

DIRECCIÓN DE GEOESTADÍSTICA

**IDENTIFICACIÓN DE AGRUPACIONES ESPACIALES
POR TÉCNICAS DE ESTADÍSTICA ESPACIAL, A
PARTIR DE LA DIFERENCIA ENTRE UNIDADES
CENSALES Y PREDIOS CATASTRALES**

MARZO de 2021

INFORME DE REVISIÓN DE DATOS ABIERTOS DE FACEBOOK



El futuro
es de todos

Gobierno
de Colombia

CONTENIDO

1. IMPLEMENTACIÓN	4
1.2. Objetivo	4
1.3. Insumos	4
1.4. Técnicas espaciales	4
2. PROCESAMIENTO	5
2.2. Procesamiento a nivel de cabeceras municipales	7
2.3. Procesamiento a nivel rural disperso	8
3. REFERENCIAS BIBLIOGRÁFICAS	11

LISTA DE FIGURAS

Figura 1: Esquema de cluster Hot/Cold spots según el nivel de significancia.....	6
Figura 2: Diagrama de flujo de procesos. Nivel urbano.....	7
Figura 3: Resultados de agrupaciones espaciales de altas/bajas diferencias UC- predios por manzana. Nivel urbano.	8
Figura 4: Diagrama de flujo de procesos. Nivel rural disperso.	9
Figura 5: Resultados de agrupaciones espaciales de altas/bajas diferencias UC- predios por manzana. Nivel rural disperso.....	10

LISTA DE TABLAS

Tabla 1: Categorías de agrupación, Indice local Gi de Getis & Ord.	6
---	---

1. IMPLEMENTACIÓN

1.1. Objetivo

Identificación de agrupaciones espaciales significativas estadísticamente sobre las diferencias numéricas entre las unidades censales y predios catastrales, a nivel submunicipal, utilizando como área de agregación la manzana censal en cabeceras municipales y la sección en áreas rurales, pertenecientes al Marco Geoestadístico Nacional.

1.2. Insumos

- ❖ Marco Geoestadístico Nacional desagregado por manzana censal y sección rural, disponible en [MGN](#),
- ❖ Unidades censales y predios catastrales, agregados por conteos en manzanas y sección rurales.

1.3. Técnicas espaciales

Vecindad espacial

La autocorrelación espacial permite comprender la variabilidad de un fenómeno que ocurre en un espacio geográfico, de tal manera que, si el fenómeno tiende a formar agrupaciones espaciales de valores similares o conglomerados, entonces se evidencia la existencia de autocorrelación positiva; por el contrario, si las medidas de la unidad espacial estudiada son disimiles de las unidades contiguas, entonces es indicador de autocorrelación negativa, es decir, se evidencia dispersión. Por último, cuando no se logra identificar un comportamiento estructurado, entonces el fenómeno se distribuye de manera aleatoria.

El concepto de vecindad es el que describe las relaciones de asociación entre las unidades espaciales (áreas o *lattices*) en el espacio analizado. Los criterios de vecindad utilizan la distancia o la contigüidad física. Estos criterios son los más extendidos ya que suelen estar implementados por defecto en los programas de análisis espacial.

El criterio de vecindad por distancia puede definirse a partir de un número definido de unidades vecinas más cercanas, y se establece a partir de un límite máximo de vecinos determinados por la distancia de separación entre las unidades, que usualmente es medida entre los centroides de los polígonos que representan dichas unidades. De otro lado, el criterio de vecindad física se basa en la

contigüidad espacial de orden n, donde el orden 1 corresponde a los vecinos dados por los límites geográficos directos.

Análisis Hot-Spots - Gi* de Getis Ord

Estadístico local que mide el grado de agrupamiento para valores altos o bajos de la variable estudiada; es decir, dado un conjunto de valores de la variable distribuidas geográficamente, identifica puntos calientes o valores altos y puntos fríos o valores bajos estadísticamente significativos.

