

ANÁLISIS DEL CONSUMO ELÉCTRICO RESIDENCIAL EN UNA REGIÓN VENEZOLANA MEDIANTE UN MODELO DE REGRESIÓN GEOGRÁFICAMENTE PONDERADA

CLAUDIO M. ROCCO S.

Escuela Básica de Ingeniería, Universidad Central de Venezuela

e-mail: croccouv@gmail.com

Recibido: agosto 2012

Recibido en forma final revisado: octubre 2013

RESUMEN

El presente trabajo muestra la aplicación de técnicas del Análisis Exploratorio de Datos Espaciales (AEDE) y de modelos de Regresión Geográficamente Ponderada (RGP) para cuantificar el efecto de variables georeferenciadas, asociadas a la población, en el consumo eléctrico residencial, para una región de servicio venezolana, conformada por varias zonas. El AEDE es una disciplina de la estadística empleada para el estudio de datos espaciales y permite identificar la presencia de efectos espaciales de autocorrelación (AE) y heteroscedasticidad (HE), que invalidan el uso del Modelo Clásico de Regresión Lineal (MCRL). Por otra parte, la RGP permite cuantificar, de forma eficiente, las relaciones funcionales en cada zona, al modelar adecuadamente los efectos espaciales. El análisis de la región de estudio permitió detectar la existencia de efectos espaciales de la AE y la HE. El modelo basado en RGP mejora el ajuste ofrecido por el MCRL en términos de la capacidad explicativa y permite cuantificar los diferentes efectos de las variables independientes en cada zona.

Palabras clave: Datos espaciales, Consumo eléctrico, Índice de Moran, Mapas LISA, Regresión geográficamente ponderada.

ANALYSIS OF THE RESIDENTIAL ELECTRIC CONSUMPTION IN A VENEZUELAN REGION USING A GEOGRAPHICALLY WEIGHTED REGRESSION APPROACH

ABSTRACT

In this paper, Spatial Data Analysis (SDA) and Geographically Weighted Regression (GWR) models are used for quantifying the effects of geo-referenced variables related to the residential electric consumption in a Venezuelan region with several zones. SDA is a statistical technique used to analyze spatial data and able to detect spatial autocorrelation (SA) and spatial heteroscedasticity (SH) effects, which invalidate classical regression model (CRM). GWR allows the efficient quantification of functional relationship in every zone, because it is able to correctly model the spatial effects. The analysis for the selected region suggests the presence of SA and SH. The regression model based on GWR is better than the CRM since it has a higher explanatory capability and allows to quantifying the different effects of the independent variables in each zone.

Keywords: Electrical consumption, Geographically weighted Regression, LISA maps, Moran Index, Spatial data.

INTRODUCCIÓN

El crecimiento planificado de un sistema eléctrico garantiza al suscriptor una calidad y continuidad del servicio adecuado y a la empresa responsable, la mayor utilización de sus recursos, manteniendo una imagen eficiente ante sus usuarios. Es por ello que la definición de modelos matemáticos adecuados, que permitan conocer la relación entre consumo y los factores más importantes que lo afectan, juega un papel preponderante en la planificación de estos sistemas.

En los modelos clásicos de regresión, ampliamente usados para la estimación de la demanda eléctrica, se supone que la relación funcional entre la variable dependiente (esto es, el consumo eléctrico) y las variables independientes (población, empleo, entre otras) es constante en toda la región de estudio.

Sin embargo, un aspecto importante que hay que considerar en este tipo de modelos, está relacionado con el hecho de que una región, conformada por varias zonas, puede presentar diferencias de orden geofísico, cultural, social

y económico. Por ejemplo, Caracas es una ciudad que se encuentra dividida en cinco municipios (Baruta, Chacao, El Hatillo, Libertador y Sucre) muy diferentes en cuanto a tamaño, número de pobladores, nivel de ingreso per cápita, actividades económicas desarrolladas, entre otras variables.

Esta situación puede provocar que el consumo eléctrico (por ejemplo, residencial) se distribuya de manera heterogénea sobre el espacio. En otras palabras, las variables que pueden explicar el comportamiento del consumo eléctrico pueden tener coeficientes variables. Esta característica es contraria a la característica básica de los modelos clásicos de regresión, en los que se definen coeficientes “globales” para cada variable independiente.

