

# Mapa y jerarquía espacial de la pobreza en México. Un nuevo procedimiento para identificar el patrón espacial de los problemas sociales\*

Mapping Poverty and its Spatial Hierarchy in Mexico.  
A New Method for Identifying the Spatial Pattern  
of Social Problems

*Jesús A. Treviño C.\*\**

## ABSTRACT

This research presents and applies a new procedure to elaborate a spatial hierarchy of poverty. The proposed method simultaneously combines variables of magnitude and intensity based on recent (bootstrapping & spatial autocorrelation) and traditional statistical techniques, and overlay routines of Geographical Information Systems.

While magnitude and intensity refer to absolute and relative data, respectively, each variable may be concentrated or agglomerated in space. In this study, concentration is the presence of high global values, regardless of their location, and agglomeration is the concentration of spatially contiguous high local values. Both agglomeration and concentration are merged through a geographical overlay procedure to create conglomerates of magnitude or intensity of poverty. Cases inside these conglomerates are classified by gaussian (natural breaks) or paretian (heads and tails) procedures to set up a spatial hierarchy.

\* Artículo recibido el 14 de mayo de 2015 y aceptado el 10 de febrero de 2016.

\*\* Profesor de posgrado, Facultad de Arquitectura, Universidad Autónoma de Nuevo León (correo electrónico: jtrevino41@hotmail.com). Los errores remanentes son responsabilidad del autor.

For the first time in the study of the spatial pattern of poverty, the resulting spatial hierarchy is based on the simultaneous combination of the concentration and agglomeration processes measured in relative and absolute terms. The benefits of the procedure for an area-based public policy are illustrated by assessing the spatial targeting of poverty in the 2,456 Mexican municipios in 2010. The suggested methodology in this research may be easily extended to identify other spatial patterns, such as crime, industry, diseases, pollution or environmental justice in different areas or countries.

*Key words:* spatial analysis, agglomeration, concentration, conglomerates, poverty.  
*JEL classification:* R58, C1, I3.

### RESUMEN

Esta investigación presenta y aplica un nuevo procedimiento para jerarquizar espacialmente la pobreza. El método propuesto combina simultáneamente variables de magnitud e intensidad utilizando técnicas estadísticas recientes (remuestreo y autocorrelación espacial) y tradicionales, así como rutinas de superposición espacial de los Sistemas de Información Geográfica.

Mientras magnitud e intensidad se refieren a datos absolutos y relativos, respectivamente, cada variable puede estar concentrada o aglomerada en el espacio. En este estudio, concentración es la presencia de valores globales altos, independientemente de su localización, y aglomeración es la concentración de valores locales altos espacialmente contiguos. Ambas, aglomeración y concentración, son fusionadas mediante un procedimiento de superposición geográfica para crear conglomerados de magnitud o intensidad de pobreza. Los casos al interior de estos conglomerados son clasificados por procedimientos gaussianos (discontinuidades naturales) o paretianos (puntas y colas) para crear una jerarquía espacial.

Por primera vez en el estudio de la pobreza, la jerarquía espacial resultante es producto de la combinación simultánea de los procesos de concentración y aglomeración medidos en términos absolutos y relativos. Los beneficios del procedimiento en una política pública espacialmente orientada son ilustrados abordando la focalización espacial de la pobreza en los 2456 municipios mexicanos en el 2010. La metodología propuesta en esta investigación puede adaptarse fácilmente para identificar los patrones espaciales de otros problemas sociales, tales como crimen, industria, enfermedades, contaminación o justicia ambiental en diferentes áreas o países.

*Palabras clave:* análisis espacial, aglomeración, concentración, conglomerados, pobreza. *Clasificación JEL:* R58, C1, I3.

## INTRODUCCIÓN

El interés por estudiar el patrón espacial de la pobreza es creciente, tanto en la academia (Partridge y Rickman, 2006) como en la política pública (Bedi, Coudouel y Simler, 2007; Coneval, 2014). El término *patrón espacial* en esta investigación se refiere a la localización o distribución geográfica de una variable en un año específico (situación) y al cambio locacional de la misma variable en un periodo dado (proceso).<sup>1</sup> La localización de la variable, a su vez, tiene dos características: magnitud e intensidad. Magnitud, generalmente expresada en términos absolutos, se refiere a la presencia nacional o interregional de la variable (pobreza, en el estudio de caso). Intensidad, por otro lado, mide la importancia local (intra-local) de la variable. En este trabajo, el análisis del patrón espacial (o simplemente, análisis espacial) se enfoca en la identificación de la jerarquía espacial de las áreas considerando la localización e intersección de la intensidad y la magnitud de la pobreza. El cambio locacional de esta variable no puede abordarse con la información disponible al nivel municipal en el estudio de caso mexicano.

La identificación del patrón espacial de la pobreza demanda una distinción clara entre magnitud e intensidad. Mientras magnitud se refiere a la cantidad, número, tamaño o volumen de la pobreza espacialmente extendida, intensidad tiene que ver con la gravedad o seriedad del problema espacialmente intensiva (Goodchild y Lam, 1980). El patrón espacial de la magnitud o la intensidad puede ser concentrado o aglomerado. El patrón espacialmente concentrado se refiere a la jerarquía de áreas considerando el nivel o grado de incidencia o magnitud de la variable (*i. e.*, muy alto, alto, medio, bajo), sin importar su localización geográfica. Por otro lado, el patrón geográfico aglomerado se refiere a la clasificación de áreas de magnitud o intensidad similar que son contiguos en el espacio, sin considerar la distribución, orden o jerarquía general de los valores. Los estudios sobre concentración y aglomeración son (o deberían ser) complementarios, no sustitutos. Esta conclusión, aunque lógica a la luz de estas definiciones, no está explícitamente presente en la literatura revisada, con la excepción del trabajo de Visvalingam (1983).

<sup>1</sup> Esta definición operativa es compatible con definiciones más amplias que consideran el patrón espacial como una expresión o manifestación material de procesos socioeconómicos esenciales subyacentes —*i. e.*, simulación espacial en O'Sullivan y Perry (2013), justicia espacial en Soja (2010), o desarrollo desigual en Gottdiener y Hutchison (2011)—.

El estudio de la dimensión espacial de la pobreza, en general, y la identificación de su jerarquía espacial, en particular, es importante por las siguientes razones:

- i)* Es la forma más eficiente de enfrentar la dimensión territorial de los retos y problemas sociales, tales como la provisión de educación secundaria y superior o servicios de salud especializados.
- ii)* Es una tarea inevitable en un país con un territorio muy amplio. La provisión de infraestructura social al estado de Chihuahua (247 460 km<sup>2</sup>), por ejemplo, es un esfuerzo espacial similar al de proporcionar ese equipamiento al Reino Unido (242 900 km<sup>2</sup>). Lo mismo puede decirse para el estado de Sonora (179 355 km<sup>2</sup>) en relación a la República del Uruguay (176 215 km<sup>2</sup>).
- iii)* Incrementa la visibilidad de la pobreza al identificar las áreas donde viven los pobres en condiciones socioeconómicas y biofísicas adversas.
- iv)* Analiza e incorpora el impacto socioeconómico de las estrategias de desarrollo regional o nacional.
- v)* Teniendo lo anterior en mente, el diseño de programas, la selección de áreas y la identificación de beneficiarios de la política social son prerrequisitos para la formulación de políticas.

Esta investigación, considerando la relevancia del estudio del patrón espacial de la pobreza y con base en las definiciones esbozadas arriba, proporciona una jerarquía espacial de las áreas de pobreza en México en las secciones siguientes. La primera sección enfoca el problema de la prioridad de las áreas en términos de valores globales altos y valores locales altos de la intensidad y magnitud. Esta primera sección, además, contiene los pasos metodológicos sustantivos para elaborar una jerarquía espacial de la pobreza. La segunda sección compara el procedimiento propuesto con otras técnicas empleadas para la identificación de conglomerados. La evaluación destaca los beneficios y omisiones de una bibliografía ilustrativa para acentuar las ventajas del método sugerido en esta investigación. La tercera sección aborda el estudio de caso en tres partes. La primera parte presenta las preguntas de investigación e hipótesis de trabajo orientadas al estudio de caso; la segunda aplica la metodología propuesta al caso mexicano y la tercera coteja los resultados con las hipótesis planteadas. Finalmente, la cuarta sección subraya los hallazgos principales, las limitaciones del estudio y las líneas futuras de investigación.

En síntesis, el escrito presenta un procedimiento nuevo para identificar y analizar patrones espaciales y lo ilustra con la pobreza en el estudio de caso mexicano. La investigación identifica una jerarquía espacial de la pobreza en México combinando simultáneamente concentración y aglomeración de valores absolutos y relativos. Los datos utilizados provienen de la información disponible más reciente sobre pobreza al nivel municipal (2010). El estudio señala que la identificación de áreas prioritarias puede y debería mejorarse o validarse con criterios de estadística espacial y no espacial. Estos criterios pueden integrarse mediante un análisis de superposición espacial, tal como se muestra en este trabajo.

## I. PROCEDIMIENTO PROPUESTO

Esta sección muestra que la pobreza se conglera en áreas que son más extensas que las aglomeraciones identificadas con estadística espacial. También presenta un portafolio de técnicas para identificar estos conglomerados y jerarquizar las áreas que los constituyen. El punto de partida en estas tareas es que hay valores globales altos que no coinciden con los valores locales altos, y viceversa, tal como se describe enseguida.

### 1. *Valores globales y locales altos*

Este trabajo considera dos tipos de valores altos: *valores globales altos* (VGA) y *valores locales altos* (VLA). Por un lado, los VGA se obtienen con estadística no espacial; son valores por encima de la media en distribuciones normales o superiores a la mediana en distribuciones asimétricas a la derecha (asimetría positiva). Son valores altos en términos del conjunto global de los datos. Por otro lado, los VLA son calificados sólo considerando los valores localizados dentro de cierta distancia, rango espacial o vecindad.<sup>2</sup> Dichos valores son identificados en estadística espacial mediante permutaciones.

Tal como se desglosa a detalle enseguida, los VGA se calculan tomando en cuenta el conjunto completo de datos, sin considerar la localización de los casos o unidades de observación. Por otro lado, los VLA se identifican teniendo como referencia algún criterio predefinido de contigüidad o vecindad. Como resultado, algunos (no todos) VGA pueden traslaparse con VLA.

<sup>2</sup> Estas definiciones se basan en Zhang *et al.* (2009). Se utiliza “valores altos” en vez de “casos atípicos” porque esta investigación se enfoca en los valores significativamente altos y no en valores extremos.

Ambos tipos de valores altos (VGA y VLA) representan procesos espaciales distintos que deberían integrarse analíticamente.

Mientras los VGA resultan del proceso de concentración general (concentración), los VLA son consecuencia de ese mismo proceso de concentración en áreas contiguas (aglomeración). Así presentado, parecería que la aglomeración es un subconjunto del proceso de concentración general. No es así. La georreferenciación (territorialización o espacialización) de la variable de interés muestra que el carácter local de la aglomeración genera una diferenciación espacial del proceso de concentración general que amerita ser atendida. Por ejemplo, la concentración general indica que la pobreza está altamente concentrada en unas cuantas áreas; el proceso de aglomeración, por otro lado, señala que algunas de esas áreas son contiguas en el espacio. Conviene advertir, sin embargo, que algunas áreas contiguas en el espacio que son altas desde el punto de vista local pueden ser bajas en el conjunto global de datos. Esta característica impide que el proceso de aglomeración sea simplemente el proceso de concentración en áreas contiguas.

Con base en pruebas empíricas preliminares, esta investigación sugiere y constata que hay valores globales altos no detectados localmente y hay valores locales altos inadvertidos globalmente (en la perspectiva nacional, en el estudio de caso). Esta situación demanda un procedimiento que simultáneamente considere valores globales y locales altos. Es una estrategia metodológica poco afortunada utilizar únicamente la estadística espacial o procedimientos no espaciales para identificar conglomerados de pobreza. Las líneas siguientes abordan la definición e identificación de VGA usando estadística no espacial (*i. e.*, remuestreo, discontinuidades naturales o puntas y colas consecutivas). Enseguida se presenta la identificación de valores locales altos utilizando la estadística espacial (*i. e.*, autocorrelación espacial). Finalmente, la sección sugiere la superposición de valores globales altos y valores locales altos para identificar conglomerados de pobreza.

a) *Valores globales altos (VGA) en el proceso de concentración. Identificación de los VGA por remuestreo.* Los VGA pueden identificarse de muchas maneras dependiendo del objetivo del estudio. Este trabajo aplica tres procedimientos con distinto fin: remuestreo, discontinuidades naturales y puntas y colas consecutivas. Se utiliza el remuestreo para la identificación de los VGA en el conjunto de datos e identificar las áreas que concentran la variable de interés, independientemente de su localización. Los dos procedimientos restantes, discontinuidades naturales y puntas y colas, se reservan para clasi-

ficar los casos dentro de los conglomerados que identifica esta investigación. Esta decisión no implica que estos dos procedimientos no se puedan utilizar como alternativas para identificar los vga en vez del remuestreo, tal como se ilustra en Pavlova *et al.* (2014) y Jiang y Yin (2014). La decisión de utilizar el remuestreo para identificar los vga que representan el proceso de concentración general radica, por un lado, en su carácter general y, por otro, en su rigor estadístico. Al compararse con el remuestreo, las discontinuidades naturales y puntas y colas generan resultados más desagregados y útiles para la identificación de estratos o clases de la variable (pobreza). La descripción del remuestreo y los pormenores de su aplicación al estudio de caso se presentan a continuación. La explicación y aplicación específica de los procedimientos mencionados se detallan en la clasificación de los conglomerados.

*Remuestreo.* Los valores globales altos son valores no espaciales que son mayores al límite superior del intervalo de confianza de la media obtenido por remuestreo, para una probabilidad determinada (e. g.,  $p = 95\%$ , con una cola).<sup>3</sup> Los intervalos de confianza no son fiables si la distribución de los datos es asimétrica. Después de verificar los resultados de las pruebas usuales de normalidad (e. g., la significancia de la asimetría y la prueba de Kolmogorov-Smirnov), esta investigación aplica dos procedimientos robustos de remuestreo para lidiar con el problema de la asimetría: el remuestreo para intervalo de sesgo ajustado y acelerado (bca, por sus siglas en inglés) y remuestreo para intervalo sesgado (*Tilting*) (Chihara y Hesterberg, 2011: 112 y Hesterberg *et al.*, 2010: 16-19 y 16-32).