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j}x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\left[n \sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j} \right)^2 \right] / (n-1)}}$$

Donde:

$$W_i^* = \frac{1}{N-1} \sum_j (x_j - \bar{x})^2$$

Si $G_i > 0$ y significativo, se identifica un cluster de valores similares elevados alrededor de la unidad i

Si $G_i < 0$ y significativo, se identifica un cluster de valores similares bajos alrededor de la unidad i

2. PROCESAMIENTO

La identificación de las agrupaciones espaciales de la variable de estudio puede ser realizados de forma masiva para todo el territorio nacional, y de manera independiente para el conjunto de datos de cada municipio; por tanto, se construyó un modelo con la herramienta Model Builder del Software ArcGIS 10.5, por medio del cual se realiza el cálculo del índice local Gi de Getis - Ord y se obtienen las capas con las categorías de significancia dadas por este índice para cada municipio procesado (ver Tabla 1). Para cada nivel de desagregación municipal se desarrolló un modelo conforme la naturaleza de los elementos geográficos.

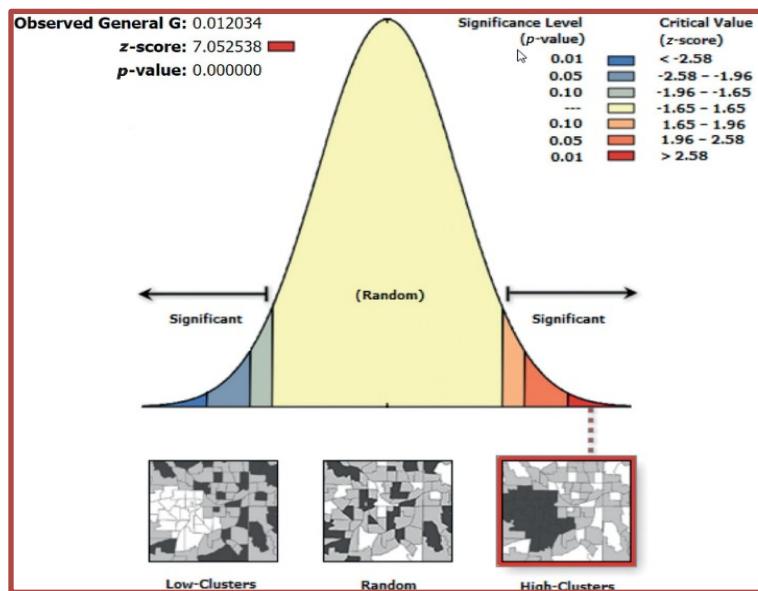


Figura 1: Esquema de cluster Hot/Cold spots según el nivel de significancia.

Tabla 1: Categorías de agrupación, Indice local Gi de Getis & Ord.

Categoría	Tipo de agrupación	Nivel de confianza
-3	Agrupaciones de baja diferencia (cold-spots) entre unidades censales-catastrales a un nivel de confianza dado	99%
-2		95%
-1		90%
0	No significativo	-
1	Agrupaciones de alta diferencia (hot-spots) entre unidades censales-catastrales a un nivel de confianza dado	90%
2		95%
3		99%

2.1. Procesamiento a nivel de cabeceras municipales

El cálculo del índice local es realizado al nivel de desagregación de manzana censal en cabecera municipal, donde en primera instancia se realiza una agrupación por el código divipola del municipio asignado a cada manzana ingresada en el conjunto de datos de entrada, con el propósito de realizar el procesamiento de forma independiente para cada municipio; luego, para cada conjunto de procesamiento (manzanas censales de cada municipio) se genera la matriz de contigüidad espacial bajo en criterio de proximidad dado por los 8 vecinos más cercanos¹. Por último, se realiza el cálculo del índice local Gi de Getis & Ord teniendo en cuenta como variable espacial la diferencia entre unidades censales y catastrales asociado a cada manzana censal.