El análisis clásico de datos estadísticos en el sector eléctrico, está asociado generalmente con series de tiempo de diferentes variables, tales como consumo de energía eléctrica, población por estratos, tipos de empleo, entre otras. Sin embargo, la realidad evidencia que gran parte de esos datos también tienen asociados características geográficas.

Por ejemplo, se puede contar con datos de consumo para cada cliente (esto es, se conoce la ubicación del suscriptor) o datos de población por municipio o segmento censal. Tradicionalmente se ha usado un conjunto de técnicas de representación y análisis de datos más enfocadas a las series de tiempo, mientras que el análisis de datos espaciales ha sido utilizado muy poco.

El AEDE es una disciplina de la estadística empleada para el estudio de datos espaciales (datos de una variable y su ubicación) y permite identificar la presencia de peculiaridades que no siempre pueden ser modeladas con técnicas econométricas tradicionales. Estas singularidades pueden resumirse en los llamados efectos de los datos espaciales: Autocorrelación Espacial (AE) y Heterogeneidad Espacial (HE) (LeSage, 1999; Chasco, 2007a; Chasco, 2003a).

La AE es la manifestación de una relación funcional entre lo que ocurre en un lugar geográfico y otro (Astorkiza, 2006; Moreno, 2002). Este efecto espacial puede manifestarse de forma positiva o negativamente; en el primer caso se observa que la presencia de un fenómeno en un lugar está acompañada con la presencia del mismo en las zonas vecinas, mientras que el segundo caso sucede lo contrario. El hecho de que una zona pueda verse afectada simultáneamente por muchas otras que la rodeen, determina el que la AE sea un fenómeno multidireccional y que por ende no pueda modelarse utilizando técnicas propias de la econometría

tradicional (LeSage, 1999; Moreno, 2002; Sánchez, 2006).

Por su parte, la HE es esencialmente la variabilidad en el espacio de las relaciones entre las variables y surge por el hecho de que cada zona estudiada tiene características propias, lo que pudiera provocar que se intentara explicar un mismo fenómeno empleando datos de unidades espaciales muy distintas (Moreno, 2002). Desde el punto de vista de los modelos de regresión clásicos, la HE puede manifestarse en forma de heteroscedasticidad (varianza no constante en el término de perturbación) y/o como inestabilidad estructural del modelo (Moreno, 2002), lo que provocaría violaciones a los supuestos del modelo clásico de regresión (LeSage, 1999; Gómez, 1999; Anselin, 1999; Chasco, 2003b).

En muchas situaciones estos efectos son tratados con técnicas propias de la econometría espacial, tales como los modelos de Retardo Espacial (correlación en las variables explicativas) o el modelo de Error Espacial (correlación en el término de error) (LeSage, 1999; Anselin, 1999; Moreno, 2000).

Otra opción que se ha utilizado recientemente, está basada en la técnica de Regresión Geográficamente Ponderada (Geographically Weighted Regression (GWR)) (Fotheringham et al. 2002), definida como una familia de modelos de regresión en los cuales los coeficientes β , asociados con cada una de las variables independientes, pueden variar en forma espacial, permitiendo representar contribuciones distintas en zonas distintas.

Este artículo analiza el consumo eléctrico residencial de una región de servicio en Venezuela (formada por varias zonas, correspondientes a segmentos censales o agregaciones de segmentos), mediante técnicas propias del AEDE para detectar la presencia de efectos espaciales. Posteriormente se comparan los resultados del enfoque tradicional de un modelo de regresión global con los resultados de un modelo GWR, resaltando las bondades de este último, en términos de la capacidad explicativa y la posibilidad de cuantificar diferencias zonales.

En la literatura existen trabajos de AEDE, Econometría espacial y GWR aplicadas al sector eléctrico, en particular al estudio de relaciones espaciales entre ingreso y consumo o pérdidas eléctricas y variables socio-económicas (Francisco, 2006; Francisco & Fagundes, 2006; Francisco et al. 2010). Sin embargo, a nuestro mejor entender, las técnicas de AEDE y GWR no han sido utilizadas en el sector eléctrico venezolano.

