changes one municipality code in a specific set by other completely different (usually geographically close to the rest). This process stops when the values of the fitness functions for all the sets cannot be improved by the MOEA. In order to guarantee unexpected bias, the global process is repeated five times (five different initial groups of 100 sets of 10 municipalities). The objectives that structure the fitness functions were:

- Maximize (for hot spots) or minimize (for cold spots) the mean of the treated prevalence of depression (*P*) in a set of *n* municipalities (*n* = 10). For hot spots:  $\text{Max}\bar{P}$  and for cold spots:  $\text{Min}\bar{P}$ .
- Minimize the Standard Deviation (*SD*) of the treated prevalence of depression  $SD_P$  in the same set of *n* municipalities (*n* = 10). For both hot and cold spots:  $\text{Min}SD_P$ .
- Minimize the minimum distance *MinD* that links all the municipalities in the solution. For both hot and cold spots:  $\text{Min}MinD$ .

Four fitness functions were designed combining the three objectives for both hot spots:  $\text{Max}\bar{P}$ ,  $\text{Min}SD_p$ , and  $\text{Min}MinD$ ; and for cold spots:  $\text{Min}\bar{P}$ ,  $\text{Min}SD_p$  and  $\text{Min}MinD$ . These fitness functions were: Fine-Grained strength Pareto, weighted objectives, standard ranking selection and fuzzy evaluation of weighted objectives [3]. The procedure is summarized in Figure 1. MOEA initially analyses 100 sets of  $n$  municipalities ( $n = 10$ ) –five times, but during the improvement process –genetic operators- the algorithm selects the best ones and, at the end, it gives an unpredictable –less than 100- number of  $n$  municipalities sets. The results are sets of municipalities where the mean of the treated prevalence of depression is high (potential hot spots) or low (potential cold spots), the standard deviation of their prevalence is low and, finally, the minimum distance that links all the  $n$  municipalities is low (therefore confirming the assumption that they are geographically close together).

Spatial units –municipalities- that appeared the most frequently, from a statistical point of view, in the potential hot and cold spots were selected as the final solution of the model: final hot and cold spots. This selection was performed using a standard procedure for identifying extreme values in a statistical distribution (Q-Q Plot method) [30]. The threshold values vary for each fitness function according to the calculated statistical frequency of the municipalities; the Q-Q Plot method selects the municipality from which the exponential and/or Pareto model provides a

plausible statistical fit for the distribution of frequencies obtained. Spatial units that rarely appeared in the spots were not included in the final solution because the statistical analysis considered them to be spurious results. Hot and cold spots were finally mapped using the Geographical Information System (GIS) ArcGIS 9©.

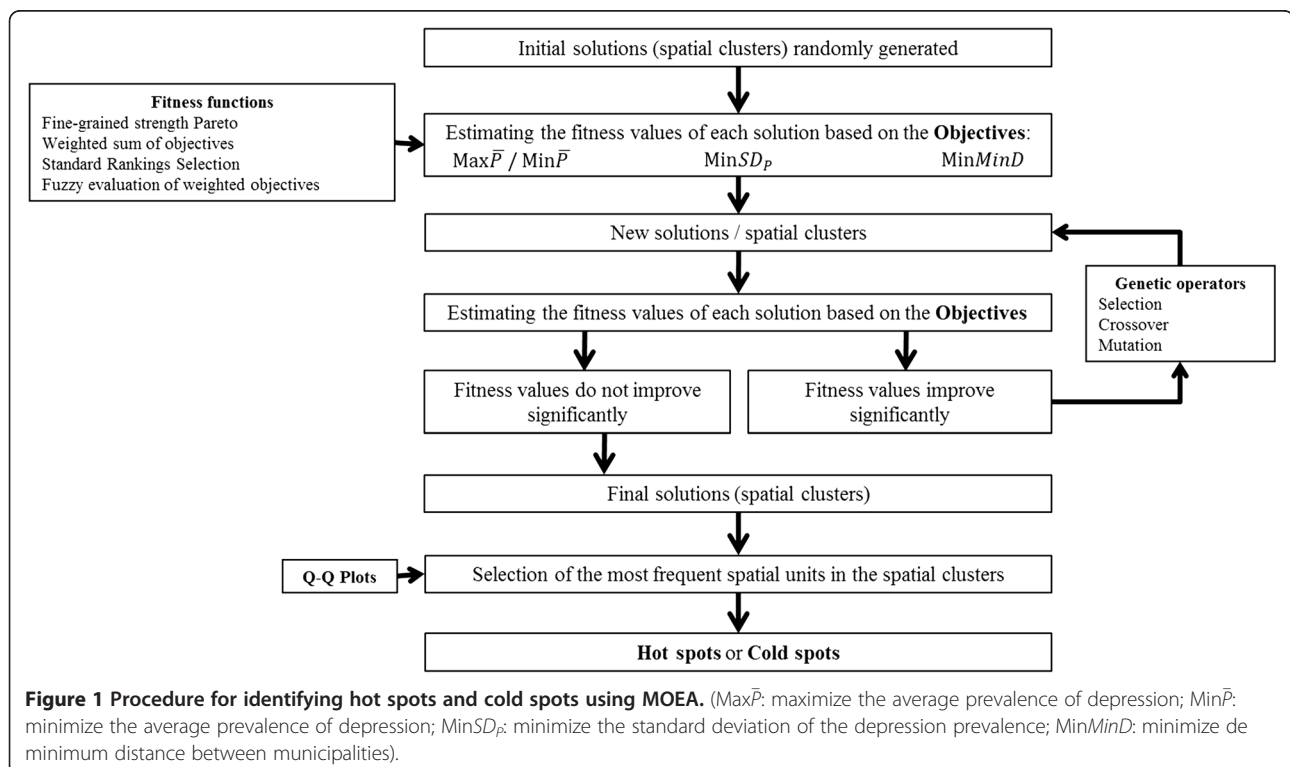
In order to check the differences in the statistical distributions of treated prevalence of depression in the hot/cold spot and in the rest of Catalonia, Kruskal-Wallis' one-way analysis of variance and Mann-Whitney  $U$  were used.