El punto de partida en el remuestreo es una muestra original representativa de la población.<sup>4</sup> El remuestreo con remplazo de la muestra original (no de la población) crea una distribución de muestras semejante a la que se obtendría si se tomaran varias muestras de la población original (distribución muestral). La forma y dispersión de la distribución remuestreada debería asemejarse a la distribución muestral. El remuestreo no resuelve los problemas de sesgo o asimetría en los datos originales. El remuestreo de intervalos para percentiles y el remuestreo de intervalos  $t$  no son confiables si la distribución del remuestreo no es normal. Por lo tanto, las dos opciones

<sup>3</sup> Hay varias alternativas para identificar los valores globales altos en distribuciones asimétricas, tales como los gráficos de cajas de bigotes o los valores- $z$  modificados con base en la mediana de las desviaciones absolutas (median absolute deviation,  $z$ -MAD).

<sup>4</sup> En inglés, *bootstraps* son las agarraderas para calzarse las botas. El término se utiliza como sinónimo de remuestreo con reemplazo. Este sinónimo es una metáfora basada en *Las sorprendentes aventuras del barón Munchausen* para decir “recobrarse o levantarse por esfuerzo propio”.

robustas en esta investigación son el remuestreo BCa y el remuestreo para intervalo sesgado (*Tilting*). La regla es muy simple: si nuestro programa de computadora proporciona estos remuestreos robustos y su muestra es representativa con más de 10 observaciones, siempre usaremos BCa o *Tilting* aplicando 10000 o más remuestras.<sup>5</sup> Ambos procedimientos generan resultados similares y mejoran la precisión de los intervalos de confianza ajustando los percentiles para corregir el sesgo y la asimetría (Hesterberg *et al.*, 2010: 16-32). En las distribuciones normales, todos los procedimientos generan resultados similares: remuestreo para intervalos de percentiles, remuestreo de intervalos *t*, BCa y *Tilting*. En esta investigación, todos los valores mayores al límite superior con una probabilidad  $p = 95\%$  con una cola, son considerados valores altos (de magnitud e intensidad). El remuestreo *Tilting* es utilizado en esta investigación para identificar los valores globales altos que caracterizan el proceso de concentración.

¿Por qué complicarse la vida con el remuestreo cuando es posible usar procedimientos no paramétricos convencionales para distribuciones asimétricas, tales como cajas de bigotes o valores *z-MAD*? Los métodos de remuestreo en la práctica funcionan significativamente mejor que los métodos alternativos. Además, el remuestreo permite inferencias sobre la población, mientras que los métodos alternativos se enfocan en (se reducen a) la muestra original. En el estudio de caso, el remuestreo identifica con bases estadísticas qué valor es significativamente alto. Esta sola característica no sólo justifica el remuestreo en esta investigación, sino en otros casos con tareas similares. Por ejemplo, Tian (2013), en sus comentarios y correcciones a O'Donoghue y Gleave (2004), sugiere este procedimiento para determinar la significancia estadística del cociente de localización.

b) *Valores locales altos (VLA) del proceso de aglomeración espacial. Identificación de los VLA por autocorrelación espacial local.* La autocorrelación espacial es una técnica estadística que mide la presencia y fortaleza de la interdependencia de los valores de una variable específica con referencia a los valores de la misma variable en las áreas aledañas o contiguas. Es la autocorrelación de una variable consigo misma en el espacio, generalmente medida por el índice global (I) o local (Ii) de Moran (Burt, Barber y Rig-

<sup>5</sup> Este trabajo utiliza SPLUS v.8.0 y su biblioteca de remuestreo que contiene las opciones BCa y *Tilting*. El paquete SPSS v.22 sólo proporciona la opción BCa. El límite de 10 o más observaciones no ha sido suficientemente probado, pero parece aceptable para el remuestreo (Hesterberg, comunicación personal).

by, 2009: 544). Una variable está autocorrelacionada si presenta un patrón espacial sistemático. En el Índice de Moran (I o Ii), este patrón puede ser identificado con el cero o con valores positivos o negativos. La hipótesis nula ( $H_0$ ) supone autocorrelación espacial cero y sugiere que el patrón espacial es aleatorio o que la variación espacial de los datos no tiene ninguna relación con su distribución espacial.

Una autocorrelación espacial positiva indica que los valores similares (altos o bajos) tienden a colocarse o a ser más similares que los más distantes. Existe autocorrelación espacial positiva si datos similares en intensidad (o magnitud) están cerca unos de otros. La autocorrelación espacial negativa, por otro lado, indica que las características o valores disímiles, como en un damero, tienden a estar cerca unos de otros: valores altos tienden a estar rodeados de valores bajos, y viceversa. Estas relaciones son la base del Diagrama de Dispersión de Moran (Anselin, 1995; 1996 y Anselin *et al.*, 2004). En el Diagrama de Moran, los valores en las abscisas están en unidades de desviación estándar, con media cero y varianza uno. En el eje de las ordenadas están los valores con rezago espacial (así se les llama a los valores de las áreas contiguas) de la variable estandarizada que aparece en el eje de las abscisas.

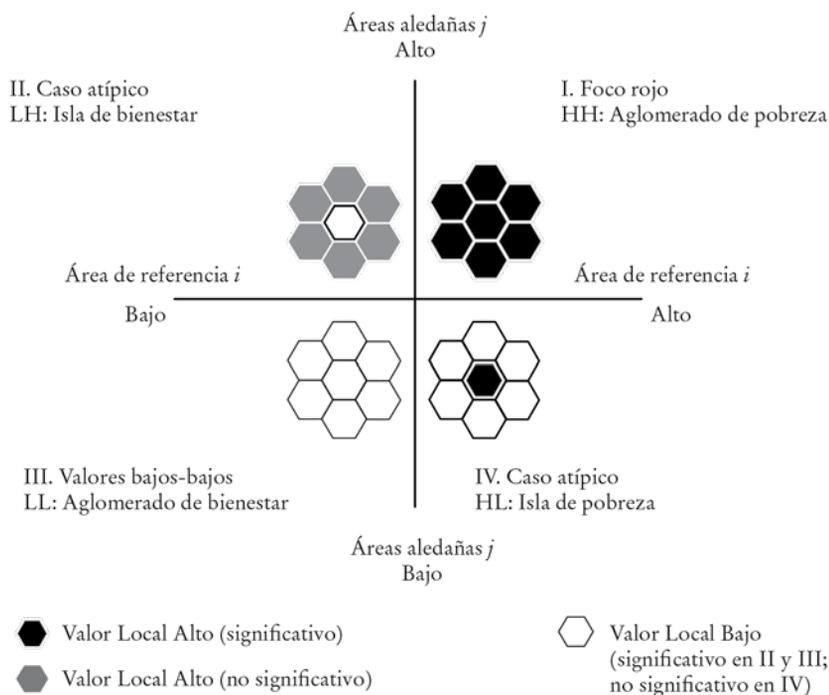
El Diagrama de Moran clasifica la autocorrelación espacial en dos categorías: aglomeraciones espaciales y casos espaciales atípicos (no deben confundirse con los casos atípicos globales superiores a las dos desviaciones estándar). Cada cuadrante en el diagrama corresponde a un tipo diferente de autocorrelación espacial (diagrama 1).

El cuadrante inferior a la izquierda (III) y el superior a la derecha (I) indican autocorrelación positiva, pero de distinto tipo. Mientras el cuadrante III contiene áreas con valores bajos rodeados de áreas con valores bajos (LL: *Low-Low*), el cuadrante I incluye áreas con valores altos rodeados por áreas con valores altos (HH: *High-High*).<sup>6</sup>

Estas diferencias entre los cuadrantes I y III muestran que las aglomeraciones identificadas por la autocorrelación espacial positiva pueden ser focos rojos (cuadrante I) o aglomeraciones de bienestar (cuadrante III). En contraste, las áreas del cuadrante superior izquierdo (II) y del inferior derecho (IV) sugieren autocorrelación espacial negativa. Los casos en los cuadrantes II y IV son casos espaciales atípicos. Mientras el cuadrante II

<sup>6</sup> Se conservan la terminología y las siglas en inglés para la fácil identificación de la literatura correspondiente, para la traducción al español y para la interpretación del estudio de caso.

DIAGRAMA 1. *Margaritas para LISA. Taxonomía espacial de la pobreza: focos rojos, núcleos de bienestar e islas de pobreza y bienestar<sup>a</sup>*



FUENTE: elaboración del autor con base en Anselin (1995).

<sup>a</sup> El título *Margaritas para LISA* busca asociar rápidamente el Diagrama de Dispersión Moran con los indicadores de autocorrelación espacial (LISA). Los centros y pétalos de las margaritas en HH y LL representan gráficamente casos que son estadísticamente significativos desde el punto de vista local: VLA en el cuadrante I y VLB en el cuadrante III. El Índice Local de Moran (Ii) es significativo para todas las observaciones que constituyen las margaritas en estos dos cuadrantes. En cambio, en LH y HL sólo el núcleo o centro es estadísticamente significativo. Los pétalos de las margaritas en los cuadrantes II y IV son VLA y VLB, respectivamente, no significativos estadísticamente. En estos valores periféricos el índice Ii no es estadísticamente significativo.

contiene valores bajos rodeados por valores altos (islas de bienestar), el cuadrante IV contiene valores altos rodeados por valores locales bajos (islas de pobreza).

Esta descripción de la autocorrelación espacial muestra que los valores locales *pueden ser* casos atípicos espaciales (LH o HL), focos rojos (HH) o aglomeraciones de bienestar (LL). El índice local de Moran (Ii), un indicador de autocorrelación espacial local (LISA, por sus siglas en inglés), proporciona la significancia estadística de las observaciones individuales en el Diagrama de Dispersión de Moran. Ii no sólo proporciona la significancia estadística

de los casos en las aglomeraciones espaciales (cuadrantes I y III) y los casos atípicos espaciales (cuadrantes II y III), sino que también identifica los casos que no son estadísticamente significativos.

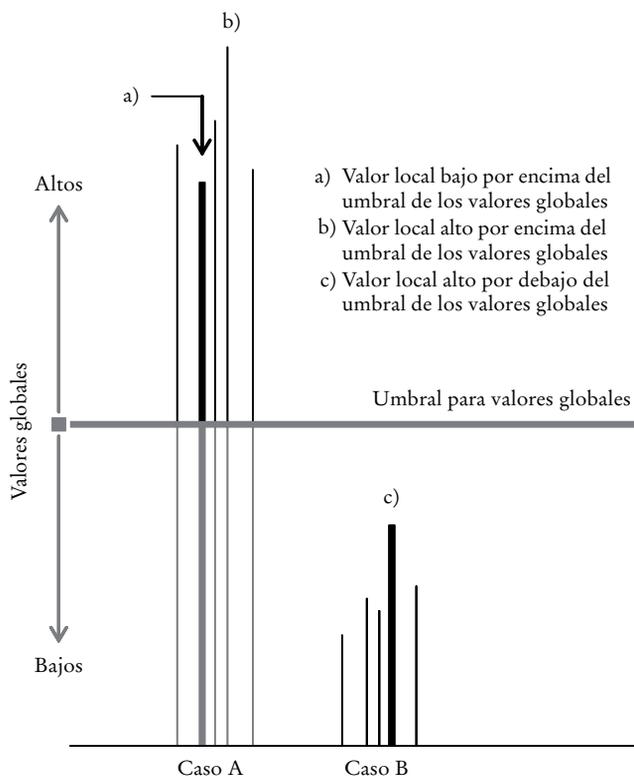
La significancia estadística del índice global I y local  $I_i$  se confronta con la hipótesis nula ( $H_0$ ) de ausencia de autocorrelación espacial (la variable sigue una distribución espacial aleatoria).  $H_0$  puede rechazarse o aceptarse bajo el siguiente principio: no significativo, no se rechaza. Por ejemplo, un índice global I significativo para una probabilidad  $p < 0.05$  indica que la ausencia de autocorrelación espacial ( $H_0$ ) se rechaza. Se concluye que el patrón espacial de la variable es aglomerado en el espacio. El programa *GeoDa* (disponible gratuitamente) genera el Diagrama de Dispersión de Moran y proporciona los valores y la significancia del índice global (I) y local de Moran ( $I_i$ ), los índices más utilizados para detectar los valores locales altos (Srinivasan, 2008: 615).

Los municipios que repiten la misma letra en esta clasificación son focos (HH—foco rojo) o núcleos de baja pobreza (LL—alto bienestar). Los municipios que mezclan letras (HL o LH) son áreas individuales diferentes a sus vecinos (casos atípicos espaciales). Hay una quinta categoría que debe considerarse en esta clasificación: los municipios que no son estadísticamente significativos. Como el interés central en este estudio es determinar los distintos grados de prioridad de las áreas de pobreza, la investigación se concentra en los focos rojos (HH) y las islas de pobreza (HL), localizados en el cuadrante I y IV, respectivamente.

c) *Identificación de conglomerados de intensidad o magnitud. Pensamiento fuera de la caja: superposición de VGA con VLA.* El análisis de superposición ya ha sido utilizado para integrar distintas variables de intensidad (e. g., educación y pobreza en Choudhury y Räder, 2014), pero, hasta donde sabemos, no ha sido utilizado para integrar simultáneamente la concentración y aglomeración de valores absolutos y relativos. Esta integración es muy importante considerando el hecho de que algunos VGA detectados por el análisis de concentración (por remuestreo o estadística descriptiva) pueden pasar desapercibidos en el análisis de aglomeración que identifica los VLA (por  $I_i$  en la autocorrelación espacial), y viceversa. Los VLA no son necesariamente valores altos desde el punto de la perspectiva nacional o global (diagrama 2).<sup>7</sup>

<sup>7</sup> Es importante mencionar que la posibilidad de tener VLA (focos rojos HH y casos del tipo HL) se incrementa si la base de datos original se suaviza o estandariza de manera que disminuya la influencia de los casos globales atípicos (Zhang *et al.*, 2008).