Una metodología relacionada con este trabajo, y

específicamente con la estimación de la demanda eléctrica, es el enfoque denominado “Spatial Load Forecasting” (SLF), propuesto desde los años 80 por Willis (2002) y recientemente documentada (Monteiro et al. 2001; Monteiro & Ramiro-Rosado, 2001; Monteiro, 2003, entre otros). Si bien la metodología SFL utiliza el término “espacial”, su enfoque, por lo general, no hace uso de modelos basados en regresiones ponderadas o en lo que se ha denominado econometría espacial. Por otra parte, el análisis espacial utilizado en SLF no utiliza los conceptos y técnicas particulares que definen el AEDE.

El resto del artículo está dividido de la siguiente manera. La sección AEDE, resume las principales técnicas que lo componen y presenta los posibles efectos espaciales que pueden ser detectados. En la sección ESPECIFICACIÓN DE MODELOS, se expone cómo los efectos espaciales afectan a los modelos clásicos de regresión lineal y se describe el modelo GWR. La sección CASO DE ESTUDIO muestra la aplicación del AEDE y GWR en el análisis de la demanda de energía eléctrica de una región atendida por la empresa estatal de energía CORPOELEC. Finalmente, la última sección presenta las conclusiones del estudio.

ANÁLISIS EXPLORATORIO DE DATOS ESPACIALES (AEDE)

Conjunto parcialmente ordenado

El Análisis Exploratorio de Datos Espaciales (AEDE) es una disciplina de la estadística empleada para el estudio de datos espaciales que permite identificar patrones en el comportamiento de variables geográficas (Chasco, 2003a). Para ello, se vale de un conjunto de técnicas que hace posible la identificación de zonas atípicas, esquemas de asociación espacial y formas de inestabilidad, a fin de lograr una completa descripción de los datos distribuidos geográficamente (Chasco, 2003a, 2005).

El AEDE combina el análisis estadístico y el gráfico, lo que permite tener una apreciación científica y al mismo tiempo visual de lo que está ocurriendo con los datos estudiados. La evolución de la variable en las distintas unidades geográficas o la tendencia espacial es representada mediante un mapa. En la Tabla 1 se presentan algunas técnicas del AEDE.

Datos Espaciales (LeSage, 1999)

Un dato espacial es la observación de una variable asociada con una localización del espacio geográfico y que, además de caracterizarse por su naturaleza georeferenciada, multidireccional y multidimensional, puede ser representada

en un mapa.

Efectos espaciales

Trabajar con datos espaciales significa que nos enfrentaremos a una serie de situaciones ausentes en la econometría tradicional, las cuales se pueden resumir en los llamados efectos espaciales: Autocorrelación Espacial (AE) y Heterogeneidad Espacial (HE).

Autocorrelación espacial

Anselin (1999: 23) define la AE como: “la coincidencia de valores similares en localidades similares”. En otras palabras, el valor de una variable en un lugar del espacio depende de su valor en otro u otros lugares.

De cierta forma, la falta de independencia entre las observaciones que describe la AE está relacionada con la primera ley de la geografía formulada por Tobler en 1979, la cual indica que todo está relacionado con todo, pero las cosas cercanas están más relacionadas entre sí que las cosas lejanas (LeSage, 1999).

Este efecto espacial puede manifestarse de forma positiva o negativa y puede clasificarse de dos maneras: sustantiva o residual. La AE sustantiva se presenta cuando el valor de una variable en una zona realmente depende del valor que tome la misma variable en unidades geográficas vecinas.

Al ignorar esta clase de AE, se introducirían sesgos en las estimaciones con mínimos cuadrados ordinarios, ya que se estaría obviando información importante acerca del comportamiento de las variables. Por otro lado, la dependencia espacial pudiera ser más bien residual, una especie de ruido dentro del análisis que debe incluirse en el término de error.

El omitir este tipo de asociación espacial viola uno de los supuestos del modelo de regresión clásico y por ende los modelos desarrollados pueden ser de poca utilidad. La Econometría Espacial presenta un tratamiento para cada uno de estos tipos de AE a la hora de especificar algún modelo econométrico (LeSage, 1999; Gómez, 1999; Anselin, 1999; Moreno, 2000).