## Results and discussion

### Hot and cold spots of treated prevalence of depression

Spatial analysis searching for geographical patterns is of growing importance in epidemiology and in the evidence-informed paradigm which regards local data as a critical component for generating knowledge for planning and health policy. As stated by Lewin and colleagues, the evidence nearest to health decision-makers is that which informs about local conditions in their environment and is necessary to judge what decisions and actions must be taken in health policy [31]. It is also important to apply these techniques to study mental disorders such as depression, given its impact on the cost and burden of diseases [32,33] and the scarcity of prior information on spatial analysis in these conditions.

Treated prevalence of depression in Catalonia (year 2009) was 3.3 per 1,000 population. The standardized treated





prevalence of depression per municipality is shown in Figure 2. Five statistical classes based on standard deviation have been represented. There is a higher prevalence of depression located in the central northern region and in many disperse areas. Although the spatial distribution of the prevalence does not show a clear territorial pattern, it cannot be attributed to a random effect (Moran's  $I=0.19$ ,  $z=12.5$ ,  $\alpha \leq 0.01$ ; Getis & Ord's  $G=0$ ,  $z=12.05$ ,  $\alpha \leq 0.01$ ).

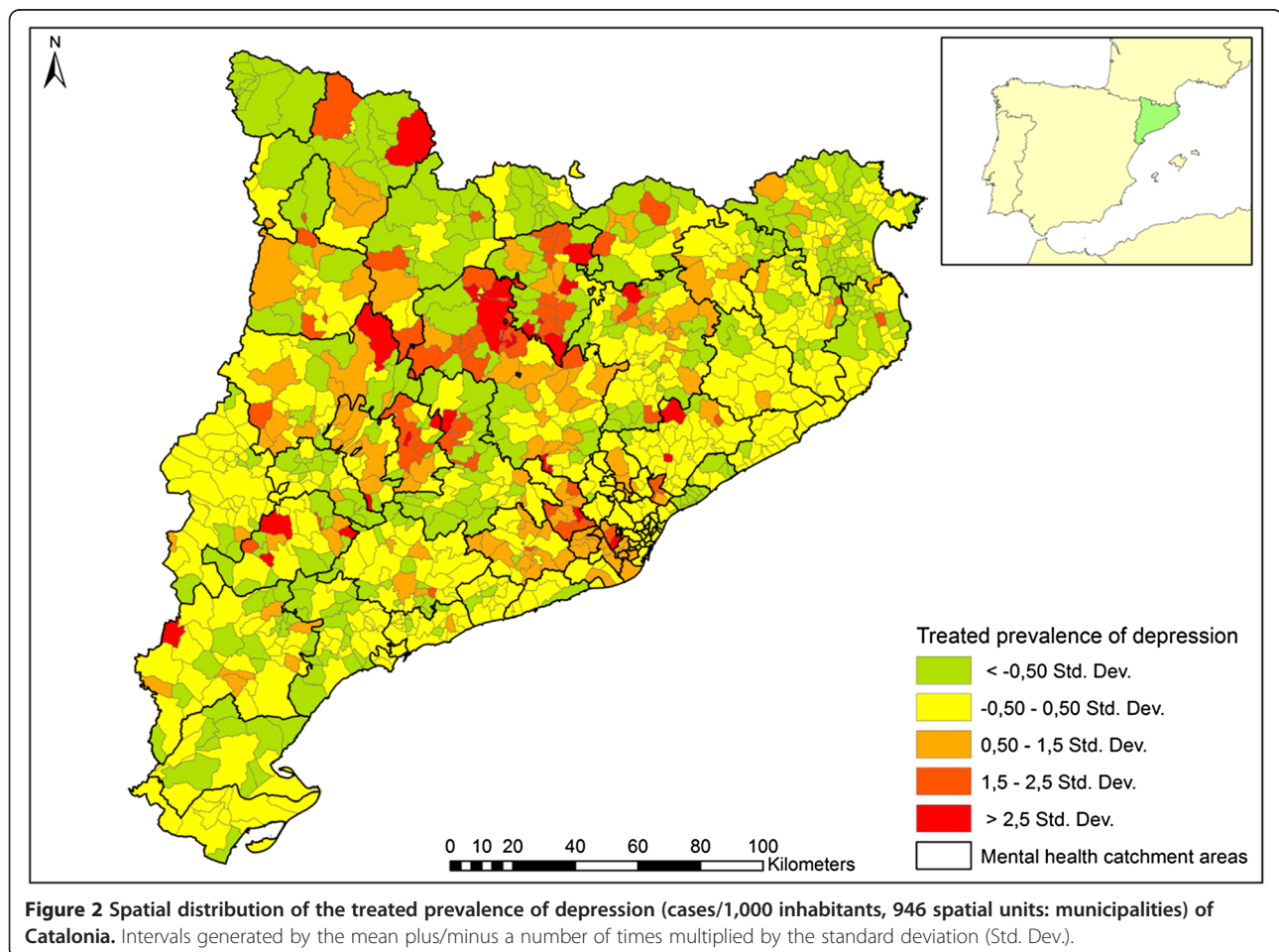
Hot spots and cold spots of treated prevalence of depression are represented in Figure 3. Five hot spots (HS1-5) and one isolated municipality (HS6) have been found. Additionally, three cold spots (CS1-3) plus a radial cluster of several municipalities (CS4) have been identified by the model. The radial cluster CS4 delimits the hot spots HS1, HS2 and HS3. Two well-defined hot spots (HS4 and HS5) are clearly identified in Figure 3. One of them (HS4) is adjacent to CS2, while the other cold spots (CS1 and CS3) are completely isolated.