DIAGRAMA 2. *Empates y discrepancias en la superposición de valores globales altos y valores locales altos<sup>a</sup>*



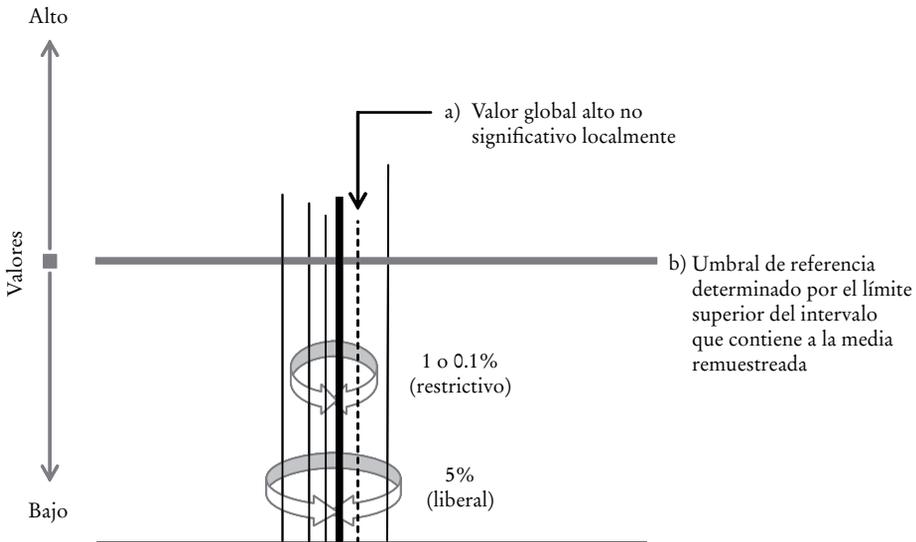
FUENTE: elaboración del autor.

<sup>a</sup> Caso A: Valor local bajo a) en un grupo de valores globales altos. No todos los valores globales altos son valores locales altos. Esto ocurre cuando el valor local es inferior al valor promedio de las áreas circundantes. Éste es el caso aun cuando los valores locales y globales sobrepasan el valor crítico obtenido por remuestreo de la media. Caso B: Valor local alto c) en un grupo de valores globales bajos.

Una crítica reciente a la estadística espacial, usando la famosa metáfora del elefante y los cinco ciegos, ilustra las limitaciones del conocimiento local en estos términos: “La imagen mental del elefante en la mente de los ciegos [...] refleja las verdades locales, que de hecho son parcialmente correctas, pero no muestran al elefante en su totalidad” (Jiang, 2015: 10). Esta crítica olvida la segunda parte del ejemplo clásico sobre los árboles y el bosque: es cierto que la atención en los árboles impide ver el bosque, pero también es cierto que la observación del bosque sacrifica la visión de los árboles. En el contexto de este trabajo, es cierto que los valores locales no muestran la perspectiva global, pero también es cierto que los valores globales ocultan la problemática local.

Estas observaciones manifiestan las limitaciones metodológicas de procedimientos que sólo utilizan estadística espacial o estadística no espacial en la identificación de *clusters* o jerarquización de una variable en el espacio. Algunos ejemplos son el uso único de la estadística espacial en la identificación de centros o subcentros urbanos —Baumont, Ertur y Le Gallo (2004: 153) y Arribas-Bel y Sanz-Gracia (2014)— o sólo la estadística descriptiva para la ubicación de *clusters* industriales (Ketels y Sölvell, 2005). Conviene anotar dos observaciones prácticas en estadística espacial: *i*) el recurso metodológico de relajar o restringir el nivel de significancia —mencionado en Baumont, Ertur y Le Gallo (2004)— no resuelve necesariamente la omisión de VGA, puesto que sólo incrementa la probabilidad de incluir más áreas aledañas (diagrama 3). Si se utilizara este procedimiento para identificar los núcleos y periferias de pobreza, el número de áreas contiguas incluidas sólo se expandiría horizontalmente (sería mayor), pero algunos VGA dentro de estos anillos de tolerancia pudieran omitirse porque no son localmente significativos. En esta opción, la inclusión potencial de valores globales altos

DIAGRAMA 3. Efecto de los criterios liberal o restrictivo en la detección de valores locales altos y posible omisión de valores globales altos<sup>a</sup>

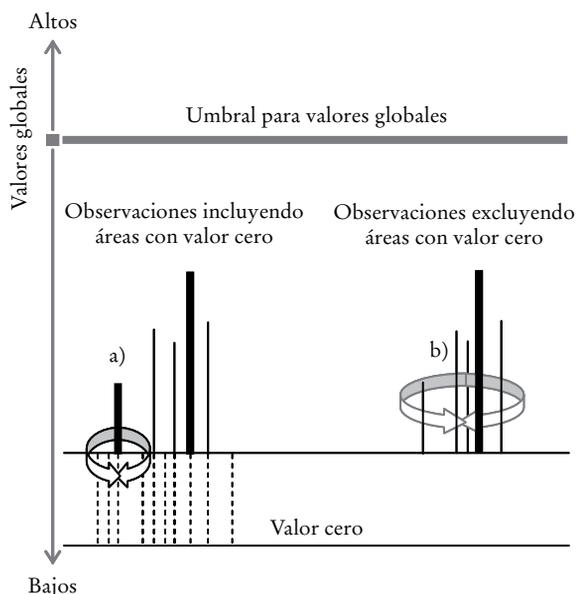


FUENTE: elaboración del autor.

<sup>a</sup> La línea vertical negra gruesa señala el área de referencia. Las líneas verticales delgadas representan los valores de las áreas vecinas. Las flechas circulares indican las áreas contiguas incluidas en un foco rojo (valores HH) bajo el criterio liberal o restrictivo.

que se hayan omitido no es segura y, de darse el caso, sería por azar; y *ii*) la ausencia de valores (valor cero) en algunas observaciones abre la posibilidad para que valores locales bajos sean considerados valores locales altos. El “salto” desde cero lleva a que un valor local bajo parezca alto cuando en realidad es bajo en el contexto del resto de valores aledaños (diagrama 4). Esta situación no existe en el estudio de caso pero conviene tenerla presente para posibles adaptaciones del procedimiento. Estas limitaciones justifican la necesidad de integrar la estadística no espacial (remuestreo) y estadística espacial (autocorrelación espacial) para medir y articular la magnitud e intensidad de dos procesos geográficos diferentes (concentración y aglomeración).

DIAGRAMA 4. *Efecto de polígonos con valor cero en la identificación de valores locales altos<sup>a</sup>*



FUENTE: elaboración del autor.

<sup>a</sup> La línea vertical negra y gruesa señala el área de referencia. Las líneas verticales delgadas representan los valores de las áreas vecinas. El inciso a) es el valor mínimo identificado como valor alto por la presencia de cedros en las áreas circundantes y el b) representa una base de datos sin ceros, valor mínimo identificado como valor bajo por la presencia de valores mayores en las áreas circundantes.

En suma, esta investigación sostiene que no todos los VLA coinciden con VGA, ni estos últimos concuerdan con los primeros. Por un lado, los VGA, identificados por remuestreo, tienen un límite estadísticamente definido

(el límite superior del intervalo de la media remuestreada, generalmente con una probabilidad de 95% con una cola). Por otro lado, los VLA, identificados mediante autocorrelación espacial, pueden existir si el valor del área de referencia es mayor que el promedio de los valores en las áreas aledañas, aunque estén por debajo del umbral que define a los valores globales altos —opción c) en el diagrama 2—. La autocorrelación local también podría omitir valores globales altos si el área de referencia está rodeada por valores globales que en promedio son aún más altos —opción a) en el diagrama 2—.

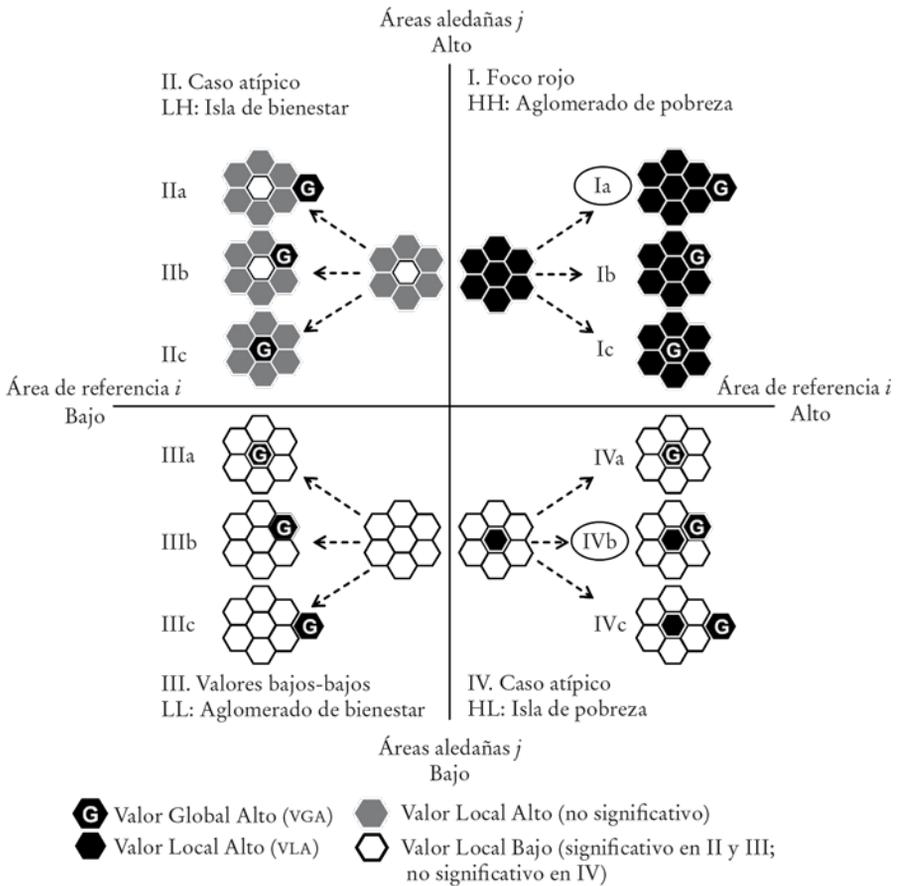
La superposición de VGA y VLA agranda los focos rojos o las islas de pobreza en dos situaciones posibles: *i*) cuando los VGA sean aledaños al foco rojo HH, o *ii*) cuando los VGA se traslapen con los valores bajos en los pétalos de las islas HL (cuadrante Ia y cuadrante IVb identificadas con un óvalo en el diagrama 5). En todos los demás casos la aglomeración inicial HH o HL identificada con autocorrelación espacial, después de la superposición de capas, es igual al conglomerado final porque los VGA y VLA se traslapan (Ib, Ic y IVa en el diagrama 5) o los VGA están distantes del núcleo —c) y d) en el diagrama 6—.

En la superposición de capas conviene destacar los siguientes criterios prácticos adicionales:

- i*) Las manchas continuas de VGA que se traslapen con la periferia de los VLA (HH), o pertenezcan a la misma, forman parte del conglomerado con núcleo HH (Ia, Ib y Ic en el diagrama 5).
- ii*) En esta investigación, no forma parte del conglomerado el resto de la mancha continua de VGA que se extiende más allá de la periferia inmediata de la aglomeración —diagrama 6, a) y b)—. Esta restricción puede relajarse a conveniencia para incluir en el conglomerado la segunda corona de VGA que aquí se excluye.
- iii*) Los VGA en el resto del país —diagrama 6, c) y d)— identifican áreas de alta concentración de la pobreza sin ser VLA.
- iv*) Los VLA que no son VGA o que se traslapan con ellos son registrados en cualquiera de los cuatro cuadrantes del Diagrama de Dispersión de Moran (diagrama 4 y 5).

Desde el punto de vista de los conglomerados de pobreza, tiene sentido que la política social incluya áreas localmente importantes aunque no

DIAGRAMA 5. Identificación de conglomerados combinando los procesos de aglomeración y concentración<sup>a</sup>

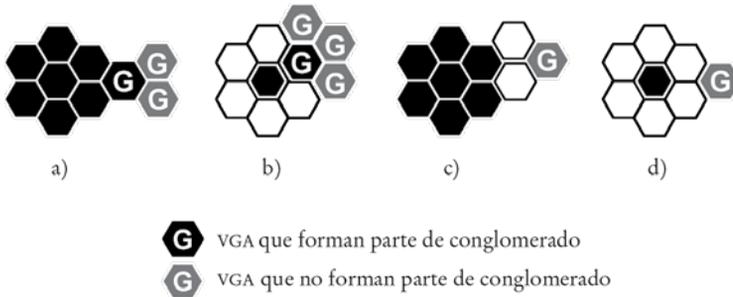


FUENTE: elaboración del autor.

<sup>a</sup> La formación de conglomerados es indicada con un óvalo (Ia y IVb). VGA: valores mayores al límite superior del intervalo de la media remuestreada en distribuciones con asimetría positiva. VLA: casos identificados por el índice local de Moran (Ii) con significancia estadística determinada mediante permutaciones condicionadas.

sean globalmente significativos. Sin embargo, sería absurdo que tal política incluyera sólo áreas importantes desde el punto de vista local al tiempo que excluyera áreas aledañas que sólo son relevantes desde la perspectiva nacional por no contener valores locales altos. Ejemplo de esta omisión importante son los valores globales altos no significativos localmente que constituyen los conglomerados que se forman a partir de los focos rojos e islas de pobreza.

DIAGRAMA 6. *VGA que forman y no forman parte de un conglomerado*



FUENTE: elaboración del autor.

En suma, la identificación de conglomerados de intensidad, por un lado, y de magnitud, por otro, es posible mediante la superposición de valores globales altos y valores locales altos (HH y HL). El análisis de superposición en los Sistemas de Análisis Geográfico integra matemáticamente (e. g., por unión o intersección) las capas de concentración y aglomeración para crear una nueva capa que contenga el conglomerado resultante. Los núcleos (focos rojos HH o casos atípicos espaciales HL) añaden a su periferia valores globales altos contiguos para formar conglomerados. Este procedimiento se aplica primero a valores relativos. Después, como una tarea independiente y separada, el análisis se repite para valores absolutos.

d) *Identificación de áreas prioritarias por superposición de conglomerados de magnitud e intensidad.* En este paso se aplican diagramas de Venn para representar los conglomerados de intensidad y magnitud. Los casos de mayor prioridad se localizan en la intersección de estos dos subconjuntos. Esta intersección contiene conglomerados de valores altos tanto de magnitud como de intensidad. Por esta razón, ellas reciben la prioridad mayor (prioridad uno).

e) *Relevancia espacial de las áreas que no se intersectan y que se localizan al interior de los conglomerados.* Una vez identificados los municipios localizados en la intersección, las áreas remanentes al interior de cada conglomerado pueden estratificarse. Esta estratificación proporciona una clasificación de áreas que complementa a los municipios de mayor prioridad. La identificación de las áreas de mayor prioridad y la estratificación de las áreas que no se intersectan proporcionan una jerarquía espacial que puede guiar la asignación de recursos sociales, la toma de decisiones en política

pública o estudios futuros sobre el patrón espacial de la pobreza. Entre las alternativas para estratificar o asignar prioridad a los conglomerados fuera de la intersección sobresalen dos procedimientos de clasificación: discontinuidades naturales y puntas y colas.