Heterogeneidad Espacial

La HE es esencialmente la variabilidad en el espacio de las relaciones entre las variables y surge por el hecho de que cada zona estudiada tiene características propias, lo que pudiera provocar que se intentara explicar un mismo

fenómeno empleando datos de unidades espaciales muy distintas (Moreno, 2002).

Desde el punto de vista de los modelos de regresión, la HE puede manifestarse en forma de heteroscedasticidad y/o como inestabilidad estructural del modelo (también conocida como no estacionariedad) (Moreno, 2002).

Esta definición puede expresarse matemáticamente de la siguiente forma:

$$y_i = f_i(\beta_i X_i + \varepsilon_i) \quad (1)$$

donde: i denota las observaciones asociadas con cada zona de la región ($i=1, \dots, N$), y_i es la variable endógena en la región i , f_i señala la posibilidad de que exista entre la variable dependiente y las independientes una relación de correspondencia diferente por cada ubicación geográfica, X_i representa el vector de k variables explicativas asociadas al vector de parámetros β_i y ε_i indica el término de perturbación estocástica.

En términos de la expresión (E.1), la inestabilidad estructural del modelo, o no estacionariedad, se origina por los diferentes valores que pueden tomar, para cada punto en el espacio, f_i y/o β_i . Esto abre la posibilidad de obtener diferentes respuestas en diferentes puntos del espacio, a partir de un mismo estímulo (Chasco, 2007a).

La heteroscedasticidad (varianza no constante en los términos estocásticos) ocurre por lo general como consecuencia de la omisión de variables y otros errores a la hora de especificar el modelo. Estos mismos factores pueden también causar la inestabilidad del comportamiento de la variable de estudio en el espacio, provocando que los parámetros estimados (o incluso la función misma) tampoco sean homogéneos en toda la región (Moreno, 2002).

Sin embargo, la Ecuación (1), muestra que ε_i puede ser diferente en cada zona. Esto provoca la posible existencia de heteroscedasticidad, toda vez que la varianza del término de perturbación estocástica puede no ser constante.

En general, la detección de la HE permite obtener modelos de consumo eléctrico más reales ya que se logra introducir, de forma explícita, las diferencias zonales en el análisis.

La consideración expresa de la HE puede poner en evidencia la presencia de la Paradoja de Simpson (Fotheringham et al. 2002), la cual teoriza sobre cambios en la asociación entre dos variables cuando los datos son desagregados por grupos. Estos cambios pueden ir desde modificaciones en

el sentido de las asociaciones hasta la desaparición de las mismas.

En el caso de los datos espaciales, la paradoja se puede manifestar tal como lo ilustra la Figura 1 (Fotheringham et al. 2002). La Figura presenta la relación existente entre dos variables cualesquiera. La Figura 1a) no considera la ubicación geográfica de cada observación, mientras que la Figura 1b) supone que los datos están localizados en dos zonas diferentes.

Es interesante notar que el signo de la posible relación (pendiente de la recta de regresión) cambia, al considerar la información geográfica.

Formulación matemática de la dependencia espacial: matriz de pesos espaciales

La manera más común de relacionar elementos espaciales es utilizando la matriz de pesos espaciales W , en la cual w_{ij} representa la intensidad de la interdependencia entre cada par de observaciones i y j .

La matriz es cuadrada, simétrica, no estocástica y debe tener todos sus elementos finitos y no negativos. La forma más sencilla para definir esa intensidad, es asignar el valor $w_{ij} = 1$ si las unidades son físicamente adyacentes y $w_{ij} = 0$ si no lo son o son elementos de la diagonal principal (LeSage, 1999; Anselin, 1999). Para definir las unidades adyacentes a la unidad i se pueden sugerir los cuatro criterios presentados en la Figura 2.

La detección de la autocorrelación espacial en el análisis exploratorio

Para la detección de autocorrelación espacial en el análisis exploratorio, el estadístico más utilizado es la I de Moran (Anselin, 1999).