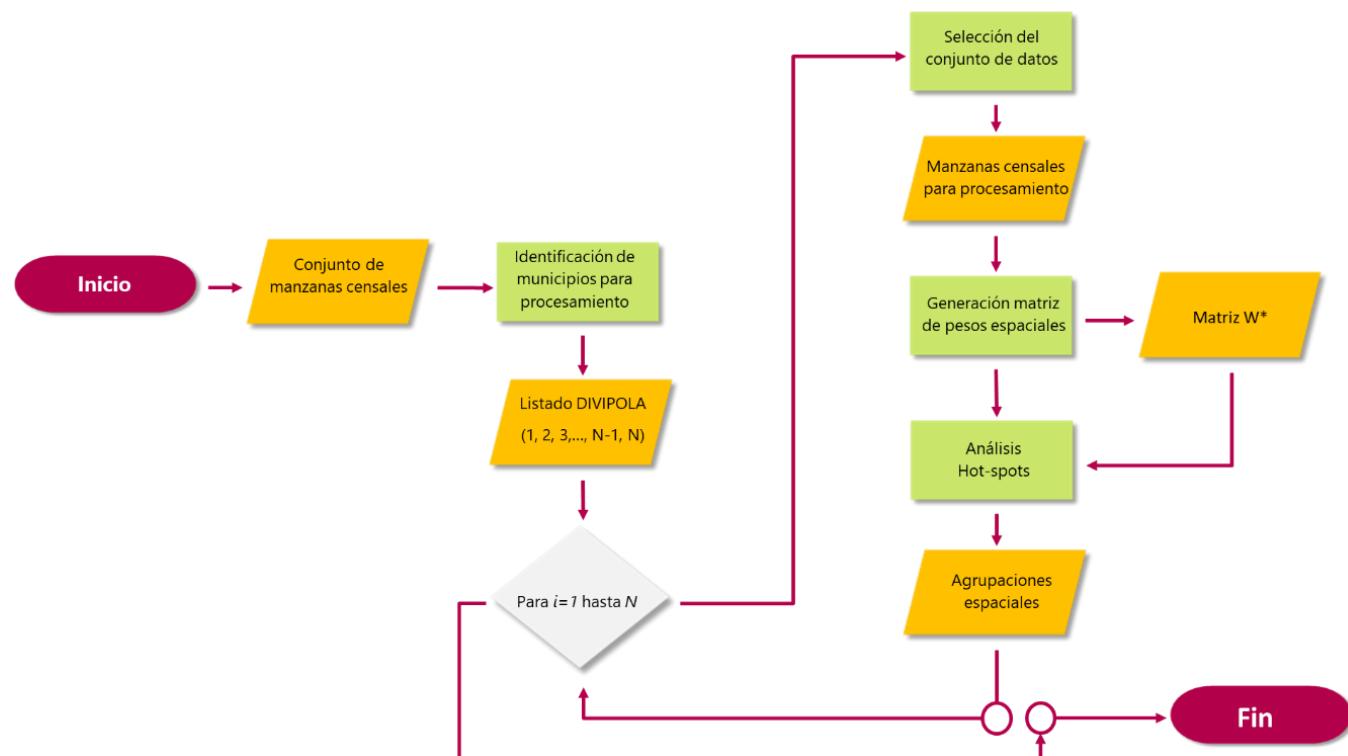


Figura 2: Diagrama de flujo de procesos. Nivel urbano

¹ Criterio de contigüidad escogido dada la naturaleza de los elementos espaciales objeto de análisis. Para el caso particular de las manzanas censales, éstas no presentan límites geográficos directos dada la separación por la cobertura vial.

Como resultado del cálculo de este índice, se obtiene para cada cabecera municipal, las agrupaciones de manzana censales cuya diferencia (alta o baja) entre unidades censales y catastrales son estadísticamente significativa, a tres niveles distintos de significancia.