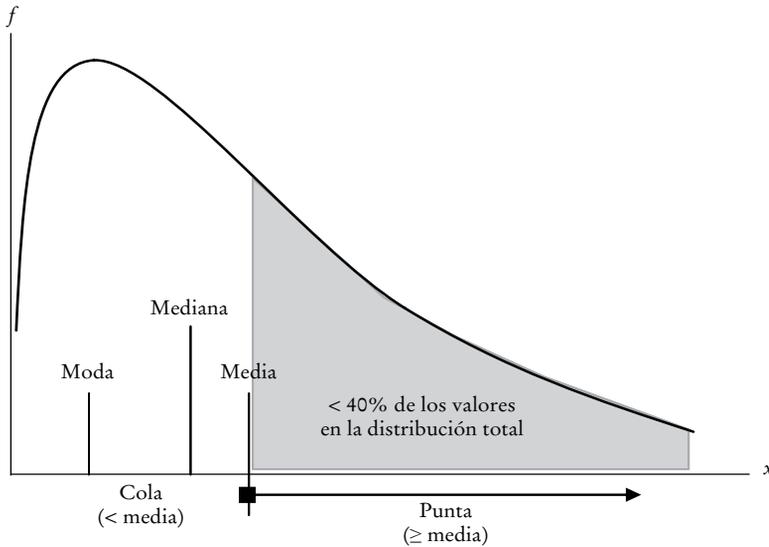
*Discontinuidades naturales o clasificación de Jenks.* Este procedimiento busca quiebres o saltos en los datos observados para agrupar casos en clases. Con base en interacciones estadísticas (optimización de Jenks), este método identifica la mejor opción definida como la clasificación que minimiza la distancia entre los valores similares al interior de cada clase y maximiza la distancia entre ellas. Aunque este método requiere de una distribución normal, generalmente se aplica a información asimétrica previamente estandarizada manipulando temporalmente los casos extremos, como en Crews y Peralvo (2008: 70).

*Puntas y colas consecutivas con base en Jiang (2015).* Los principales pasos en este método especialmente diseñado para distribuciones muy asimétricas son:

- i) Encontrar la media del total de valores (absolutos o relativos). En una distribución sesgada a la derecha (asimetría positiva), los valores igual o superiores a la media son el pico o punta y los que están por debajo de la media son la cola. Nótese que los términos “puntas” y “colas” se refieren a la distribución rango-tamaño y no a la función de densidad de la probabilidad. La punta o pico contiene los pocos valores altos, mientras que la cola contiene la mayoría de los valores menores (gráfica 1).
- ii) La punta se subdivide de nuevo en pocos valores grandes y muchos valores pequeños, con base en la media aritmética de los valores contenidos en ella. Esto genera otra punta y cola “anidadas” dentro de la punta identificada en el paso previo, como ocurre con las muñecas rusas que contienen dentro de sí otra muñeca y dentro de ésta otra y otra.
- iii) El proceso continúa para cada punta sucesivamente identificada hasta que las proporciones de valores en la punta y la cola sean aproximadamente las mismas o ya no haya valores que subdividir.

Como regla de dedo, “los porcentajes [de casos] en las puntas deben ser menores a 40%. Esta condición puede extenderse para muchos aspectos geográficos hasta 50% o más si la punta retiene menos de 40% en niveles jerárquicos posteriores” (Jiang y Yin, 2013: 8).

GRÁFICA 1. *Puntas y colas en el procedimiento de clasificación de puntas/colas de una distribución paretiana con asimetría a la derecha*



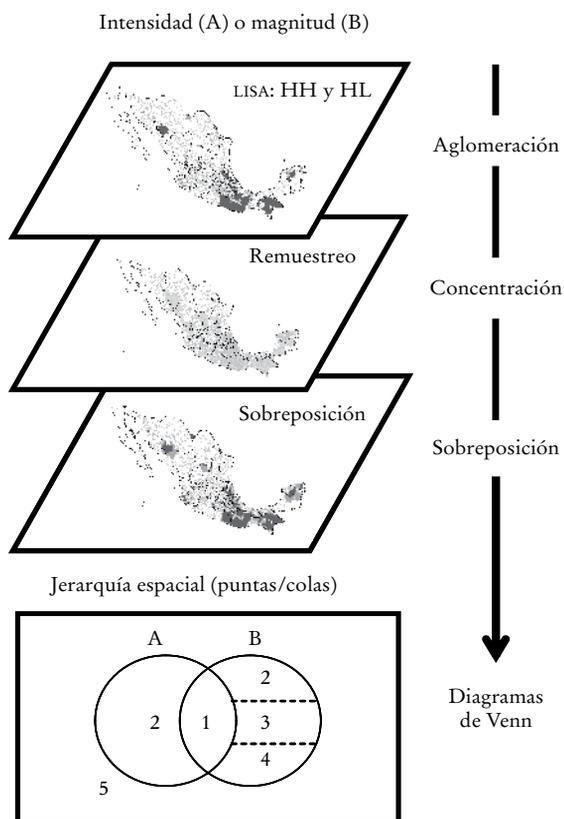
FUENTE: elaboración del autor con base en Jiang (2015). Nótese que la punta en esta gráfica corresponde a la cola en la distribución gaussiana.

e) *Relevancia espacial de las áreas que se localizan fuera de los conglomerados.* Las áreas fuera de los conglomerados tienen una importancia espacialmente diferenciada porque contienen valores globales altos. Estos valores globales altos, si fuera necesario, también pueden clasificarse por los métodos sugeridos para las áreas que no se intersectan dentro de los conglomerados, dependiendo de la simetría o asimetría de su distribución —véase punto d) previo—.

En suma, un enfoque que diferencie y combine los procesos de concentración y aglomeración de valores relativos y absolutos tiene varias alternativas para asignar un orden espacial a la variable de interés. La opción sugerida en este trabajo puede confrontarse con los procedimientos que utilizan valores críticos, como el utilizado por Sedesol (2013), o criterios mixtos (Graw y Husmann, 2014). Los resultados también podrían compararse con variaciones del procedimiento propuesto. Por ejemplo, si se utilizan datos estandarizados de magnitud, el número de áreas en los núcleos correspondientes podría incrementarse.

El diagrama 7 sintetiza de manera gráfica los principales pasos metodológicos descritos en esta sección.

DIAGRAMA 7. Esquema metodológico para identificar la jerarquía espacial de la pobreza



FUENTE: elaboración del autor.

## II. COMPARATIVO DEL PROCEDIMIENTO PROPUESTO CON OTRAS TÉCNICAS EMPLEADAS PARA LA IDENTIFICACIÓN DE CONGLOMERADOS

En general, tres enfoques con diversas variantes sobresalen en la literatura que identifica las concentraciones o aglomeraciones de variables sociales, incluyendo la pobreza: *i*) enfoque del valor crítico de valores globales para áreas individuales, *ii*) enfoque de valores locales altos para áreas contiguas, y *iii*) enfoque de valores globales altos para variables relativas y absolutas en polígonos contiguos o aislados. Esta investigación propone un cuarto enfoque: superposición de VGA y VLA de variables absolutas y relativas.

### 1. *Valores críticos*

Es un método muy controvertido y tiene muchas versiones en la literatura actual. En particular, dos características definen el enfoque del valor crítico e identifican sus problemas principales:

*i) Valor crítico por ensayo y error.* Esta versión determina valores críticos para identificar áreas con base en un proceso de ensayo y error, guiado por el conocimiento local (McMillen, 2001). Limitaciones: este proceso lleva a la selección de diversos valores críticos para áreas específicas de estudio, por lo que se limita la comparabilidad de resultados en el tiempo y el espacio (Riguelle, Thomas y Verhetsel, 2007: 198).

*ii) Valor crítico con técnicas estadísticas.* Existen versiones que imponen un valor crítico global único para todas las subáreas con asistencia estadística formal. La selección de este valor crítico único puede definirse por medio de estadística paramétrica o no paramétrica con base en una probabilidad específica (*e. g.*,  $p = 95\%$ ), no requiere de un conocimiento pleno del área de estudio y típicamente utiliza datos relativos —*e. g.*, selección de áreas por encima del 90° percentil, como en Ketels y Sölvell (2005)—.

### 2. *Identificación de valores locales altos según áreas contiguas*

Este enfoque utiliza la autocorrelación espacial para explorar las áreas aleatorias a cada observación con el fin de identificar VLA. Este procedimiento puede registrar VLA que no alcancen un umbral global predefinido pero puede omitir VGA en las circunstancias que se revisan en este escrito. Limitaciones: aunque en autocorrelación espacial pueden utilizarse valores absolutos, los estudios generalmente emplean sólo variables relativas (*i. e.*, Riguelle, Thomas y Verhetsel, 2007). También existe la posibilidad de que algunos VGA no sean detectados por el escrutinio estadístico que detecta los valores locales altos.

### 3. *Identificación de valores globales altos para variables relativas y absolutas en polígonos contiguos o aislados*

Este procedimiento es una variación reciente del enfoque de valor crítico con técnicas estadísticas (Van Den Heuvel *et al.*, 2012). El procedimiento

sólo incluye las áreas donde existe la variable y se resume en dos pasos consecutivos:

i) Establecer valores críticos para datos absolutos (número de pobres) y relativos (coeficiente de localización, LQ por sus siglas en inglés). Seleccionar, por ejemplo, las áreas con un  $LQ > 1$  o si el valor absoluto es superior a la media (mediana, en distribuciones asimétricas). Descartar para el resto del análisis las áreas por debajo de estos valores críticos.

ii) En las áreas seleccionadas en el paso previo, combinar las áreas contiguas y establecer nuevos criterios de selección. Los autores citados arriba descartan la posibilidad de utilizar la autocorrelación espacial porque sólo opera con una variable y el procedimiento necesita combinar simultáneamente dos variables (valores relativos y absolutos de una misma variable). Los autores optan por valores críticos de conveniencia, según las necesidades de la investigación (Van Den Heuvel, 2012: 7-8). En su ejemplo para todas las áreas del segundo paso (combinadas e individuales), los autores establecen como valor crítico el 90° percentil para los valores absolutos (sólo las áreas en 10% superior califican en este paso). El valor crítico para el LQ es el segundo o 90° percentil, dependiendo de cuál se alcance primero. Si el LQ es 1.9 en el 90° percentil, el valor crítico lo determina dicho percentil. Este ejemplo indica que un  $LQ = 2$  está en un nivel superior al 90° percentil. Los criterios de selección en este procedimiento se parecen mucho al enfoque no espacial de “agrupación de tres estrellas” de Ketels y Sölvell (2005: 16), pero aplicado a valores absolutos y relativos en una preselección de áreas contiguas combinadas y de áreas aisladas.

Limitaciones: el método sólo aplica criterios globales aunque se combinen áreas contiguas. Por esta razón, el procedimiento detecta aglomeraciones de vga, no de vla. Esto puede llevar a que se omita la selección de áreas importantes en el ámbito local que no alcanzan el umbral global predefinido.

#### 4. *Superposición de vga y vla de variables absolutas y relativas, propios de esta investigación*

Ningún estudio revisado combina simultáneamente los procesos de concentración y aglomeración en variables absolutas y relativas. Hay versiones mixtas en estudios previos que aplican *parcialmente* el enfoque propuesto en esta investigación. La reseña y evaluación de la bibliografía ilustrativa

muestra que no hay un procedimiento que combine todo a la vez (cuadro 1). Los procedimientos que sólo utilizan estadística no espacial, como Ketels y Sölvell (2005), omiten el efecto de la autocorrelación espacial, relevante en el ámbito local. Los procedimientos que sólo utilizan estadística espacial, como Riguelle, Thomas y Verhetsel (2007), corren el riesgo de omitir vga que son relevantes en una política nacional.

Los procedimientos que utilizan tablas cruzadas para incluir tanto el proceso de concentración espacial como de aglomeración espacial, como Carroll, Reid y Smith (2008), omiten la magnitud o la intensidad en uno de los dos procesos. Estos autores identifican *clusters* regionales potenciales para la industria combinando valores absolutos y relativos en tablas de doble entrada. Al efecto cruzan vga del cociente de localización (LQ, una medida local expresada en términos globales) y vla de magnitud (número de empleados). La tipificación de Carroll, Reid y Smith (2008) omite el proceso de aglomeración en la variable relativa y el proceso de concentración en la variable absoluta.

Las investigaciones que combinan magnitud e intensidad sólo consideran la concentración y omiten la aglomeración (Chainey y Desyllas, 2008; Van Den Heuvel *et al.*, 2012 y Graw y Husmann, 2014). Por una lado, en un estudio sobre el proceso de concentración de la criminalidad, Chainey y Desyllas (2008) combinan valores relativos (niveles alto, medio y bajo de la tasa de robos) y absolutos (niveles alto, medio y bajo del número de robos). Esta clasificación de la criminalidad omite el proceso de aglomeración tanto en la magnitud como en la intensidad. El trabajo de Graw y Husmann (2014), por otro lado, utiliza técnicas de los sistemas de información geográfica, juicio de experto y criterios estadísticamente definidos para clasificar variables de magnitud e intensidad. Esta investigación, sin embargo, también pasa por alto el proceso de aglomeración. Finalmente, la metodología propuesta por Van Den Heuvel *et al.* (2012), descrita previamente, también omite el proceso de la aglomeración espacial.

Los estudios que analizan la concentración y la aglomeración simultáneamente en variables de intensidad (LQ) mediante técnicas estadísticas tradicionales y autocorrelación espacial, omiten la magnitud, como De Dominicis, Arbia y De Groot (2013). Una tabla de doble entrada facilita la interpretación de estas omisiones, sintetiza la bibliografía ilustrativa de esta reseña y muestra comparativamente las bondades de la metodología propuesta (cuadro 1).