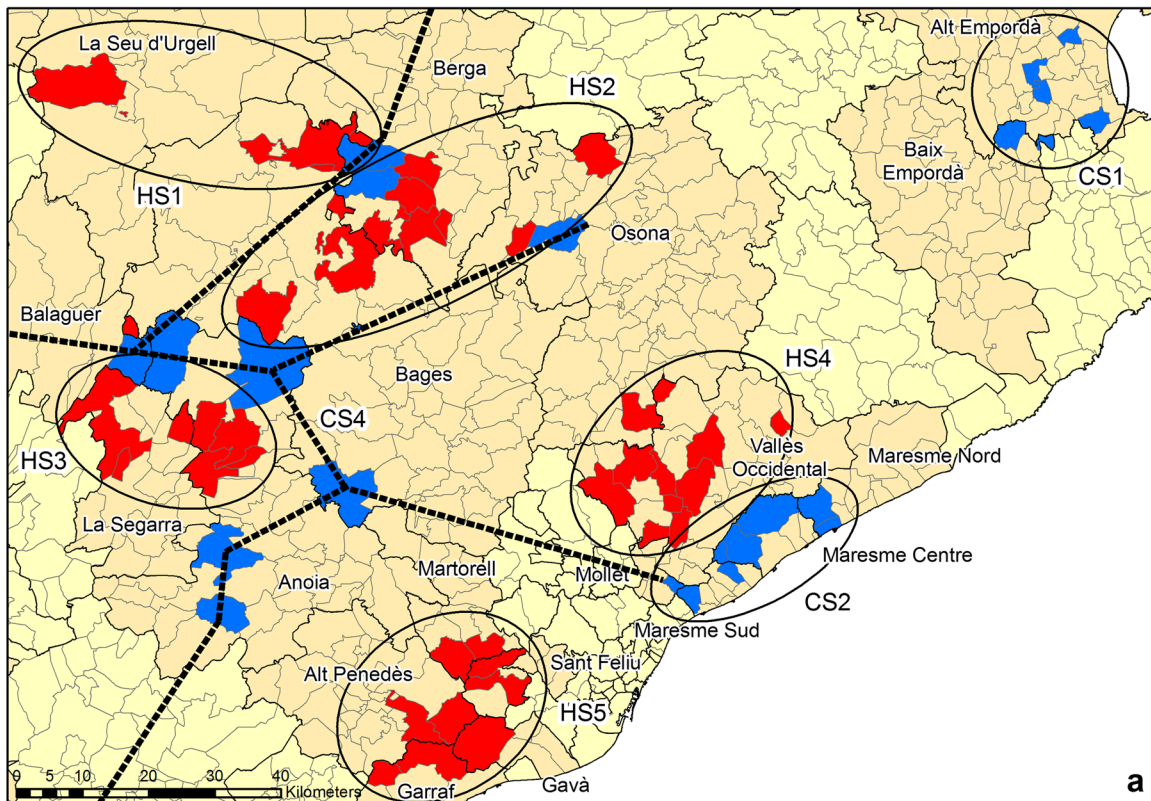
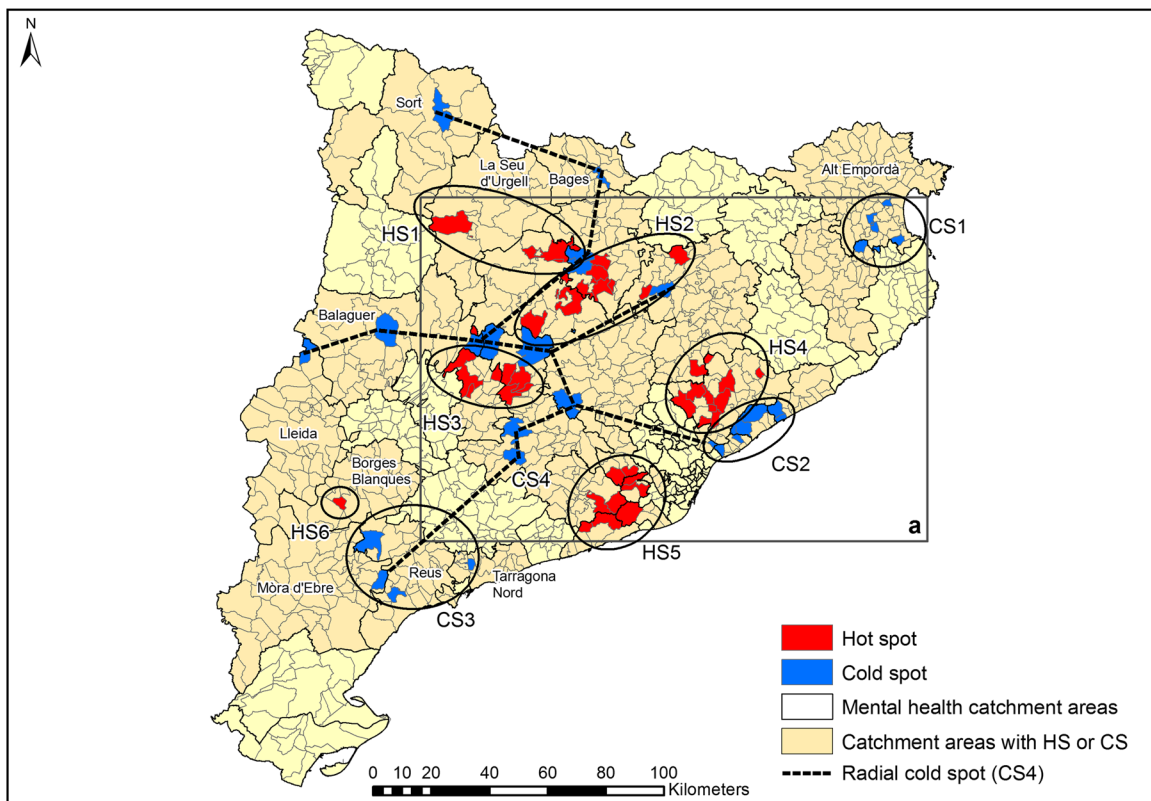
Table 1 shows the basic statistics of the treated prevalence of depression in the hot/cold spots identified, as well as the catchment areas. There are two hot spots with a mean prevalence over 10 per 1,000 inhabitants (HS2 and HS3). It is important to note that in these hot spots the