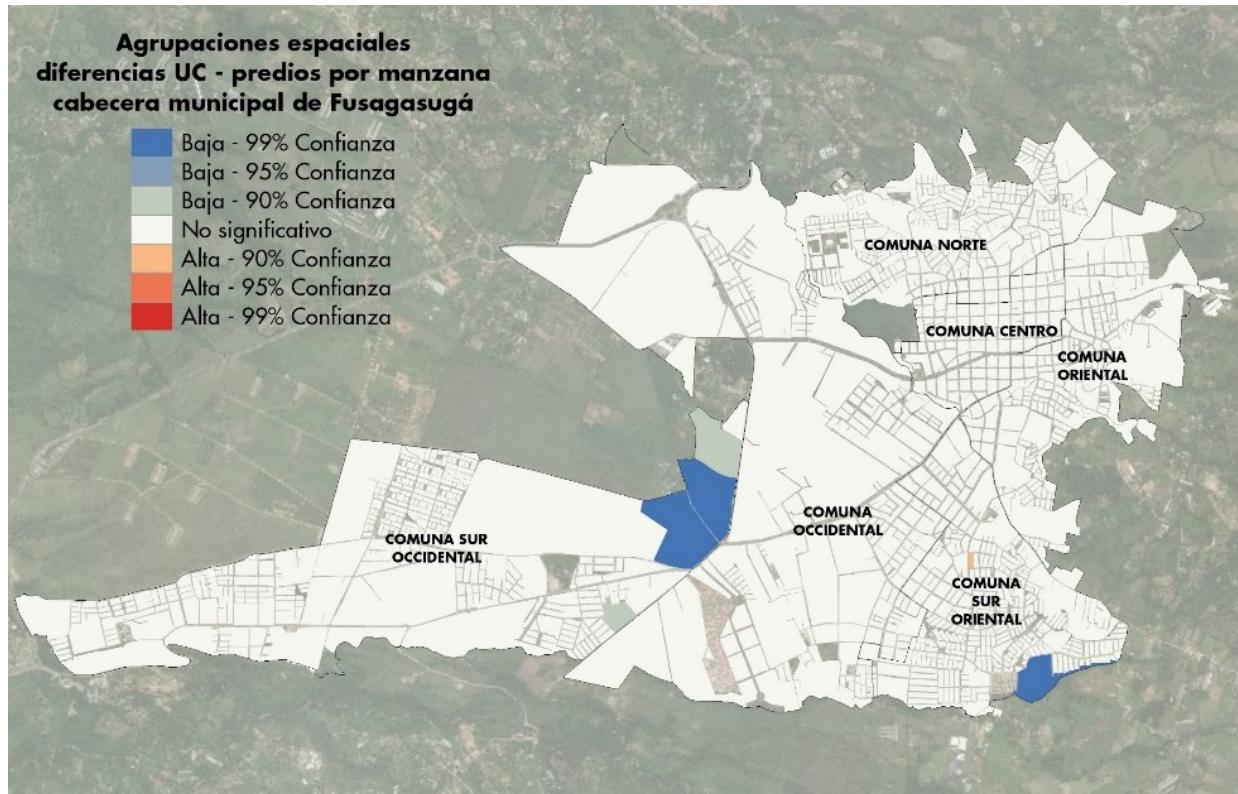


Figura 3: Resultados de agrupaciones espaciales de altas/bajas diferencias UC- predios por manzana. Nivel urbano.

2.2. Procesamiento a nivel rural disperso

El procesamiento de datos al nivel de desagregación de sección rural aprovecha los límites geográficos directos de los elementos espaciales objeto de este análisis, con el propósito de ampliar el número de observaciones para garantizar la apropiada identificación de agrupaciones espaciales en los distintos niveles de significancia. Primero, se realiza una agrupación de secciones rurales que pertenecen a cada municipio ingresado a partir del código divipola asignado. La ampliación del número de secciones rurales a procesar para un municipio en particular es realizada a partir de la selección de las áreas rurales de municipios vecinos que presenten límites espaciales directos. Se genera la matriz de contigüidad espacial teniendo en cuenta el criterio reina, que considera a cualquier par de secciones

rurales como vecinas si comparten mínimo un nodo o vértice. Se realiza el cálculo del índice local Gi de Getis & Ord teniendo en cuenta como variable espacial la diferencia entre unidades censales y catastrales asociado a cada sección. Por último, se crea una capa geográfica con los resultados de este índice únicamente para los elementos espaciales del municipio objeto de análisis.