CUADRO 1. *Comparación de procedimientos sobre la elaboración de conglomerados en bibliografía ilustrativa<sup>a</sup>*

Tema del estudio	Proceso espacial			
	Concentración (no espacial)		Aglomeración (espacial)	
	Absolutos	Relativos	Absolutos	Relativos
Industria				
Ketels y Sölvell, 2005	X	✓	X	X
Industria				
Riguelle, Thomas y Verhetsel, 2007	X	X	X	✓
Industria				
Van Den Heuvel <i>et al.</i> , 2012	✓	✓	X	X
Marginación				
Graw y Husmann, 2014	✓	✓	X	X
Crimen				
Chainey y Desyllas, 2008	✓	✓	X	X
Industria				
Carroll <i>et al.</i> , 2008	X	✓	✓	X
Industria				
De Dominicis, Arbia y De Groot, 2013	X	✓	X	✓
Pobreza				
Esta investigación	✓	✓	✓	✓

FUENTE: elaboración del autor.

<sup>a</sup> Una vez establecidos los criterios de evaluación, el término “bibliografía ilustrativa” es preferido al de “bibliografía representativa” en la perspectiva metodológica de este trabajo y en una época de información abundante.

La metodología sugerida en esta investigación incorpora tanto el proceso de concentración como el de aglomeración en variables absolutas y relativas. El procedimiento sugerido registra las deficiencias y evita las omisiones en la bibliografía revisada mediante la superposición de capas, al tiempo que incorpora ideas y técnicas estadísticas de temas diversos para identificar conglomerados y establecer una jerarquía espacial de la pobreza. Las limitaciones del procedimiento se presentan en las conclusiones del estudio de caso.

### III. ESTUDIO DE CASO

#### 1. Preguntas de investigación e hipótesis de trabajo

En un escrito sobre métodos, la estructura lógica, las aportaciones, bondades y desventajas del procedimiento demandan la atención principal. Lo que importa es la sustancia del contenido y, en este caso, la novedad que se presume. En este trabajo, se plantean preguntas de investigación e hipótesis

de trabajo asociadas para facilitar el proceso de investigación, la organización del escrito y la comunicación de sus resultados. En realidad el uso de preguntas de investigación e hipótesis de trabajo agrega poco al valor de la investigación y es más un asunto de preferencia del investigador y del estilo de redacción (Yeager, 2008: 52). Las preguntas de investigación y la especificación de las hipótesis correspondientes en este estudio no son muy distintas a las planteadas por estudios similares para una variable georeferenciada, aunque con procedimientos e inquietudes distintos (Anselín, Murray y Rey, 2007):

*i) ¿Está la pobreza aleatoriamente distribuida entre los municipios de México?*

*Hipótesis:*  $H_0$ : La pobreza está distribuida aleatoriamente entre los municipios de México.  $H_1$ : No se acepta  $H_0$ .

*Implicación.* Si no se rechazara  $H_0$ , la política contra la pobreza espacialmente focalizada no tendría ningún sentido, puesto que la pobreza estaría aleatoriamente distribuida.

*ii) ¿Hay aglomeraciones de municipios pobres o focos rojos espaciales de pobreza compuestos por varios municipios?*

*Hipótesis:*  $H_0$ : No hay aglomeraciones de municipios pobres.  $H_1$ : Hay municipios con altos niveles de pobreza rodeados por municipios igualmente pobres.

*Implicación.* Si no se rechazara  $H_0$ , no existirían aglomeraciones de áreas de pobreza ni necesidad de intervenir o focalizar espacialmente los recursos y esfuerzos sociales.

*iii) ¿Pueden los municipios ser ordenados o jerarquizados considerando simultáneamente la intensidad de la pobreza, el número de pobres y su distribución espacial?*

*Hipótesis:*  $H_0$ : La ausencia de una metodología que articule simultáneamente VGA y VLA de variables en términos relativos y absolutos, demuestra que su existencia no es posible; su inexistencia así lo evidencia.  $H_1$ : No se acepta  $H_0$ . La ausencia de una metodología como la descrita sólo es señal de la evolución incompleta del pensamiento espacial integrado a temas sociales.

*Implicación:* Si no se rechazara  $H_0$ , se tendrían dos problemas importantes: *a)* el uso único de valores relativos subestimaría el problema de la pobreza en las áreas densamente pobladas y lo sobrestimaría en áreas pe-

queñas o escasamente habitadas (lo contrario ocurre para el uso único de valores absolutos), y *b*) además, la falta de correspondencia entre los vga y vLa, no sólo omite la complementariedad de los procesos de concentración y de aglomeración, sino que identifica conglomerados incompletos. Estos dos problemas complican el trabajo técnico y la existencia de criterios objetivos para asignar los recursos sociales, como ocurre actualmente.

## 2. Resultados y prueba de hipótesis

Esta sección reporta los resultados obtenidos con el procedimiento esbozado en la sección metodológica. Tiene tres partes principales. Siguiendo los pasos sugeridos en el diagrama 8, la intensidad se aborda primero. Se mide con el cociente de localización. Se aplica el remuestreo para determinar los valores globales de intensidad que son significativamente altos. Este procedimiento estadístico identifica los valores globales altos independientemente de su localización, como resultado del proceso de concentración de la pobreza. Se remuestran las 2456 observaciones, como lo sugiere Tian (2013) en su opción alternativa al cociente de localización estandarizado (SLQ, *Standardized Location Quotient*) de O'Donoghue y Gleave (2004).<sup>8</sup> En el caso del LQ, no tiene sentido considerar como “valor alto” un cociente menor a uno. Esta investigación verifica que el valor crítico definido por el límite superior de la media remuestreada siempre sea igual o mayor a uno. Todos los LQ remuestreados reportados como valores altos deben ser igual o mayor a uno.

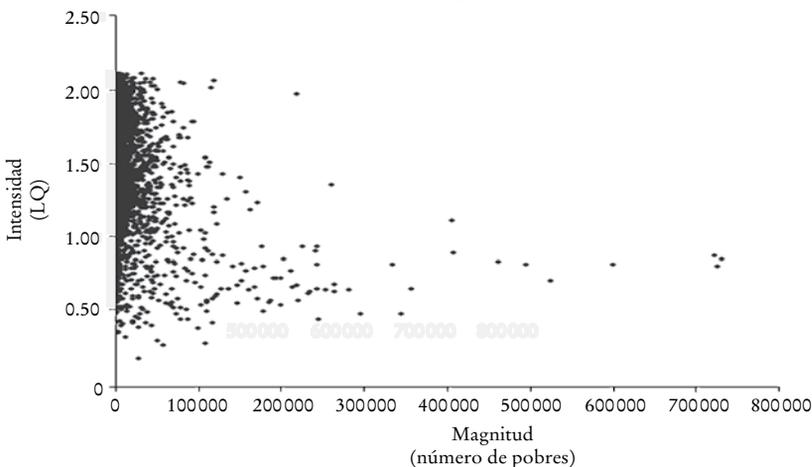
La autocorrelación espacial identifica la aglomeración espacial de los vLa (núcleos reconocidos como valores HH y HL). Los vga y vLa representan el proceso de concentración y aglomeración, respectivamente. Los valores para estos dos procesos son sobrepuestos para identificar los conglomerados de intensidad. La segunda parte sigue los mismos pasos descritos para la intensidad a fin de identificar los conglomerados de magnitud. La sección final sobrepone los conglomerados de intensidad y magnitud creando una jerarquía espacial. Las áreas de pobreza con prioridad más alta se localizan en la intersección de estos dos conglomerados. Las áreas remanentes dentro de los conglomerados (fuera de la intersección) son estratificadas siguiendo

<sup>8</sup> Debe notarse que Tian (2013) considera el remuestreo *en general* como una alternativa del SLQ. Como se establece en la metodología, el remuestreo no corrige la asimetría en la muestra original. La significancia estadística en las distribuciones asimétricas demanda el uso de las opciones robustas BCa o *Tilting* porque las opciones del percentil remuestreado y los intervalos *t* remuestreados no son confiables para distribuciones no normales.

do el procedimiento de “puntas y colas” sugerido recientemente por Jiang (2015). El producto final es un orden territorial integrado resultante de la combinación simultánea de la concentración y aglomeración de los valores absolutos y relativos de la pobreza.

a) *Magnitud e intensidad*. La mayoría de las personas pobres vive en áreas con baja intensidad de pobreza (gráfica 2). Por lo tanto, se espera que en su mayoría los conglomerados de intensidad sean distintos de los conglomerados de magnitud. Los valores altos de intensidad están en municipios de baja magnitud. En contraste, un gran número de personas pobres vive en áreas de intensidad baja o moderada. Las áreas de alta intensidad de la pobreza contienen sólo un pequeño número de pobres. Tal como lo notó Visvalingam (1983) hace tres décadas, estas líneas muestran que las políticas espacialmente orientadas que se basan solamente en intensidad (capturada por medidas relativas, tales como porcentajes, proporciones o tasas) exageran el problema en asentamientos pequeños y lo minimizan en áreas con una gran masa de pobres. Por otro lado, las políticas de combate a la pobreza solamente basadas en magnitud (expresada en números absolutos) discriminan contra la gran proporción de pobres en asentamientos pequeños. Por lo tanto, los valores absolutos y relativos deben combinarse simultáneamente en una sola clasificación o jerarquía territorial para sustentar políticas espacialmente focalizadas.

GRÁFICA 2. Pobreza. Intensidad (LQ-pobres) vs magnitud  
(número de pobres)



FUENTE: elaboración del autor.

b) *Intensidad (LQ). Proceso de aglomeración y núcleos. Autocorrelación espacial.* El índice global de Moran (I) para intensidad (LQ), utilizando una matriz de contigüidad del tipo reina, rechaza la hipótesis nula de aleatoriedad. El índice global I indica que la intensidad de la pobreza está espacialmente concentrada en el país ( $I = 0.6796, p = 0.0001$ ; para 9 999 permutaciones). El índice local de Moran (Ii), con una probabilidad  $p < 5\%$ , identifica 565 núcleos de intensidad (HH) y 20 islas de pobreza (HL). Aunque los municipios del tipo HH y HL son distintos, por razones prácticas, esta investigación considera los dos tipos como núcleos de conglomerados potenciales de pobreza.

La mayoría de los núcleos de intensidad se localiza en el sur de México, la Sierra de Puebla y algunos municipios del sureste del país (gráfica 2). Estas áreas son la semilla o punto de partida para la identificación de los conglomerados de intensidad. La periferia de estos núcleos la constituyen los valores globales altos del proceso de concentración, que se obtienen por el remuestreo de la media de los LQ.

*Proceso de concentración. Remuestreo.* Los procedimientos robustos más directos para encontrar el límite superior del intervalo de confianza de la media son el bca o el de distribuciones sesgadas (*Tilting*). Como los resultados de ambos procedimientos son similares, el estudio de caso toma de los resultados del *Tilting* la opción más refinada. En el caso de incidencia, el valor obtenido por el *Tilting* con 95% en la cola derecha, para 10 000 réplicas de la media inicial de los 2 456 casos, es  $LQ = 1.4705$ . Hay 1 303 municipios por encima de este valor (el mapa 1 muestra su ubicación geográfica). Algunos de los VGA empatan con los núcleos, otros son su periferia o son meramente áreas no aglomeradas de alta concentración localizadas en el resto del país.

Todos los 2 456 casos son incluidos. Si sólo se consideraran los casos con  $LQ \geq 1$ , el valor crítico sería 1.59 para 95% de confianza en la cola derecha utilizando la opción *Tilting*. Como el promedio nacional (el denominador) en LQ se obtiene con base en todos los casos, se considera conveniente tomar el resultado del remuestreo para el conjunto completo de datos en vez de asumir sólo los valores donde  $LQ \geq 1$ .

*Superposición de la concentración y aglomeración.* El análisis de superposición integra los núcleos (VLA: HH y HL de LQ) con VGA ( $LQ \geq 1.4705$ ) para generar conglomerados de intensidad. La superposición identifica los conglomerados de intensidad en dos pasos:

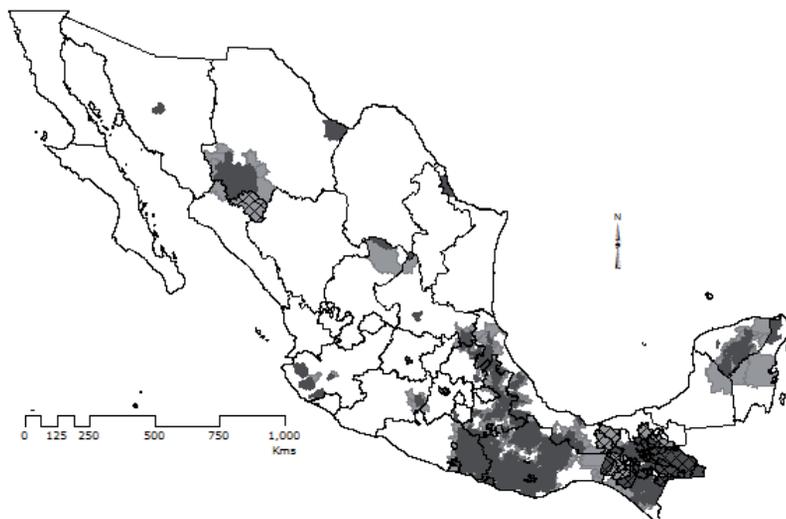
- i) Preparación de los núcleos. Intersección de las áreas HH y HL con la capa de municipios que contiene los valores  $LQ \geq 1.4705$ . Los núcleos son los valores HH y HL de LQ. El mapa de integración muestra los VGA que también son núcleos porque empatan con los VLA. En esta superposición de capas, conviene aclarar que los valores HH y HL son núcleos, aunque no empaten con los valores globales altos.
- ii) Identificación de conglomerados. Un conglomerado está constituido por un núcleo y su periferia. La periferia son VGA (áreas en gris claro, donde  $LQ \geq 1.4705$ ) *contiguos* a las áreas HH y HL (en gris oscuro en el mapa 1). Como lo establece el paso previo, los VLA que coinciden con los valores HH y HL son considerados núcleos.

c) *Magnitud. Proceso de aglomeración y núcleos. Autocorrelación espacial.* Aunque los pasos metodológicos para la magnitud e intensidad son los mismos, los resultados a nivel local son distintos. El índice global de Moran (I) para magnitud (número de pobres), utilizando una matriz de contigüidad del tipo reina, rechaza la hipótesis nula de aleatoriedad. El índice global I indica que la magnitud de la pobreza está espacialmente concentrada en el país ( $I = 0.279$ ;  $p = 0.0001$ ; para 9999 permutaciones). El índice local de Moran (Ii), con una probabilidad  $p < 5\%$ , identifica 146 núcleos de magnitud (HH) y 12 islas de pobreza (HL) (mapa 1).