standard deviations of prevalence are also very high as both hot spot group municipalities have very different rates of depression. On the other hand, HS1, HS4 and HS5 show lower standard deviations. The mean and standard deviation of some cold spots are zero (CS1 and CS3) or very low (CS2 and CS4). In these cases, our methodology was seeking geographical zones with very low prevalence, so the algorithm was successful in identifying cold spots.

Some spatial units with small populations –numbers of inhabitants– have been included in hot/cold spots. Treated prevalence might show a high longitudinal variation in areas with this characteristic where a variation of a few patients can greatly influence the overall treated prevalence. This behaviour might be due to a random variation in depression patients throughout the time span. MOEA searches clusters of close spatial units in the space, so the appearance of significantly high or low treated prevalence cannot be considered to be due to random effects and shows potential areas of interest for decision makers.

Statistical tests indicate that there are significant differences between hot and cold spots in the distribution of treated prevalence of depression in comparison with the rest of the municipalities in Catalonia ( $\alpha \leq 0.05$ ). Hot/





**Figure 3** Spatial distribution of hot spots and cold spots of depression treated prevalence.

**Table 1 Basic statistics, geographical location and characteristics of the hot/cold spots of treated prevalence of depression in Catalonia (2009) (946 municipalities)**

SU	Number of municipalities and basic statistics in HS/CS	Location of HS/CS (Catchment area name)	Urbanicity of catchment areas where HS/CS are located (Type)	Availability of MHCC [25] (Rate per 100,000 pop.)	Road accessibility to a MHCC in catchment areas in minutes [24]	Adequacy of provision of mental health services [25]
HS1	<i>N: 3 m; Mean: 9.9; Median: 8.3; St. Dev: 3.4</i>	Bages	Significantly rural	0.6	0 - 60	Very high
		Seu d'Urgell	Significantly rural	5.6	0 - >60	Low
HS2	<i>N: 8 m; Mean: 12.0; Median: 9.0; St. Dev: 10.1</i>	Bages	Significantly rural	0.6	0 - 60	Very high
		Berga	Significantly rural	3.0	0 - 45	High
		Osona	Predominantly urban	0.8	0 - 45	Very high
HS3	<i>N: 7 m; Mean: 11.6; Median: 9.6; St. Dev: 6.6</i>	Anoia	Significantly rural	1.1	0 - 60	Low
		La Segarra	Significantly rural	6.0	0 - 45	Medium
HS4	<i>N: 11 m; Mean: 4.6; Median: 3.3; St. Dev: 3.2</i>	Vallès Oriental	Predominantly urban	0.4	0 - 45	Very high
		Osona	Predominantly urban	0.8	0 - 45	Very high
HS5	<i>N: 9 m; Mean: 5.2; Median: 5.4; St. Dev: 2</i>	Alt Penedès	Significantly rural	1.3	0 - 45	High
		Garraf	Predominantly urban	0.8	0 - 30	High
		Gavà	Predominantly urban	1.1	0 - 15	High
		Martorell	Predominantly urban	0.8	0 - 30	Very high
HS6	<i>N: 1 m; P: 13.9</i>	Sant Feliu	Predominantly urban	1.1	0 - 30	Medium
		Borges Blanques	Predominantly rural	6.0	0 - 45	Medium
		-	-	-	-	-
No HS	<i>N: 908 m; Mean: 2.3; Median: 1.9; St. Dev: 2.4</i>	-	-	-	-	-
CS1	<i>N: 6 m; Mean: 0; Median: 0; St. Dev: 0</i>	Alt Empordà	Significantly rural	0.9	0 - 60	Medium
		Gironès	Significantly rural	0.6	0 - 30	Very high
CS2	<i>N: 8 m; Mean: 1.1; Median: 1; St. Dev: 0.9</i>	Maresme Nord	Predominantly urban	0.9	0 - 45	Very high
		Maresme Centre	Predominantly urban	0.8	0 - 30	Very high
		Maresme Sud	Predominantly urban	1.1	0 - 15	Medium.
		Mollet	Predominantly urban	1.2	0 - 15	High
CS3	<i>N: 4 m; Mean: 0; Median: 0; St. Dev: 0</i>	Mòra d'Ebre	Significantly rural	2.8	0 - 45	Medium low
		Reus	Predominantly urban	0.6	0 - 45	Low
		Tarragona Nord	Predominantly urban	0.6	0 - 30	Medium low