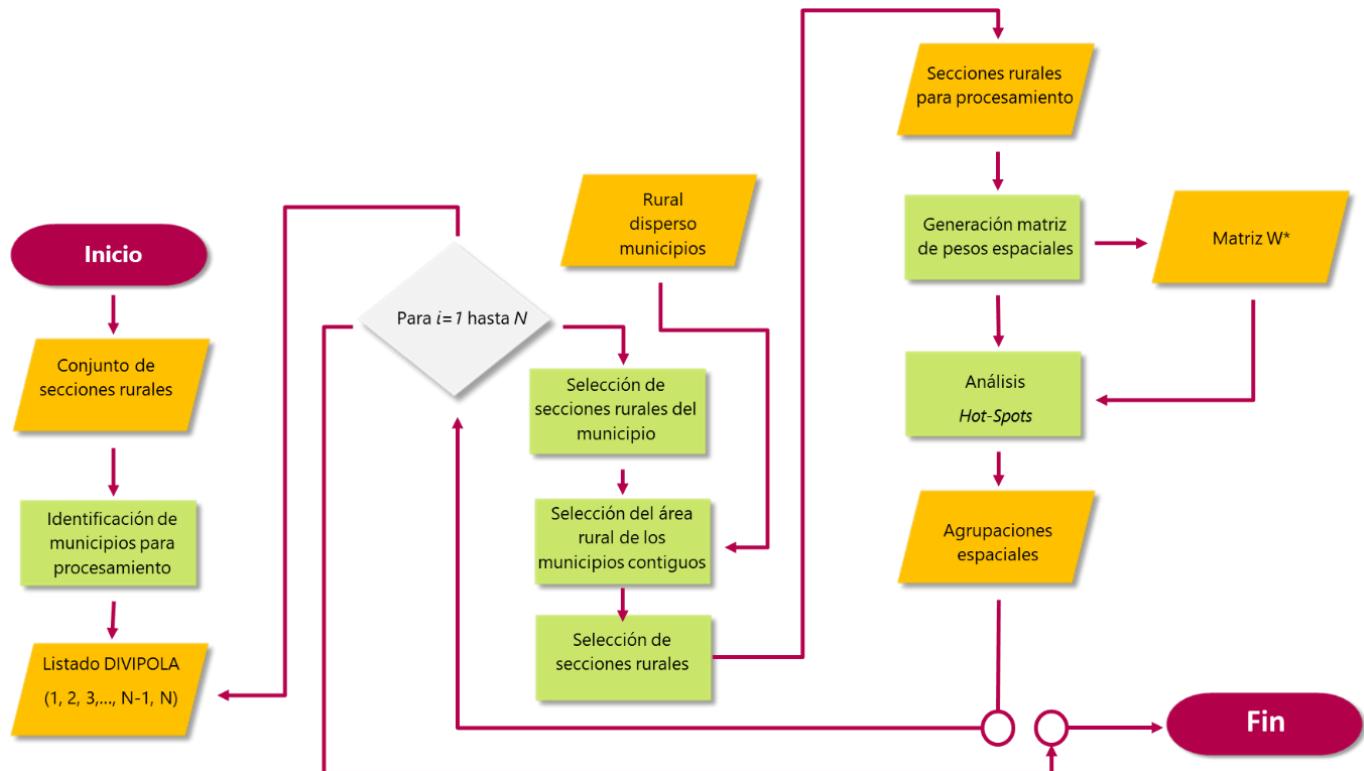


Figura 4: Diagrama de flujo de procesos. Nivel rural disperso.

Como resultado del cálculo de este índice, se obtiene para cada municipio, las agrupaciones de secciones rurales cuya diferencia (alta o baja) entre unidades censales y catastrales son estadísticamente significativa, a tres niveles distintos de significancia.