La mayoría de los núcleos de magnitud se localiza en áreas densamente pobladas, tales como las ciudades capital. La periferia de estos núcleos la constituyen los VGA que tipifican el proceso de concentración; éstos son obtenidos por remuestreo de la media de la magnitud.

*Proceso de concentración. Remuestreo. Patrón espacial de la magnitud en municipios con población igual o mayor a 23 232 pobres.* En el caso de la magnitud, el remuestreo por *Tilting* a 95% en la cola derecha, para 10000 muestras, es 23 232 pobres. Como en LQ, el remuestreo utiliza todos los 2456 casos. Hay 534 municipios por encima de este valor (mapa 2). Como es de esperar, los VGA de magnitud se localizan en las áreas más pobladas. Resultados similares se esperarían para la venta de Coca-Cola o de cualquier otra variable relacionada directamente con la población. Un patrón similar se confirma para el proceso de aglomeración de los valores absolutos de la pobreza.

*Superposición de concentración y aglomeración de la magnitud.* La superposición integra los núcleos (HH y HL) con los VGA para identificar los

MAPA 1. *Conglomerados de intensidad identificados por superposición de capas<sup>a</sup>*

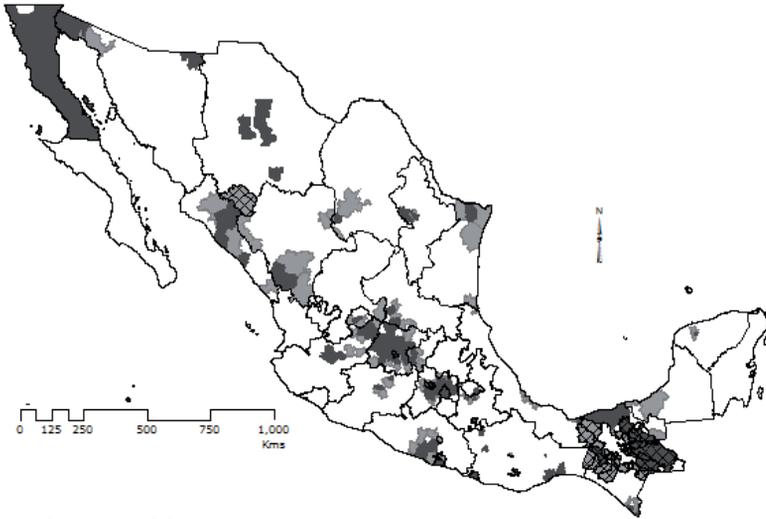
FUENTE: elaboración del autor.

<sup>a</sup> Núcleos (HH y HL en gris oscuro) y periferias (en color gris claro). Los municipios cuadriculados se localizan en el área de intersección de los conglomerados de intensidad y magnitud (prioridad 1, en el cuadro 2). Las áreas fuera de la intersección son estratificadas conforme a los límites de clase en el cuadro 3 (no representados en el mapa). La magnitud se refiere al número de pobres en el 2010 tal como lo reporta el Coneval.

conglomerados de magnitud (mapa 2). La superposición identifica los conglomerados de magnitud en dos pasos (mapa 2):

- i) Intersección de áreas HH y HL con áreas que contienen 23 232 pobres o más. Este paso integra en un solo mapa valores locales y globales superpuestos.
- ii) Unión de áreas HH y HL (áreas en gris oscuro) con las áreas aledañas con VGA (áreas en gris claro, con 23 232 pobres o más).

d) *Áreas de mayor prioridad. Superposición de los conglomerados de intensidad y magnitud.* Una vez identificados los conglomerados de magnitud e intensidad, es posible identificar y ordenar las áreas de pobreza en términos de ambas características. Esta sección superpone los conglomerados de datos relativos (mapa 1) y absolutos (mapa 2) para identificar las áreas de pobreza con la prioridad más alta. El procedimiento completo es esbozado en el diagrama 8. Los resultados registran 44 áreas de mayor prioridad (prioridad 1), localizadas en nueve estados: Oaxaca, Chihuahua, Veracruz, Guerrero, Guanajuato, Puebla, México, Tabasco y Chiapas (diagrama 8 y

MAPA 2. Conglomerados de magnitud identificados por superposición de capas<sup>a</sup>

FUENTE: elaboración del autor.

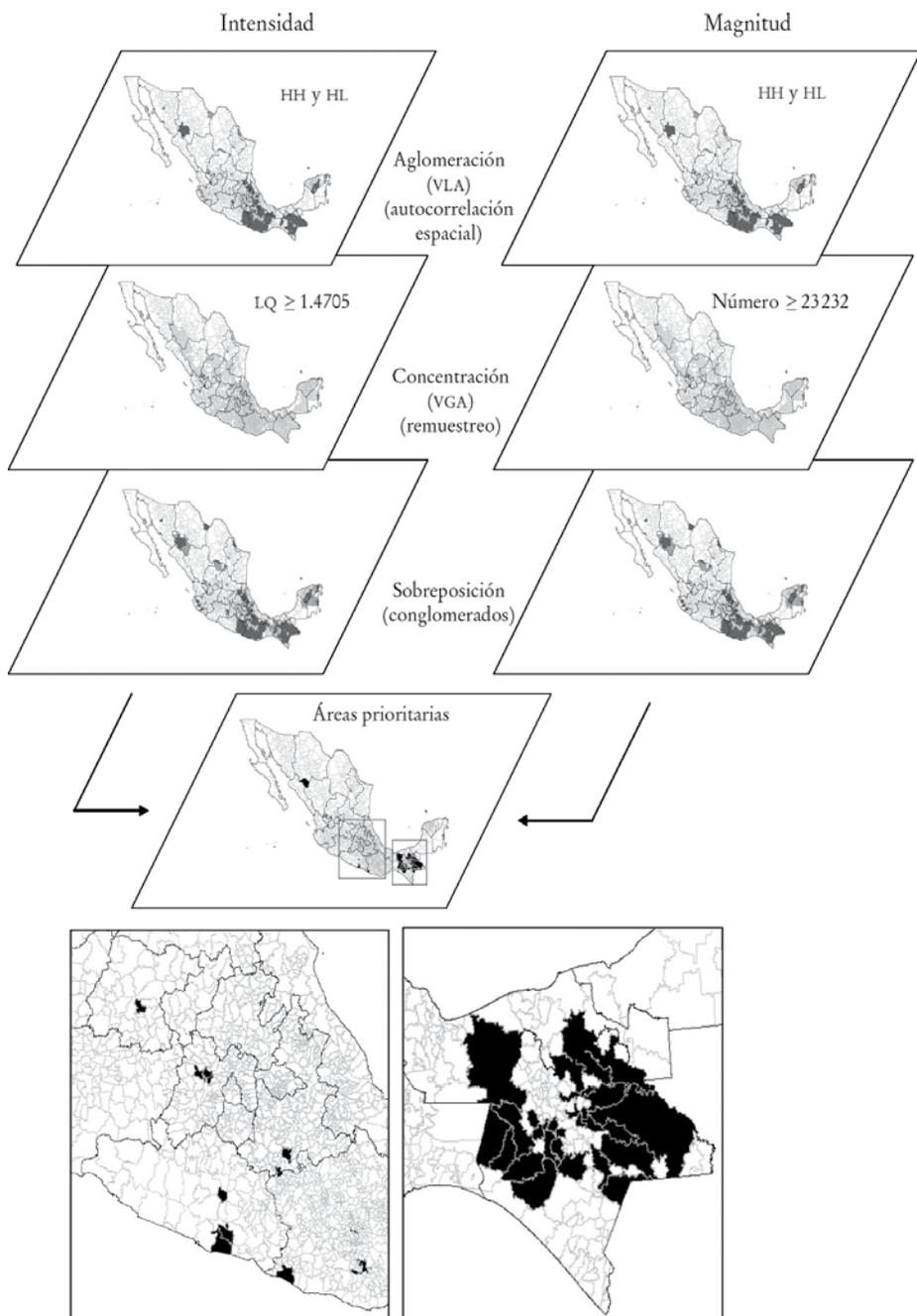
<sup>a</sup> Núcleos (HH y HL en gris oscuro) y periferias (en color gris claro). Los municipios cuadriculados se localizan en el área de intersección de los conglomerados de intensidad y magnitud (prioridad 1, en el cuadro 2). Las áreas fuera de la intersección son estratificadas conforme a los límites de clase en el cuadro 3 (no representados en el mapa). La magnitud se refiere al número de pobres en el 2010 tal como lo reporta el Coneval.

cuadro 2). De estos 44 municipios de mayor prioridad 28 están en Chiapas (65%). Estas 28 áreas contienen 70% (1 874 943) de los pobres localizados en las áreas de mayor prioridad (2 678 769).

*Municipios fuera del área de intersección.* En este trabajo, los conglomerados de intensidad o magnitud, según corresponda, son resultado de la superposición de las capas de concentración y aglomeración. Estos conglomerados *per se* son áreas importantes de pobreza relativa o absoluta. Es razonable asignar la prioridad espacial mayor para fines de política social a la intersección de ambos conglomerados (áreas urgentes o de prioridad uno). Los municipios fuera (a la vez que dentro) de la intersección de los conglomerados de intensidad y magnitud, por ser también importantes en términos relativos o absolutos, son estratificados para sugerir una jerarquía espacial para la toma de decisiones.

*Discontinuidades naturales (Jenks) y puntas y colas.* El uso de estos procedimientos requiere de algunas observaciones previas. En presencia de asimetría y valores atípicos extremos, el procedimiento de discontinuidades naturales, también conocido como método de bondad de ajuste de la va-

DIAGRAMA 8. *Superposición de los conglomerados de intensidad y magnitud*



FUENTE: elaboración del autor.

CUADRO 2. *Lista de las cuarenta y cuatro áreas donde se intersectan la magnitud e intensidad de la pobreza en México*

Huimanguillo	] Tabasco	Chiapas	La Trinitaria
Tacotalpa			Ocosingo
Macuspana			Tuxtla Gutiérrez
			Chiapa de Corzo
Temoaya	] México		Tecpatán
Jiquipilco		San Fernando	
Ixtlahuaca		Ocozacoautla de Espinosa	
			Chilón
Acatlán	] Puebla		Chenalhó
			Comitán de Domínguez
Santa Cruz de Juventino Rosas	] Guanajuato		Altamirano
			Acala
			San Juan Cancuc
San Marcos	] Guerrero		Zinacantán
Tecoanapa		Yajalón	
Tixtla de Guerrero		Villaflores	
			Villa Corzo
			Venustiano Carranza
Las Choapas	] Veracruz		Tenejapa
			Simojovel
			Salto de Agua
Guadalupe y Calvo	] Chihuahua		Palenque
			Oxchuc
			Las Margaritas
Villa de Zaachila	] Oaxaca		Ixtapa
Miahuatlán de Porfirio Díaz		Tila	
Santiago Pinotepa Nacional		Jiquipilas	
			Cintalapa

FUENTE: elaboración del autor.

rianza (GVF, por sus siglas en inglés) o de Jenks (por su autor), demanda modificaciones. Los casos extremos deben ser excluidos temporalmente para aplicar las discontinuidades naturales a datos previamente transformados (*e. g.*, logaritmos, raíces cuadradas o valores inversos). Una vez que la clasificación es obtenida, los valores extremos son reintegrados al estrato menor o mayor, según corresponda. Además, si la distribución es asimétrica, la bondad de ajuste de la desviación absoluta (GADF por sus siglas en inglés) en vez de la GVF debería utilizarse para evaluar la estratificación.

La estadística descriptiva muestra que la distribución de los valores en las áreas fuera de la intersección es asimétrica, lo mismo para los conglomerados

dos de magnitud que intensidad. Las cajas de bigotes (*box-plot*) en *ArcGis* 10.2 y la estandarización  $z$ -MAD registran casos extremos en la distribución de la magnitud. El valor crítico para casos extremos (límite superior) es 241 325 y 191 029 pobres para las cajas de bigotes y los valores  $z$ -MAD, respectivamente. La prueba  $\kappa$ -s rechaza la hipótesis de normalidad a 5%. La sección del conglomerado de magnitud que está fuera de la intersección es asimétrica y presenta casos extremos.

Por otro lado, hay 902 casos en el conglomerado de intensidad que están fuera del área de intersección, con una asimetría de  $-0.290$  y un error estándar de asimetría de  $0.081$ . El valor  $z$  resultante es  $-3.58$ , menor a  $-1.96$ , el valor requerido para ser estadísticamente significativo a 5% en una prueba de dos colas. La prueba  $\kappa$ -s rechaza la hipótesis de normalidad a 5%. Las gráficas de bigotes y los valores  $z$ -MAD no reportan casos extremos. La sección del conglomerado de intensidad que está fuera de la intersección es asimétrica sin casos extremos.

El método de puntas y colas está especialmente diseñado para distribuciones asimétricas. Es un procedimiento robusto para distribuciones estadísticas no normales con casos extremos. El número de estratos en el procedimiento de puntas y colas puede identificarse por una regla empírica o una condición metodológica (Jiang y Yin, 2013: 8). Aplicando la regla empírica de que “las puntas deben contener menos de 40% de los casos”, la magnitud tiene tres estratos. Si se aplica la condición de que la mayoría (aunque no necesariamente todas) de las clases jerárquicas cumple con el principio de “mucho más cosas pequeñas que las más grandes”, hay cinco estratos (cuadro 3). Las siguientes líneas se basan en la clasificación de tres estratos obtenida por la regla de 40% del procedimiento de puntas y colas (cuadro 4). Un número similar de estratos se obtiene con discontinuidades naturales para fines comparativos, incluyendo y excluyendo los casos extremos de la magnitud (cuadro 3).