CS4	<i>N: 13 m; Mean: 0.2; Median: 0; St. Dev: 0.8</i>	Anoia	Significantly rural	1.1	0 - 60	Low
		Bages	Significantly rural	0.6	0 - 60	Very high
		Balaguer	Predominantly rural	3.2	0 - >60	Medium
		Berga	Significantly rural	3.0	0 - 45	High
		Lleida	Significantly rural	0.6	0 - 45	High
		Osona	Predominantly urban	0.8	0 - 45	Very high
No CS	<i>N: 915 m; Mean: 2.6; Median: 2.0; St. Dev: 2.9</i>	Sort	Predominantly rural	11.8	0 - >60	Low
		-	-	-	-	-
		-	-	-	-	-
		-	-	-	-	-
		-	-	-	-	-
		-	-	-	-	-
		-	-	-	-	-
Catalonia	<i>N: 946 m; Mean: 2.6; Median: 2.0; St. Dev: 2.9</i>	-	-	1.2	-	-

CS, cold spot; HS, hot spot; m, municipalities; MHCC, mental health community centre; No HS, Municipalities not included in hot spots; No CS, municipalities not included in cold spots; SU, spatial units.

cold spots can be considered independent groups of spatial units, with different geographical location and rates of treated prevalence of depression.

#### Catchment areas and hot/cold spots

The hot/cold spots of treated prevalence of depression in Catalonia are located in 25 of the 74 mental health

catchment areas. The existence of hot/cold spots could not be attributed to the characteristics of the administrative division of mental health care in Catalonia as hot/cold spots could not be assigned to individual catchment areas. The spots include municipalities within different catchment areas. Therefore, the existence of hot/cold-spots cannot be attributable to a variation in clinical practice in specific MHCC.

On the other hand, municipalities in hot/cold spots inherit the main characteristics of the catchment areas in which they are included: urbanicity, availability, accessibility and adequacy. These domains are shown in Table 1.

Hot spots and cold spots are mainly located in the central-northern and eastern regions of Catalonia except for CS3 in the south. The size of the whole region does not suggest any relationship with the geographical characteristics of the territory because the affected catchment areas are urban and rural, industrial and agriculture-based, etc. (Table 1). HS1, HS2 and HS3 are located in mainly rural catchment areas although HS4 and HS5 have been identified in urban areas. On the other hand, CS1 and CS4 are located in predominantly rural areas while CS2 and CS3 are in urban ones. Treated prevalence of depression in Catalonia cannot easily be associated with urbanicity as has been claimed in other studies [34].

It has been stated that a long distance to a mental health specialist may reduce the numbers of visits of rural patients with depression [35]. There is no clear relationship between accessibility to MHCC and the location of hot/cold spots of treated prevalence of depression in Catalonia. According to Olivet et al. (2008), HS3, HS4, HS5, CS1 and CS2 are located in areas less than 30 minutes of travel time away. HS1, HS5, CS3 and CS4 are over 30 minutes away.

Areas with very low or very high accessibility to mental health services could be related to changes in the rates of treated prevalence for several reasons [36]. Perhaps MHCC could have an overload of cases due to the lack of other intermediate services. On the other hand, less availability of services may generate less demand and therefore a lower treated prevalence. However, the door-keeper effect of MHCC in a system organised by sectors and the lack of very-low adequacy areas rules out this possibility. On the other hand, previous studies have shown a relation between higher service utilisation and higher availability and provision. However, better accessibility may generate more cases and service use and an increase in cases attended in MHCC due to there being less diversity of services [37].