La I de Moran se expresa formalmente como:

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

donde:

- N : el número de observaciones,
- w_{ij} : el elemento de la matriz de ordenación espacial correspondiente al par de observaciones i y j ,
- x_i y x_j : las observaciones para las localizaciones i y j ,
- \bar{x} : media muestral de las observaciones
- S_0 : constante de escala $S_0 = \sum_i \sum_j w_{ij}$

Cuando N es suficientemente grande, la I de Moran estandarizada sigue una distribución asintótica normal:

$$Z(I) = \frac{I - E(I)}{[V(I)]^{1/2}} \approx N(0,1) \quad (3)$$

donde:

$E(I)$: la esperanza de I

$V(I)$: la varianza de I

Para este caso, un valor no significativo de $Z(I)$ lleva a no rechazar la hipótesis nula de no autocorrelación espacial, mientras que un valor significativo positivo (o negativo) informará sobre la presencia de autocorrelación espacial positiva (o negativa).

Métodos gráficos de representación de la dependencia espacial

Diagrama de dispersión de Moran

El diagrama de dispersión de Moran es un gráfico que representa en el eje horizontal, la variable a estudiar, previamente estandarizada, y en el eje vertical representa el retardo espacial de esa misma variable estandarizada (Figura 3). El retardo espacial representa el promedio ponderado de los valores que adopta una variable en el subconjunto de observaciones cercanas a una dada.

En este tipo de diagrama, la pendiente de la recta de regresión representa el valor denominado Estadístico I de Moran de autocorrelación espacial global: cuanto mayor sea ese valor, más fuerte es el grado de autocorrelación espacial en la variable y viceversa (LeSage, 1999; Anselin, 1999).

Este diagrama de dispersión divide el tipo de asociación espacial en cuatro categorías, con sus respectivos

cuadrantes: dos para autocorrelación espacial positiva (cuadrantes I y III); y dos para autocorrelación espacial negativa (cuadrantes II y IV).

En el diagrama dispersión de Moran bivalente se representa en el eje vertical, el retardo espacial de la variable que se desea explicar y en el eje horizontal, la variable explicativa (tal como el diagrama de dispersión de Moran univalente).

La pendiente de la línea de regresión indica el grado de relación lineal existente entre la variable del eje horizontal y los valores de la variable del eje vertical en los puntos cercanos a uno dado.

Mapa LISA

Las siglas LISA se refieren a “Local Indicator of Spatial Association” (Indicador Local de Asociación Espacial) (Anselin, 1999). En un mapa LISA se representan las zonas con valores significativos del estadístico I de Moran. Este estadístico mide el grado de concentración de valores altos/bajos de una variable en el entorno geográfico de cada una de las observaciones de la muestra. Para cada valor del estadístico, es posible evaluar el nivel de significancia estadística de rechazo de la hipótesis nula de ausencia de similitud o disimilitud de valores en una localización geográfica. De este modo, se pone de manifiesto la presencia de puntos calientes (“hot spots”) o atípicos espaciales, cuya mayor o menor intensidad dependerá de la significancia asociada de los citados estadísticos (LeSage, 1999; Anselin, 1999; Moreno, 2000; Chasco, 2003a).

En la Figura 4 puede observarse un mapa LISA en el que, a través de una gradación de colores, se representan las zonas calientes de concentraciones de localidades con alto valor (color rojo) y bajo valor (azul) en la tasa de la variable estudiada. El mapa destaca también aquellas zonas en las que se producen discontinuidades significativas en los valores de la variable.

Tabla 1. Técnicas del análisis exploratorio de datos espaciales (AEDE)

Distribución espacial	Univariante	Diagrama de mapa/caja
		Histograma de frecuencias
		Mapa de percentiles
		Cartograma
	Multivariante	Diagrama de mapa/dispersión
Asociación espacial	Global	Diagrama de dispersión de Morán
		Mapa LISA
	Local	Diagrama de mapa/caja LISA
		Puntos atípicos en el diag. de dispersión de Moran
	Multivariante	Diagrama de dispersión multivariante de Moran
Heterogeneidad espacial		Mapa del histograma de frecuencias
		Diagrama de dispersión