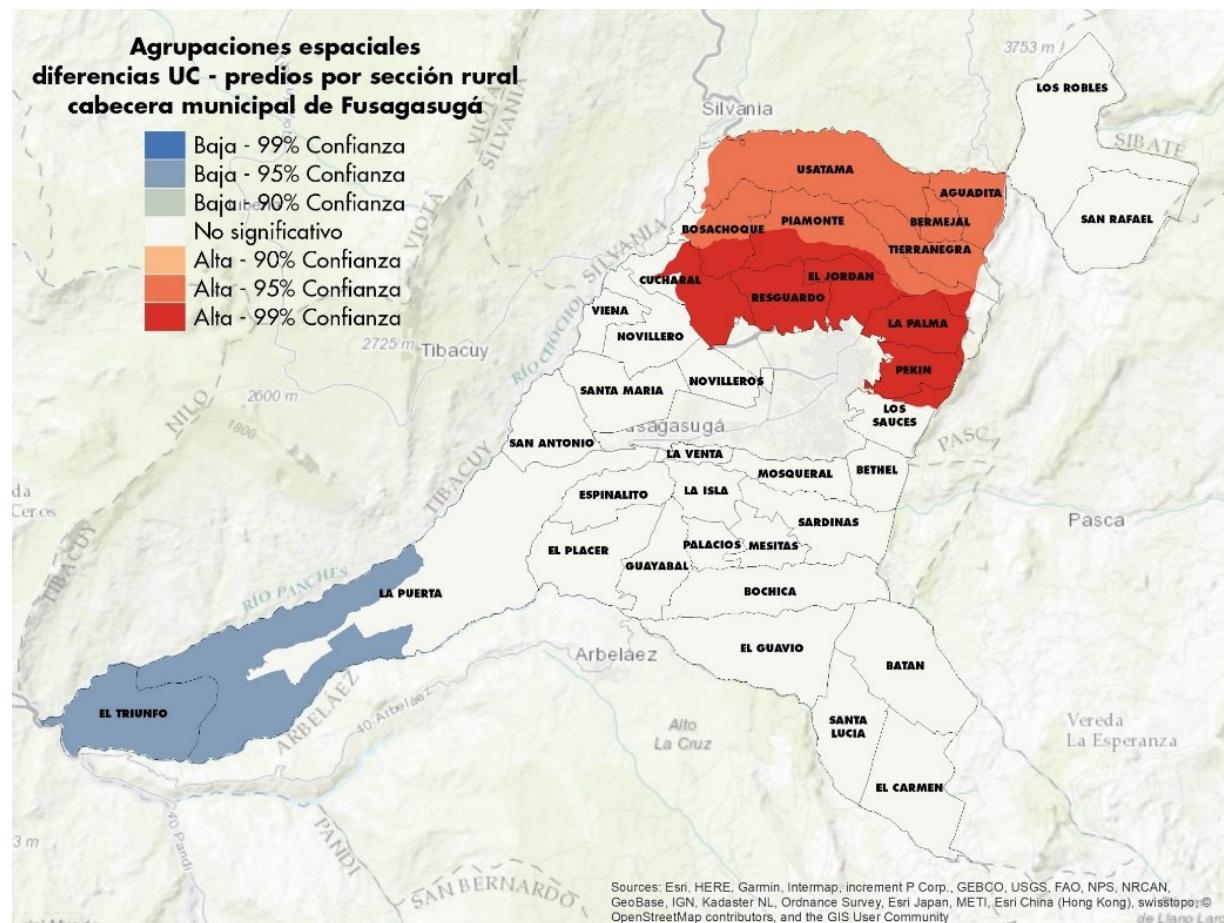


Figura 5: Resultados de agrupaciones espaciales de altas/bajas diferencias UC- predios por manzana. Nivel rural disperso.

3. REFERENCIAS BIBLIOGRÁFICAS

- Celemín, J. P. (2009). Autocorrelación espacial e indicadores locales de asociación espacial. Importancia, estructura y aplicación. *Revista Universitaria de Geografía, Universidad Nacional del Sur*, 18, 11-31. Obtenido de <https://www.redalyc.org/pdf/3832/383239099001.pdf>
- Esri. (2020). *ArcGIS Desktop*. Obtenido de Análisis de punto caliente (Gi* de Getis-Ord): <https://desktop.arcgis.com/es/arcmap/10.3/tools/spatial-statistics-toolbox/hot-spot-analysis.htm>
- Herrera Gómez, M. (2013). Análisis de Estructuras Espaciales Persistentes. Desempleo Departamental en Argentina. Obtenido de <https://mpra.ub.uni-muenchen.de/49407/>
- Moreno Serrano, R., & Vayá Valcarce, E. (2002). Econometría espacial, nuevas técnicas para el análisis regional. Una aplicación a las regiones europeas. 83 -106. Obtenido de <https://www.redalyc.org/articulo.oa?id=28900104>
- Ramirez Yara, Y. (2012). Análisis de autocorrelación espacial en el desarrollo sostenible respecto al componente social, ambiental, institucional y económico de los municipios de Corgochivor para el año 2009. Bogotá.
- Siabato, W., & Guzmán-Manrique, J. (2018). La autocorrelación espacial y el desarrollo de la geografía cuantitativa. *Cuadernos de Geografía: Revista Colombiana de Geografía*, 1-22. doi: 10.15446/rcdg.v28n1.76919