The availability of Adult MHCC per population of 100,000 in catchment areas with hot spots or cold spots is different, as seen in Table 1. The adequacy of the provision of all specialized mental health services is high or very high in two-thirds of the catchment areas with hot spots while 13.3% of the areas have low provision.

HS2, HS4 and HS5 have the highest adequacy while the lowest is in HS3. Half of the catchment areas with cold spots have high or very high provision while 31.3% are low or medium low. The cold spots with the best provision are CS1 and CS2, while CS3 is the lowest and CS4 shows a great degree of variation.

#### Limitations of the study

This study is not aimed at identifying the causes of hot/cold spots. It is not possible to infer individual level relationship from relationship observed at the aggregate level due to the ecological fallacy [38,39]. The analysis of the Catalonia Health Survey indicates that the local burden of depression is associated with a low educational level and living alone [40]. The ESEMeD project shows the relationship between the prevalence of depression and sex, unemployment, civil status and disabilities [34].

On the other hand, the modifiable spatial unit problem is an additional difficulty in spatial analysis [1]. This problem refers to the variation in interpretations of statistics and results due to the size of the geographical area where individual data have been aggregated. The impact of both problems has been reduced in our study using the municipality scale. This aggregate unit is the best spatial unit available that does not compromise individual identification and confidentiality.

The spatial units located at the borders of the region are an important constraint for spatial analysis because the values of their neighbouring areas located in other regions or countries are not known. If these values were available, additional hot/cold spots could appear in the territory. However, this problem is reduced in Catalonia due to its geographical characteristics (a long seacoast) and the organisation of care by defined sectors.

This study uses specialised health databases; this information should be completed with the analysis of depression treated in primary care. Unfortunately primary care databases do not cover the whole territory of Catalonia and only partial information is available on mental disorders treated in the primary care system [41]. Furthermore, the specialised care registries are not complete in 7 out of the 74 mental health catchment areas.

#### Conclusions

Hot spots and cold spots have previously been identified in a number of studies [42,43]. However, to our knowledge this is the first analysis that combines the optimization of basic statistics (mean and standard deviation) and the geographical location of hot and cold spots in one single procedure based on a hybrid model. A number of cold spots delimits different hot spots which would have been regarded as a single cluster otherwise. This analysis has also identified radial cluster patterns of cold spots which have not previously been described and which improve the identification of hot spots. It is especially interesting because it allows the

identification of hot spots that could be considered to be one and the same due to their proximity, though each of them could be generated by different factors.

The relationship between hot and cold spots may have appeared in previous studies although their tentative relationship and meaning were not described [43]. In order to better understand the spatial distribution of a disease in the territory, hot and cold spots should be described together.

The location of both hot/cold spots may require specific actions including flexible health programs, plans and priority settings [44]. The visual representation of the results on maps can be a relevant component of the Knowledge Discovery Data applied to Health System research. It facilitates the elicitation of implicit expert knowledge to better understand complex information using, for example, Expert-based Cooperative Analysis (EbCA) [45].

Future studies may include the combined analysis of different databases (e.g. in primary and in tertiary care), the probing of the relationship between hot/cold spots in other diagnosis and territories, and the relationships between spatial clusters of treated prevalence of depression and the characteristics of mental health catchment areas and socioeconomic indicators through regression methods and ordinal classification.

#### Abbreviations

CS: Cold spot; EbCA: Expert-based Cooperative Analysis; GIS: Geographic Information System; HS: Hot spot; ICD-10: International Statistical Classification of Diseases and Related Health Problems. 10th Revision; LISA: Local Indicators of Spatial Association; MHCC: Mental Health Community Centre; CMBD-SMA: Minimum Data Set for Outpatient Mental Health Centres; MOEA: Multi-Objective Evolutionary Algorithm.

#### Competing interests

The authors declare that they do not have competing interests with regard to this manuscript.

#### Authors' contributions

CRGA and LSC designed the tool for the spatial analysis, JASP performed the data mining, data analysis and maps, and CMP and EJS analyzed and interpreted the results. All authors contributed to the writing of the manuscript. All authors read and approved the final manuscript.

#### Authors' information

for the GEOSCAT Group  
Antoni Serrano, Ana Fernández, Teresa Marfull, Miriam Poole, Mencia Ruiz, María Luisa Rodero, Javier Álvarez, Josep María Haro, Esther Rovira, Josep Fusté, Cristina Romero and Bibiana Prat.

#### Acknowledgements

This study is part of the project entitled 'Development of a health map of services and facilities for the integral care of people with mental illness and the application of geographic information systems for decision support in planning services in Catalonia' [Project: CPA 73.10.15], which was funded by the Health Department of Catalonia. This study is also partly subsidised by the Carlos III Health Institute (Ministry of Health of Spain) [project P111/02008] and co-funded by FEDER funds. We also thank the reviewers for their helpful comments that improved the quality of our final manuscript.

#### Author details

<sup>1</sup>Universidad Loyola Andalucía, Business Administration Faculty, Sevilla, Córdoba, Spain. <sup>2</sup>Direcció General de Regulació, Planificació i Recursos

Sanitaris, Departament de Salut, Generalitat de Catalunya, Barcelona, Spain. <sup>3</sup>Faculty of Health Sciences, University of Sydney, Sydney, Australia.