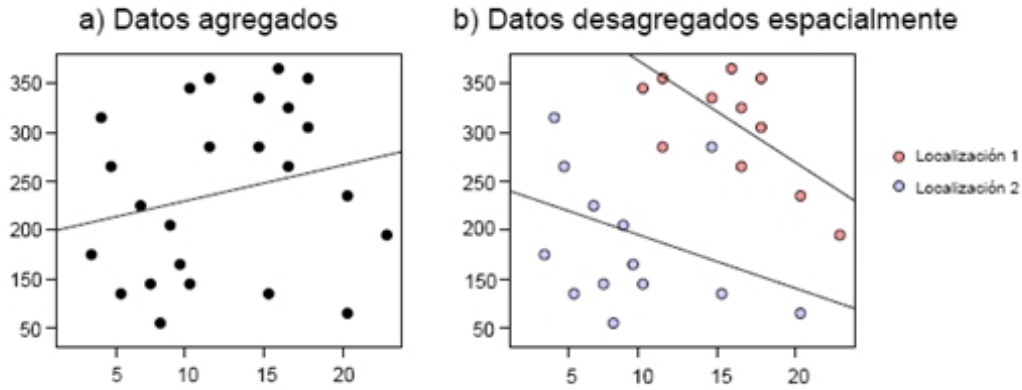


Figura 1. La Paradoja de Simpson: a) Datos agregados; b) datos desagregados (Fotheringham et al. 2002)

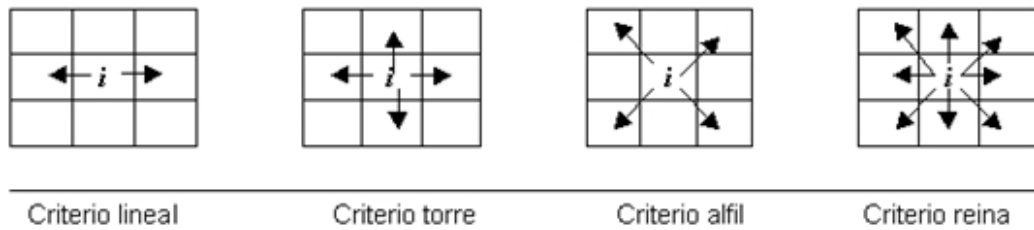


Figura 2. Criterios de vecindad

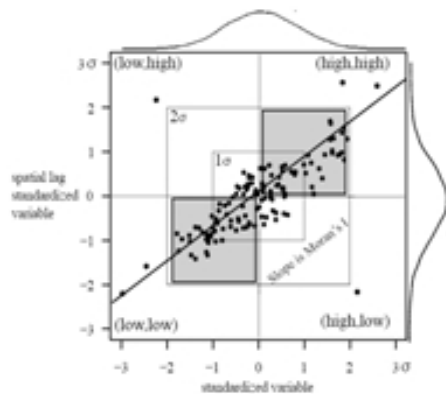


Figura 3. Diagrama de dispersión de Morán (Anselin, 2005)

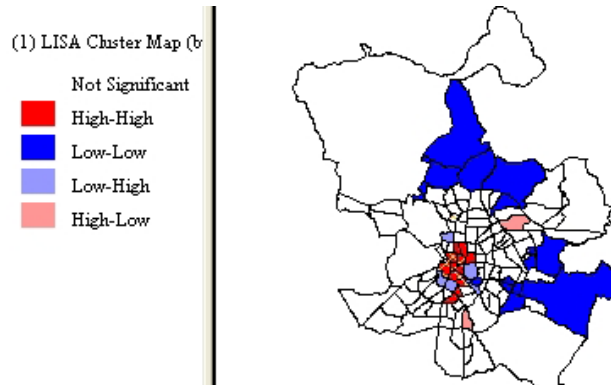


Figura 4. Gráficos LISA (Anselin, 2005)

ESPECIFICACIÓN DE MODELOS EN PRESENCIA DE INESTABILIDAD PARAMÉTRICA

Una vez que el AEDE ha descrito el panorama desde una perspectiva geográfica y evidenciado la existencia de efectos espaciales, surge la necesidad de especificar modelos adecuados.

El presente artículo se centra en el tratamiento de la HE y, en particular, en la forma de inestabilidad espacial. Sin embargo, como menciona Anselin (1999), es común que ambos efectos espaciales se presenten de manera conjunta, e incluso es posible que el impacto de uno se confunda con

el impacto del otro.

Los modelos para considerar la HE en su forma de inestabilidad paramétrica se basan en la (E.1). Como se puede apreciar, no es posible estimar los β_i , puesto que únicamente se tienen N observaciones (las N zonas que conforman la región) y se requiere estimar, al menos $2*N$ coeficientes. En otras palabras, no se tiene información suficiente para obtener una estimación para cada observación espacial.