Como el procedimiento de puntas y colas sugerido por Jiang y Yin (2013) es muy reciente, las discontinuidades naturales resultan ser el procedimiento más común para clasificar (con algunas adaptaciones) distribuciones asimétricas con casos extremos. La adaptación básica del procedimiento sugiere una estandarización de datos para cumplir con el requisito de normalidad e identificar casos extremos potenciales.<sup>9</sup> Después de la estandarización, los

<sup>9</sup> En esta investigación, estandarización no equivale a transformación. La estandarización de datos (valores  $z$  o  $z$ -MAD) reescala la información sin cambiar la forma de la distribución original; la asimetría es la misma y los casos extremos, si existen, también permanecen. La estandarización  $z$ -MAD, basada en

CUADRO 3. Sección del conglomerado de magnitud que está fuera del área de intersección. Clasificación de puntas y colas y discontinuidades naturales<sup>a</sup>

Estrato	Puntas y Colas (regla de 40%)	Discontinuidades Naturales	
		Con casos extremos	Sin casos extremos <sup>b</sup>
		<i>Límites de clase</i>	
1	23 271-101 809	23 271-134 177	23 271-53 987
2	101 810-236 467	134 178-356 328	53 988-103 550
3	236 468-732 154	356 329-732 154	103 551-732 154
GADF	0.5382	0.5506	0.4846

FUENTE: elaboración del autor con el programa *ArcGis 10.2*.

<sup>a</sup> Puntas = Valores  $\geq$  media; Colas = Valores  $<$  media. Cálculos verificados con el programa de Puntas y Colas disponible en <https://www.dropbox.com/sh/k2aaop1y888r45/AAACV2jsvP8sOy8a-5reRgrda>

<sup>b</sup> Los casos extremos son los valores-z modificados ( $z\text{-MAD}$ )  $>$   $|3.5|$ . Estos casos son excluidos temporalmente y reintegrados en el estrato superior para aplicar la estratificación de Jenks. Los límites de clase no cambian si se utilizan los valores absolutos originales o su  $z\text{-MAD}$  equivalente.

casos extremos son temporalmente excluidos para aplicar la estratificación de Jenks y reintegrados posteriormente al estrato mayor o menor, según corresponda. En la opción de discontinuidades naturales para la sección del conglomerado de magnitud que está fuera del área de intersección, la investigación estandariza los datos brutos en valores  $z\text{-MAD}$  para remediar la asimetría y excluye temporalmente los casos extremos. Esta opción, sin embargo, no genera estratos de amplitud similar a los obtenidos con el procedimiento de puntas y colas (cuadro 4).

La evaluación GVF no se aplica para el procedimiento de puntas y colas porque esta última no está basada en la distribución normal de los valores. Los valores GADF, la alternativa robusta a la GVF para evaluar distribuciones asimétricas, son 0.55 y 0.48 para valores con y sin casos extremos, respectivamente. Estos valores GADF sugieren que el número de clases en magnitud identificadas por el método puntas y colas debería incrementarse (hasta alcanzar un GADF de 0.8, al menos) (cuadro 3). Teniendo en mente la superioridad de este método sobre las discontinuidades naturales para clasificar distribuciones asimétricas (Jiang y Yin, 2013), estos resultados muestran que el GADF no parece apropiado para comparar la eficiencia de ambos procedimientos.

la mediana, se recomienda para distribuciones asimétricas porque es menos afectada por valores extremos que los valores  $z$ , basados en la media. En distribuciones asimétricas, la  $z\text{-MAD}$  detecta más casos extremos que los valores  $z$ . Por otro lado, la transformación de los valores, tales como logaritmos, raíz cuadrada, transformación Box-Cox, cambia la forma de la distribución y modifica (o desaparece) la existencia de casos extremos.

CUADRO 4. *Prioridades de política espacial considerando simultáneamente la magnitud e intensidad de la pobreza en los municipios de México, 2010*

	<i>Magnitud (número de pobres)</i>	<i>Intensidad (LQ—pobreza)</i>
Prioridad 1	Intersección	Intersección
Prioridad 2	236 468-732 154 (áreas fuera de la intersección pero dentro del conglomerado)	Todas las áreas fuera de la intersección pero dentro del conglomerado
Prioridad 3	101 810-236 467 (áreas fuera de la intersección pero dentro del conglomerado)	
Prioridad 4	23 271-101 809 (áreas fuera de la intersección pero dentro del conglomerado)	
Prioridad 5	Todos los valores globales altos ( $\geq 23\,232$ pobres) fuera del conglomerado	Todos los valores globales altos ( $LQ \geq 1.4705$ ) fuera del conglomerado
Prioridad 6	Áreas restantes	Áreas restantes

FUENTE: elaboración del autor con base en el procedimiento explicado en el texto.

Por otro lado, bajo la regla de 40% para las puntas, el procedimiento puntas y colas clasifica los valores de intensidad fuera de la intersección en un estrato único. Por lo tanto, en la sección del conglomerado de intensidad que está fuera del área de intersección no se utiliza la clasificación por discontinuidades naturales para fines comparativos. Todos los valores del conglomerado de intensidad fuera del área de intersección tienen prioridad 2 (cuadro 4 y diagrama de Venn en el diagrama 8).<sup>10</sup>

La estratificación de puntas y colas es la mejor opción para distribuciones asimétricas, como en la pobreza. Una vez que el método de las discontinuidades naturales es reemplazado por el procedimiento de puntas y colas, la jerarquía espacial de los conglomerados de magnitud e intensidad

<sup>10</sup> La intensidad fuera de la intersección pero dentro del conglomerado correspondiente, de ser necesario, puede transformarse para obtener una distribución gaussiana que se estratifique con el método de discontinuidades naturales. Este método es el más apropiado para distribuciones simétricas, con el debido manejo de los casos extremos (Galant, 2006). Un criterio, por ejemplo, de política sería dar igual jerarquía espacial al estrato tres de la intensidad (obtenido por Jenks) y al estrato tres de la magnitud (obtenido por puntas y colas). En la opción actual, la intensidad en todas las áreas fuera de la intersección, pero dentro del conglomerado, tienen prioridad 2. Es decir, las prioridades 3 y 4 de magnitud no tienen equivalente en el conglomerado de intensidad.

está lista para orientar la política pública que se interese en los conglomerados de valores altos, sin importar si son globales o locales. La jerarquía espacial de la pobreza sugerida en esta investigación se sintetiza de la siguiente manera:

### 3. Recuento para las hipótesis de trabajo

En ningún caso se acepta la hipótesis nula ( $H_0$ ) que da cauce empírico a cada una de las tres preguntas de investigación.

*Primera hipótesis.*  $H_0$ : La pobreza está distribuida aleatoriamente entre los municipios de México.  $H_1$ : No se acepta  $H_0$ .

La política contra la pobreza espacialmente focalizada tiene sentido puesto que no se acepta  $H_0$ . El remuestreo indica que la pobreza no sólo se concentra en algunos municipios del país sino que su distribución geográfica no es aleatoria, tal como lo muestra el Índice Global de Moran (I). Los argumentos sobre concentración y aglomeración son válidos lo mismo para la intensidad como para la magnitud de la pobreza.

*Segunda hipótesis.*  $H_0$ : No hay aglomeraciones de municipios pobres.  $H_1$ : Hay municipios con altos niveles de pobreza rodeados por municipios igualmente pobres.

No se acepta  $H_0$  porque existen aglomeraciones de municipios pobres. Es necesario focalizar geográficamente la política social. El Índice Local de Moran ( $I_i$ ) no sólo confirma la existencia de focos rojos, sino que al combinarse con VGA identificados por remuestreo evidencia que esas aglomeraciones se extienden a su periferia para formar conglomerados de intensidad o magnitud de pobreza. La combinación de VGA y VLA no sólo muestra que la pobreza no está aleatoriamente distribuida en el espacio, sino que se aglomera en algunas áreas que son más extendidas que lo que sugiere la estadística espacial ( $I_i$ ).

*Tercera hipótesis.*  $H_0$ : No es posible una metodología que articule simultáneamente los valores globales altos y valores locales altos de variables en términos relativos y absolutos.  $H_1$ : No se acepta  $H_0$ .

No se acepta  $H_0$ . La ausencia de una metodología como la descrita sólo es señal de la evolución incompleta del pensamiento espacial que puede mejorar. Esta investigación muestra que es posible elaborar una jerarquía espacial que combine simultáneamente valores absolutos y relativos del proceso de concentración y aglomeración. El método seleccionado depende del objetivo del estudio y las características de los datos. Si el interés es tener una jerarquía espacial de todos los municipios, el método de Jenks o el de puntas y colas pueden ser útiles, con las precauciones estadísticas del caso. Estas precauciones están directamente vinculadas a las características de los datos. Si la distribución de los datos es asimétrica, con casos atípicos, el procedimiento paretiano de puntas y colas es apropiado. Si la distribución es simétrica sin datos atípicos, el procedimiento de Jenks puede ser útil. Si existieran casos atípicos en una distribución simétrica, el procedimiento de Jenks puede aplicarse excluyendo temporalmente los valores extremos. Estos valores se reintegran después al extremo más alto o más bajo, según corresponda.

#### IV. NOTAS FINALES Y LÍNEAS FUTURAS DE INVESTIGACIÓN

Esta investigación presenta un nuevo procedimiento para identificar la jerarquía municipal de la pobreza combinando simultáneamente magnitud e intensidad. El estudio, por el principio de independencia de la población, se enfoca primero en la intensidad. Identifica concentraciones de pobreza y las integra con aglomeraciones (núcleos) para formar conglomerados de intensidad (núcleos y periferias). En este proceso de jerarquización, la concentración se refiere a la localización de los vga en unos municipios de México, independientemente de su localización geográfica. La aglomeración, por un lado, se refiere a localización de vla en municipios contiguos. Por un lado, los vga, identificados por remuestreo, pueden localizarse donde sea; crean conglomerados cuando son contiguos a las aglomeraciones, identificadas con autocorrelación espacial. Por otro lado, considerando que el tamaño importa, los conglomerados de magnitud se obtienen repitiendo los pasos metodológicos para crear los conglomerados de intensidad.

El punto de partida en esta investigación es la diferenciación explícita entre concentración y aglomeración (Arbia, 2001) y los trabajos pioneros (Linge, 1960) y recientes (Van den Heuvel *et al.*, 2012) que combinan valores absolutos y relativos. Sin embargo, para establecer un orden jerárquico

de áreas, este estudio, por primera vez en la literatura sobre análisis espacial, combina simultáneamente concentración y aglomeración de los valores absolutos y relativos de la pobreza utilizando el análisis de superposición de capas. Se utilizan diagramas de Venn para identificar las áreas en la intersección entre los conglomerados de magnitud e intensidad. Las áreas en esta intersección tienen máxima prioridad (prioridad 1). Las áreas fuera de la intersección se clasifican en subconjuntos o estratos de los conglomerados de intensidad y magnitud.

Como es de esperarse en la mayoría de los problemas sociales, el estudio de caso mexicano rechaza la hipótesis nula de aleatoriedad espacial (hipótesis 1) y concluye que la intensidad y la magnitud, cada una por su lado, tienen un patrón espacial concentrado y aglomerado (hipótesis 2). En el área de intersección de estos conglomerados de magnitud e intensidad están 44 municipios mexicanos (prioridad 1). De estos cuarenta y cuatro municipios tres cuartas partes contienen 70% de la pobreza en las áreas de prioridad más alta, y se localiza completamente dentro del estado de Chiapas. Ocosingo, tiene “el valor más alto entre los más altos” de magnitud e intensidad. Atendiendo simultáneamente a los valores absolutos y relativos de la magnitud e intensidad de la pobreza, Ocosingo debería tener la prioridad más alta en la política pública respectiva.

Una vez identificada la intersección entre los dos conglomerados, el estudio sugiere una jerarquía espacial de los municipios restantes del conglomerado. Las áreas de magnitud e intensidad fuera de la intersección son clasificadas con el método de puntas y colas. Como lo demuestra Jiang y Ying (2013), los resultados obtenidos con discontinuidades naturales no se acercan a los obtenidos con el procedimiento de puntas y colas. Considerando la gran asimetría de las distribuciones de magnitud e intensidad, el procedimiento de puntas y colas es la opción más apropiada para clasificar los valores fuera de la intersección de los conglomerados. Este procedimiento genera tres estratos por magnitud y sólo un estrato para la intensidad. Las áreas en la intersección y la estratificación de las que están fuera de ella (dentro y fuera de los conglomerados) proporcionan una jerarquía espacial de la pobreza en México en el 2010 (hipótesis 3: es posible elaborar una metodología que combine simultáneamente VGA y VLA de variables en términos relativos y absolutos).

Existe un número importante de VGA fuera de los conglomerados de magnitud e intensidad que pueden clasificarse con los métodos sugeridos para

los municipios dentro de los conglomerados, dependiendo de la simetría o asimetría de su distribución. Esta clasificación adicional depende de las necesidades que se manifiesten. No tiene caso multiplicar innecesariamente una clasificación si no existe una necesidad técnica o social que la justifique.

Las tres contribuciones principales del estudio son:

- i)* Integra un cuerpo conceptual comparativo para identificar áreas de pobreza con una jerarquía espacial. Este marco es relevante para definir las políticas enfocadas espacialmente, y puede ser insumo para modelos explicativos posteriores (mínimos cuadrados ordinarios, regresión espacial y regresión geográficamente ponderada). En el nivel estadístico, la mera existencia de aglomeraciones y conglomerados es evidencia de la heterogeneidad espacial (regímenes espaciales) que debe tomarse en cuenta en los modelos de regresión espacial y no espacial.
- ii)* Articula conceptual y empíricamente dos procesos espaciales (concentración y aglomeración), dos técnicas estadísticas complementarias (remuestreo y autocorrelación espacial), un procedimiento de sistemas de información geográfica (análisis de superposición de capas), un procedimiento reciente de clasificación (puntas y colas) y dos tipos diferentes de datos (absolutos y relativos) para elaborar una jerarquía espacial de la pobreza. Este procedimiento también puede aplicarse a cada una de las variables que definen la pobreza, tales como personas en carencia de cierto número de necesidades básicas o por debajo de cierto nivel de ingreso.
- iii)* Retoma conceptos básicos (*i. e.*, patrón espacial, concentración, aglomeración) y repasa medidas de intensidad y magnitud que pueden utilizarse para identificar y explicar diversos patrones geográficos, tales como pobreza, desarrollo humano, mortalidad infantil, fertilidad adolescente, enfermedad o crimen. Los resultados también deben abordarse a la luz de la experiencia en otros estudios de caso.

Se espera que la metodología en esta investigación ayude a crear una jerarquía espacial útil para focalizar territorialmente los programas sociales. Las dos variables en esta investigación (intensidad de la pobreza y número de pobres) pueden ser sustituidas por las consideradas en los programas Sedesol, tales como la Cruzada contra el Hambre (gente en pobreza extrema o en pobreza alimentaria) o Prospera. La investigación también presenta

criterios objetivos que no requieren familiaridad con el estudio de caso y proporciona valores críticos estadísticamente respaldados.