Received: 8 June 2012 Accepted: 10 August 2012

Published: 24 August 2012

#### References

1. Elliott P, Wartenberg D: **Spatial epidemiology: current approaches and future challenges.** *Environ Health Perspect.* 2004, **112**:998–1006.
2. Ward M: **Geospatial Technologies and Homeland Security.** In *Spatial Epidemiology: Where Have We Come in 150 Years? Volume 94.* Edited by Sui DZ. Dordrecht: Springer; 2008:257–282.
3. García-Alonso CR, Salvador-Carulla L, Negrín-Hernández MA, Moreno-Küstner B: **Development of a new spatial analysis tool in mental health: identification of highly autocorrelated areas (hot-spots) of schizophrenia using a Multiobjective Evolutionary Algorithm model (MOEA/HS).** *Epidemiol Psychiatr Soc.* 2010, **19**:302–313.
4. Auchincloss AH, Gebreab SY, Mair C, Diez Roux AV: **A Review of Spatial Methods in Epidemiology, 2000–2010.** *Annu Rev Public Health.* 2012, **33**:107–122.
5. Bithell JF: **A classification of disease mapping methods.** *Stat Med.* 2000, **19**:2203–2215.
6. Curtis S, Copeland A, Fagg J, Congdon P, Almog M, Fitzpatrick J: **The ecological relationship between deprivation, social isolation and rates of hospital admission for acute psychiatric care: a comparison of London and New York City.** *Health Place.* 2006, **12**:19–37.
7. Kirkbride JB, Fearon P, Morgan C, Dazzan P, Morgan K, Murray RM, Jones PB: **Neighbourhood variation in the incidence of psychotic disorders in Southeast London.** *Soc Psychiatry Psychiatr Epidemiol.* 2007, **42**:438–445.
8. Fortney JC, Rushton G, Wood S, Zhang L, Xu S, Dong F, Rost K: **Community-Level Risk Factors for Depression Hospitalizations.** *Adm Policy Ment Health.* 2007, **34**:343–352.
9. Fortney JC, Xu S, Dong F: **Community-Level Correlates of Hospitalizations for Persons With Schizophrenia.** *Psychiatr Serv.* 2009, **60**:772–778.
10. Zhen H, McDermott S, Lawson A, Aelion M: **Are clusters of mental retardation correlated with clusters of developmental delay?** *Geospat Health.* 2009, **4**:17–26.
11. Chaix B, Leyland AH, Sabel CE, Chauvin P, Råstam L, Kristersson H, Merlo J: **Spatial clustering of mental disorders and associated characteristics of the neighbourhood context in Malmö, Sweden, in 2001.** *J Epidemiol Community Health.* 2006, **60**:427–435.
12. Gruebner O, Khan MMH, Lautenbach S, Muller D, Kramer A, Lakes T, Hostert P: **A spatial epidemiological analysis of self-rated mental health in the slums of Dhaka.** *Int J Health Geogr.* 2011, **10**:36.
13. Moscone F, Knapp M, Tosetti E: **Mental health expenditure in England: A spatial panel approach.** *J Health Econ.* 2007, **26**:842–864.
14. Moreno B, García-Alonso CR, Negrín Hernández M, Torres-González F, Salvador-Carulla L: **Spatial analysis to identify hotspots of prevalence of schizophrenia.** *Soc Psychiatry Psychiatr Epidemiol.* 2008, **43**:782–791.
15. Torabi M, Rosychuk RJ: **An examination of five spatial disease clustering methodologies for the identification of childhood cancer clusters in Alberta, Canada.** *Spat Spatiotemporal Epidemiol.* 2011, **2**:321–330.
16. Jacques GM, Kaufmann A, Goovaerts P: **Boundaries, links and clusters: a new paradigm in spatial analysis?** *Environ Ecol Stat.* 2008, **15**:403–419.
17. Caçado AL, Duarte AR, Duczmal LH, Ferreira SJ, Fonseca CM, Gontijo EC: **Penalized likelihood and multi-objective spatial scans for the detection and inference of irregular clusters.** *Int J Health Geogr.* 2010, **9**:55.
18. Eurostat: *Eurostat regional yearbook 2010.* Luxembourg: Publications Office of the European Union; 2010.
19. Salvador-Carulla L, Costa-Font J, Cabases J, McDaid D, Alonso J: **Evaluating mental health care and policy in Spain.** *J Ment Health Policy Econ.* 2010, **13**:73–86.
20. Health and Social Security Department: *Notification Handbook of the Register of the Minimum Basic Data Set: Outpatient Mental Health Centers.* Barcelona: Catalan Health Service; 2003.
21. World Health Organization: *International Statistical Classification of Diseases and Related Health Problems, 10th Revision.* 2010th edition. Geneva: World Health Organization; 2011. Volume 2. Instruction manual.
22. Rezaeian M, Dunn G, St Leger S, Appleby L: **Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary.** *J Epidemiol Community Health.* 2007, **61**:98–102.