Existen varios métodos para especificar la inestabilidad paramétrica en forma discreta y otros para modelar la

no estacionariedad de los parámetros en forma continua (LeSage, 1999; Moreno, 2000). En este artículo se considera este último tipo de inestabilidad y se presenta el modelo GWR (Fotheringham et al. 2002), como la técnica que permite resolver el problema de la falta de grados de libertad.

Modelo de Regresión Geográficamente Ponderada

Las regresiones geográficamente ponderadas (GWR en inglés) fueron desarrolladas por Fotheringham, Brunson y Charlton (FBC) (Fotheringham et al. 2002) y constituyen, probablemente, el método más empleado para analizar el problema de la HE. A partir de la Ecuación (1), se tiene:

$$y_i = X_i \beta_i + \mu_i \quad (4)$$

Ya sea por la variabilidad de la muestra, por errores en la especificación del modelo o simplemente porque existen relaciones intrínsecamente diferentes a lo largo del espacio, Fotheringham et al. (2002) consideran que no tiene sentido mantener el supuesto de estacionariedad empleado regularmente al analizar las series. Estos autores suponen que un mismo estímulo puede provocar diferentes respuestas en distintas regiones estudiadas.

En el GWR, la regresión global se convierte en un conjunto de regresiones, a través de la siguiente especificación:

$$y_i = \beta_{0i} + \beta_{1i}x_{1i} + \beta_{2i}x_{2i} + \dots + \beta_{ki}x_{ki} + u_i \quad (5)$$

Esto quiere decir que, mediante GWR, es posible obtener estimaciones para cada punto en el espacio, usando para ello una sub-muestra de los datos disponibles, que, como se verá, está referida básicamente a observaciones relativamente cercanas a dichos puntos (Chasco, 2007a). En general el punto en el espacio se hace coincidir con las coordenadas x - y del centroide de cada zona.

El procedimiento es descrito por Fotheringham et al. de la siguiente manera:

“Si trazamos un círculo de radio, por ejemplo, r alrededor de un punto concreto, p_i y resolvemos una regresión de mínimos cuadrados ordinarios del modelo con base solamente en las observaciones que están geográficamente dentro del círculo, entonces el β_i obtenido puede interpretarse como un estimador de las variables independientes alrededor de, p_i . En definitiva, son estimaciones de β_{ij} . Evaluando β_{ij} para cada p_i , es posible obtener un conjunto de estimadores de los parámetros con variabilidad espacial” (Fotheringham et al. 2002: 53).

En términos matriciales, la solución a la regresión geográficamente ponderada, correspondiente al punto i , se obtienen a partir de la siguiente expresión:

$$\beta_i = [X'W_iX]^{-1}X'W_iy \quad (6)$$

donde: W_i es una matriz diagonal de pesos especificados previamente por el analista.

El problema de la falta de grados de libertad es resuelto por el GWR a través de la asignación de una estructura de pesos para la matriz W_i , a través de funciones que dependen inversamente de la distancia y de funciones “kernel” (Chasco, 2007a).

La idea es determinar el número de observaciones con la que se ajustarán las regresiones, de manera que una distancia grande implicará más observaciones, lo que pudiera provocar, en casos extremos, que las regresiones locales sean muy similares a la regresión global. Por otra parte, con un valor de distancia pequeño, las regresiones se basarán en pocos datos, los que producirá varianzas elevadas en los modelos.

Las funciones “kernel” pueden llegar a especificar los pesos w_{ij} como una función continua monótonamente decreciente, lo que, al menos teóricamente, concordaría con la Ley de Tobler.

Por ejemplo, la función Gaussiana de decaimiento: $w_{ij} = \exp(-d_{ij}^2/h^2)$ muestra que la influencia de una observación en una zona con respecto a otra, se va “desgastando” a medida que se incrementa la distancia entre ellas. El término h es un parámetro no negativo conocido como el ancho de banda (Fotheringham et al. 2002; Mur, 2006; Chasco, 2007a, 2007b) y es de suma importancia pues es el que produce el decaimiento de la influencia entre las variables con la distancia.

En el límite cuando h tiende a cero, la banda es extremadamente angosta y las observaciones más próximas al punto de referencia recibirán una ponderación muy superior al resto de los datos (Mur, 2006).