Como en cualquier otro método, la metodología espacial sugerida tiene sus propias limitaciones:

- i)* Depende de la desagregación y exactitud de la unidad espacial de análisis (e. g., municipios, AGEB, manzanas).
- ii)* Es posible que los resultados no sean los mismos para datos suavizados en el tiempo (e. g., promedios móviles) y el espacio (e. g., suavización espacial bayesiana). Además de utilizar tasas espaciales suavizadas en esta investigación, de estar disponibles, los estudios también pueden utilizar la media estadística para un periodo de tiempo (i. e., la tasa promedio de la pobreza para el 2003-2012) o los promedios móviles en el tiempo. Desafortunadamente, la base de datos espacialmente desagregada para la pobreza en el estudio de caso mexicano sólo existe para el 2010.
- iii)* Deja sin los beneficios de los programas sociales a personas en la pobreza que no se localizan en las áreas seleccionadas. Son necesarios los ajustes regionales basados en la historia, cultura y economía comunes. Las regiones y sus municipios generalmente comparten características naturales similares, factores sociales y experiencias comunes. Por ejemplo, 23 municipios integran la región tarahumara, pero sólo algunos son incluidos en el conglomerado de intensidad de la pobreza tarahumara. Por lo tanto, es necesario visitar la región y verificar lo que pasa ahí para ver si la inclusión/exclusión de los municipios tiene sustento local. Esta tarea está fuera del alcance del estudio de caso y queda como una asignatura pendiente para estudios subsecuentes.

Los resultados principales de esta investigación no son comparables o compatibles con los reportados en los documentos de organismos oficiales tales como Coneval o Sedesol. En el caso de Coneval:<sup>11</sup>

<sup>11</sup> La mención crítica a la estratificación del Coneval no es exclusiva para este organismo, sino que se aplica a cualquier procedimiento de jerarquización espacial elaborada con procedimientos espaciales que, centrados exclusivamente en la intensidad, omiten por regla general los valores absolutos y el proceso de aglomeración. Si las unidades espaciales (estados, municipios o AGEB) fueran mezclados como las fichas de dominó o barajadas como las cartas de un mazo de naipes, los estratos generados por cualquier método espacial serían los mismos.

- i) Mientras esta investigación incluye simultáneamente magnitud (número absoluto de pobres) e intensidad (LQ) en su clasificación, el Coneval no incluye magnitud. Sólo utiliza incidencia (porcentaje de pobres) para estratificar los municipios mexicanos.
- ii) A diferencia de esta investigación, el Coneval no considera la aglomeración en su clasificación.

Por otro lado, aunque la selección municipal de la Cruzada contra el Hambre en Sedesol incluye criterios de magnitud e intensidad, la política social federal utiliza la pobreza alimentaria, una variable distinta a la utilizada en este trabajo. La pobreza alimentaria es un subconjunto de la pobreza multidimensional generada por el Coneval que se utiliza en esta investigación.

En general, el estudio sugiere que las investigaciones realizadas hasta hoy sobre patrones espaciales utilizando sólo estadística espacial o únicamente estadística no espacial sean reconsideradas con el procedimiento propuesto. La misma invitación se extiende a las investigaciones futuras sobre el tema.

#### REFERENCIAS BIBLIOGRÁFICAS

- Anselin, L., A. T. Murray y S. J. Rey (2013), "Spatial Analysis", en T. D. Little (ed.), *The Oxford Handbook of Quantitative Methods: Vol. 2*, Oxford University Press, Oxford, Nueva York.
- , Y. W. Kim y I. Syabri (2004), "Web-based Analytical Tools for the Exploration of Spatial Data", *Journal of Geographical Systems*, vol. 6, núm. 2, pp. 197-218.
- (1995), "Local Indicators of Spatial Association-LISA", *Geographical Analysis*, vol. 27, núm. 2, pp. 93-115.
- (1996), "The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association", en M. M. Fischer, H. J. Scholten y D. Unwin (eds.), *Spatial Analytical Perspectives on GIS*, Taylor & Francis., Londres.
- (2003), "Spatial Econometrics: Methods and Models", en B. H. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Blackwell Publishing, Reino Unido, pp. 310-330.
- Arbia, G. (2001), "The Role of Spatial Effects in the Empirical Analysis of Regional Concentration", *Journal of Geographical Systems*, vol. 3, núm. 3, pp. 271-281.
- Arribas-Bel, D., y F. Sanz-Gracia (2014), "The Validity of the Monocentric City Model in a Polycentric Age: US Metropolitan Areas in 1990, 2000 and 2010", *Urban Geography*, vol. 35, núm. 7, pp. 980-997.
- Baumont, C., C. Ertur y J. L. Gallo (2004), "Spatial Analysis of Employment and

- Population Density: The Case of the Agglomeration of Dijon 1999”, *Geographical Analysis*, vol. 36, núm. 2, pp. 146-176.
- Bedi, T., A. Coudouel y K. Simler (2007), *More than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*, Banco Mundial, Washington D. C.
- Bourdieu, P. (2000), “Site Effects”, en P. Bourdieu, A. Accardo y P. Parkhurst Ferguson (eds.), *The Weight of the World: Social Suffering in Contemporary Society*, Stanford University Press, California.
- Burt, J. E., G. M. Barber y D. L. Rigby (2009), *Elementary Statistics for Geographers* (3ª ed.), Guilford Press, New York.
- Carroll, M. C., N. Reid y B. W. Smith (2008), “Location Quotients versus Spatial Autocorrelation in Identifying Potential Cluster Regions”, *Annals of Regional Science*, vol. 42, núm. 2, pp. 449-463.
- Coneval (2014), “Medición de la Pobreza. Pobreza a nivel municipio 2010”, disponible en <http://www.coneval.gob.mx/Medicion/Paginas/Medici%C3%B3n/Medicion-de-la-pobreza-municipal-2010.aspx>
- Crews, K. A., y M. F. Peralvo (2008), “Segregation and Fragmentation: Extending Landscape Ecology and Pattern Metrics Analysis to Spatial Demography”, *Population Research and Policy Review*, vol. 27, núm. 1, pp. 65-88.
- Cromley, R., y D. Hanink (2012), “Focal Location Quotients: Specification and Applications”, *Geographical Analysis*, vol. 44, núm. 4, pp. 398-410.
- Chainey, S., y J. Desyllas (2008), “Modelling Pedestrian Movement to Measure On-street Crime Risk”, en L. Liu y J. Eck (eds.), *Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems*, Information Science Reference, Pensilvania.
- Chihara, L., y T. Hesterberg (2011), *Mathematical Statistics with Resampling and R*, Wiley, Nueva Jersey.
- Choudhury, N. Y., y C. Räder (2014), “Targeting the Poorest and Most Vulnerable: Examples from Bangladesh”, en J. Von Braun y F. W. Gatzweiler (eds.), *Marginality: Addressing the Nexus of Poverty, Exclusion and Ecology*, Springer Netherlands, Dordrecht.
- De Dominicis, L., G. Arbia y H. L. De Groot (2013), “Concentration of Manufacturing and Service Sector Activities in Italy: Accounting for Spatial Dependence and Firm Size Distribution”, *Regional Studies*, vol. 47, núm. 3, pp. 405-418.
- Frizado, J., B. W. Smith, M. C. Carroll y N. Reid (2008), “Impact of Polygon Geometry on the Identification of Economic Clusters”, *Letters in Spatial and Resource Sciences*, vol. 2, núm. 1, pp. 31-44.
- , B. Smith y M. Carroll (2007), “Identification of Economic Clusters Using ArcGIS Spatial Statistics”, documento presentado en la Proceedings of the ESRI User Conference, disponible en [https://s3.amazonaws.com/webapps.esri.com/esri-proceedings/proc07/papers/papers/pap\\_1367.pdf](https://s3.amazonaws.com/webapps.esri.com/esri-proceedings/proc07/papers/papers/pap_1367.pdf)

- Galant, K. (2006), "Data Classification from Cartographic Point of View", *Electronic Journal of Polish Agricultural Universities*, vol. 9, núm. 4.
- Goodchild, M. F., y N. S. Lam (1980), "Areal Interpolation: A Variant of the Traditional Spatial Problem", *Geo-Processing*, 1, pp. 297-312.
- Gottdiener, M., y R. Hutchison (2011), *The New Urban Sociology*, Westview Press, Colorado.
- Graw, V., y C. Husmann (2014), "Mapping Marginality Hotspots", en J. Von Braun y F. W. Gatzweiler (eds.), *Marginality: Addressing the Nexus of Poverty, Exclusion and Ecology*, Springer Netherlands, Dordrecht.
- Hesterberg, T., D. S. Moore, S. Monaghan, A. Clipson, R. Epstein, B. A. Craig y G. P. McCabe (2010), "Bootstrap Methods and Permutation Tests", en D. S. Moore, G. P. McCabe y B. A. Craig (eds.), *Introduction to the Practice of Statistics*, 7ª ed., W. H. Freeman, Nueva York.
- Holt, J. B. (2007), "The Topography of Poverty in the United States: A Spatial Analysis Using County-Level Data from the Community Health Status Indicators Project", *Preventing Chronic Disease*, vol. 4, núm. 4, pp. 1-9.
- Jiang, B. (2013), "Head/Tail Breaks: A New Classification Scheme for Data with a Heavy-Tailed Distribution", *The Professional Geographer*, vol. 65, núm. 3, pp. 482-494.
- , y J. Yin (2014), "Ht-index for Quantifying the Fractal or Scaling Structure of Geographic Features", *Annals of the Association of American Geographers*, vol. 104, núm. 3, pp. 530-541.
- (2015), "Geospatial Analysis Requires a Different Way of Thinking: The Problem of Spatial Heterogeneity", *GeoJournal*, vol. 80, núm. 1, pp 1-13.
- Ketels, C., y Ö. Sölvell (2005), "Clusters in the EU 10 New Member Countries", Europe INNOVA, disponible en <http://www.europe-innova.eu>
- Linge, G. J. (1960), "The Concentration and Dispersion of Manufacturing in New Zealand", *Economic Geography*, vol. 36, núm. 4, pp. 326-343.
- Liu, Z. (2013), "Global and Local: Measuring Geographical Concentration of China's Manufacturing Industries", *The Professional Geographer*, vol. 66, núm. 2, pp. 284-297.
- McMillen, D. P. (2001), "Polycentric Urban Structure: The Case of Milwaukee", *Economic Perspectives*, vol. 25, núm. 2, pp. 15-27.
- North, B., D. Curtis y P. Sham (2002), "A Note on the Calculation of Empirical P-Values from Monte Carlo Procedures", *The American Journal of Human Genetics*, vol. 71, núm. 2, pp. 439-441.
- O'Donoghue, D., y B. Gleave (2004), "A Note on Methods for Measuring Industrial Agglomeration", *Regional Studies*, vol. 38, núm. 4, pp. 419-427.
- O'Sullivan, D., y G. L. Perry (2013), *Spatial Simulation: Exploring Pattern and Process*, Wiley Blackwell, New Jersey.
- Partridge, M. D., y D. S. Rickman (2006), *The Geography of American Poverty: Is*

- There a Need for Place-Based Policies?*, W. E. Upjohn Institute for Employment Research, Michigan.
- Pavlova, S. *et al.* (2014), “Geographical Pattern of STDs in Texas Counties”, en J. A. Treviño (ed.), *Geography of Three Social Issues: STDs, Energy Consumption, & TOD Impacts*, UANL/UTA/Medical University/Varna, Nuevo León.
- Riguelle, F., I. Thomas y A. Verhetsel (2007), “Measuring Urban Polycentrism: A European Case Study and its Implications”, *Journal of Economic Geography*, vol. 7, núm. 2, pp. 193-215.
- Sedesol (2013), “Procedimiento de selección de los 400 municipios prioritarios de la Cruzada Nacional Contra el Hambre”, disponible en [http://www.sedesol.gob.mx/work/models/SEDESOL/Cruzada/4\\_NOTA\\_TECNICA\\_DEL\\_PROCEDIMIENTO\\_DE\\_SELECCION\\_DE\\_LOS\\_400\\_MUNICIPIOS\\_PRIORITARIOS\\_DE\\_LA\\_CNCH.pdf](http://www.sedesol.gob.mx/work/models/SEDESOL/Cruzada/4_NOTA_TECNICA_DEL_PROCEDIMIENTO_DE_SELECCION_DE_LOS_400_MUNICIPIOS_PRIORITARIOS_DE_LA_CNCH.pdf)
- Soja, E. W. (2010), *Seeking Spatial Justice*, University of Minnesota Press, Minnesota.
- Srinivasan, S. (2008), “Local and Global Spatial Statistics”, en S. Shekhar y H. Xiong (eds.), *Encyclopedia of GIS*, Springer, Nueva York.
- Tian, Z. (2013), “Measuring Agglomeration Using the Standardized Location Quotient with a Bootstrap Method”, *Journal of Regional Analysis and Policy*, vol. 43, núm. 2, pp. 186-197.
- Van den Heuvel, F. P., P. W. De Langen, K. H. Van Donselaar y J. C. Fransoo (2012), “Identification of Employment Concentration Areas”, *European Planning Studies*, vol. 22, núm. 1, pp. 204-226.
- Visvalingam, M. (1983), “Area-Based Social Indicators: Signed Chi-Square as an Alternative to Ratios”, *Social Indicators Research*, vol. 13, núm. 3, pp. 311-329.
- Yeager, S. J. (2008), “Where Do Research Questions Come from and How Are They Developed?”, en G. J. Miller y K. Yang (eds.), *Handbook of Research Methods in Public Administration*, 2a ed., CRC Press, Florida
- Zhang, C. (2008), “An Analysis of Urban Spatial Structure Using Comprehensive Prominence of Irregular Areas”, *International Journal of Geographical Information Science*, vol. 22, núm. 6, pp. 675-686.
- , L. Luo, W. Xu y V. Ledwith (2008), “Use of Local Moran’s I and GIS to Identify Pollution Hotspots of Pb in Urban Soils of Galway, Ireland”, *Science of The Total Environment*, vol. 398, núm. 1, pp. 212-221.
- , Y. Tang, L. Luo y W. Xu (2009), “Outlier Identification and Visualization for Pb Concentrations in Urban Soils and its Implications for Identification of Potential Contaminated Land”, *Environmental Pollution*, vol. 157, núm. 11, pp. 3083-3090.