23. OECD: *Creating rural indicators for shaping territorial policy*. Paris: Organisation for Economic Co-operation and Development; 1994.
24. Olivet M, Aloy J, Prat E, Pons X: **Health services provision and geographic accessibility**. *Med Clin (Barc)*. 2008, **131**(Suppl 4):16–22.
25. GEOSCAT Group: *Integral Map of Mental Health Resources of Catalonia*. Barcelona: Department of Health of Catalonia; In press.
26. Anselin L: **Local indicators of spatial association-LISA**. *Geogr Anal* 1995, **27**:93–115.
27. Ord JK, Getis A: **Local spatial autocorrelation statistics: distributional issues and an application**. *Geogr Anal* 1995, **27**:186–306.
28. García-Alonso CR, Pérez-Naranjo LM, Fernández-Caballero JC: **Multiobjective evolutionary algorithms to identify highly autocorrelated areas: the case of spatial distribution in financially compromised farms**. *Ann Oper Res* 2011, **1**:16.
29. Coello-Coello C, Lamont G, Van Veldhuizen D: *Evolutionary algorithms for solving multi-objective problems*. New York: Springer; 2007.
30. Beirlant J, Goegebeur Y, Teugels J, Segers J: *Statistics of extremes: theory and applications*. Chichester, West Sussex: Wiley; 2004.
31. Lewin S, Oxman AD, Lavis JN, Fretheim A, Garcia Marti S, Munabi-Babigumira S: **SUPPORT tools for evidence-informed policymaking in health 11: Finding and using evidence about local conditions**. *Health Res Policy Syst* 2009, **7**(Suppl 1):S11.
32. Gustavsson A, Svensson M, Jacobi F, Allgulander C, Alonso J, Beghi E, Dodel R, Ekman M, Faravelli C, Fratiglioni L, Gannon B, Jones DH, Jennum P, Jordanova A, Jönsson L, Karampampa K, Knapp M, Kobelt G, Kurth T, Lieb R, Linde M, Ljungcrantz C, Maercker A, Melin B, Moscarelli M, Musayev A, Norwood F, Preisig M, Pugliatti M, Rehm J, Salvador-Carulla L, Schlehofer B, Simon R, Steinhausen H-C, Stovner LJ, Vallat J-M, den Bergh PV, van Os J, Vos P, Xu W, Wittchen HU, Jönsson B, Olesen J: **Cost of disorders of the brain in Europe 2010**. *Eur Neuropsychopharmacol* 2011, **21**:718–779.
33. Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jönsson B, Olesen J, Allgulander C, Alonso J, Faravelli C, Fratiglioni L, Jennum P, Lieb R, Maercker A, van Os J, Preisig M, Salvador-Carulla L, Simon R, Steinhausen H-C: **The size and burden of mental disorders and other disorders of the brain in Europe 2010**. *Eur Neuropsychopharmacol* 2011, **21**:655–679.
34. Gabilondo A, Rojas-Farreras S, Vilagut G, Haro JM, Fernández A, Pinto-Meza A, Alonso J: **Epidemiology of major depressive episode in a southern European country: results from the ESEMeD-Spain project**. *J Affect Disord* 2010, **120**:76–85.
35. Fortney JC, Rost K, Zhang M, Warren J: **The impact of geographic accessibility on the intensity and quality of depression treatment**. *Med Care*. 1999, **37**:884–893.
36. Marín I, Briones E: **Variability and clinic management. Concerning the use of the atlas for clinical Ulysses to overcome cyclop's vision**. *Atlas of Variations in Medical Practice* 2007, **2**:139–141.
37. Gittelsohn A, Powe NR: **Small area variations in health care delivery in Maryland**. *Health Serv Res*. 1995, **30**:295–317.
38. Macintyre S, Ellaway A, Cummins S: **Place effects on health: how can we conceptualise, operationalise and measure them?** *Soc Sci Med* 2002, **55**:125–139.
39. Ocaña-Riola R: **Common errors in disease mapping**. *Geospat Health*. 2010, **4**:139–154.
40. Sabes-Figuera R, Knapp M, Bendeck M, Mompart-Penina A, Salvador-Carulla L: **The local burden of emotional disorders. An analysis based on a large health survey in Catalonia (Spain)**. *Gac Sanit* 2012, **26**:24–29.
41. Aragonès E, Salvador-Carulla L, López-Muntaner J, Ferrer M, Piñol JL: **Registered prevalence of borderline personality disorder in primary care databases**. *Gac Sanit*. In press.
42. Cheng C-L, Chen Y-C, Liu T-M, Kao-Yang Y-H: **Using Spatial Analysis to Demonstrate the Heterogeneity of the Cardiovascular Drug-Prescribing Pattern in Taiwan**. *BMC Publ Health* 2011, **11**:380.
43. Sridharan S, Koschinsky J, Walker JJ: **Does context matter for the relationship between deprivation and all-cause mortality? The West vs. the rest of Scotland**. *Int J Health Geogr*. 2011, **10**:33.
44. Koschinsky J: **The case for spatial analysis in evaluation to reduce health inequities**. *Eval Program Plann*. In press.
45. Gibert K, García-Alonso CR, Salvador-Carulla L: **Integrating clinicians, knowledge and data: expert-based cooperative analysis in healthcare decision support**. *Health Res Policy Syst*. 2010, **8**:28.

doi:10.1186/1476-072X-11-36

**Cite this article as:** Salinas-Pérez et al.: Identification and location of hot and cold spots of treated prevalence of depression in Catalonia (Spain). *International Journal of Health Geographics* 2012 **11**:36.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