En (Fotheringham et al. 2002) se sugiere obtener el valor óptimo del ancho de la banda minimizando la validación mínimo-cuadrática:

$$CVRSS(h) = \sum [y_i - y_{\neq i}(h)]^2 \quad (7)$$

donde: $y_{\neq i}$ es el valor estimado de y_i omitiendo el punto i en

la calibración del modelo.

Para determinar el h óptimo, se requiere realizar una serie de regresiones ponderadas en las que se emplearían diferentes radios de influencia, para finalmente seleccionar el que ofrezca el valor mínimo de $CVRSS(h)$. En Fotheringham et al. (2002) se indica que también es posible obtener el h óptimo a través del Criterio de Información de Akaike corregido (AICc).

CASO DE ESTUDIO

El caso de estudio está referido al consumo eléctrico residencial de la región geográfica correspondiente a la Gran Caracas, conformada por 613 zonas. La Tabla 2 muestra el resumen de las variables a ser utilizadas (INSURBECA, 2011).

El sufijo 08A en cada variable indica que se hace referencia al año base, esto es 2008. Así por ejemplo, PABC08A y PD08A hacen referencia a la población de estrato ABC y D, respectivamente, referidas al año 2008.

La variable dependiente CONS_RES, representa el consumo anual en kWh asociado a una zona específica, para el año 2008. Esta información se recopila a partir de la lectura de los medidores de energías asignados a cada suscriptor. Por razones de confidencialidad, toda la información procesada, se presenta como agregado estadístico.

Las Figuras 5 a 8, presentan los mapas temáticos de percentiles asociados con las variables: Consumo Residencial, Población ABC, Población D y Población E. El AEDE se realiza mediante el paquete GEODA (Anselin, 2005).

El patrón observado en todas estas figuras es básicamente el mismo: no existe uniformidad espacial de las correspondientes variables. En general se puede notar la presencia de zonas con percentiles similares, pero rodeadas

de zonas con valores distintos (superiores o inferiores). Esta distribución de valores, sugiere la existencia de una posible autocorrelación y/o /heterogeneidad espacial.

Tabla 2. Diccionario de variables (INSURBECA, 2011)

VARIABLE	DESCRIPCION
ID	IDENTIFICADOR DE ZONA
COD_ZONA	DATOS BASICOS DE ZONA
NOM_ZONA	
COD_REGION	
NOM_REGION	
PARROQUIA	
MUNICIPIO	
CONS_RES	ENERGIA ANUAL kWh 2008
PABC08A	POBLACION DE LOS ESTRATOS ABC AÑO 2008
PD08A	POBLACION DEL ESTRATO D PARA LOS AÑOS MODELADOS
PE08A	POBLACION DEL ESTRATO E PARA LOS AÑOS MODELADOS
XOO	COORDENADA X DEL CENTROIDE DE LA ZONA
Y00	COORDENADA Y DEL CENTROIDE DE LA ZONA

A continuación se presentan los diagramas de Moran y su indicador I , con la intención de evaluar la posible existencia de fenómenos de autocorrelación espacial. Las Figuras 9 a 12 muestran los diagramas de Moran para las variables: Consumo Residencial, Población ABC, Población D y Población E. Todos los cálculos están basados en una vecindad tipo reina. En el eje de la abscisa se muestra el promedio de la variable en estudio y en el eje de la ordenada el promedio de las zonas vecinas, de acuerdo con el tipo de vecindad utilizado.

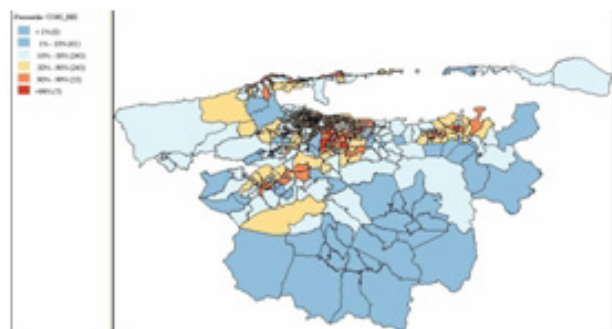


Figura 5. Mapa temático con Percentiles asociados con el Consumo Residencial (año base 2008)

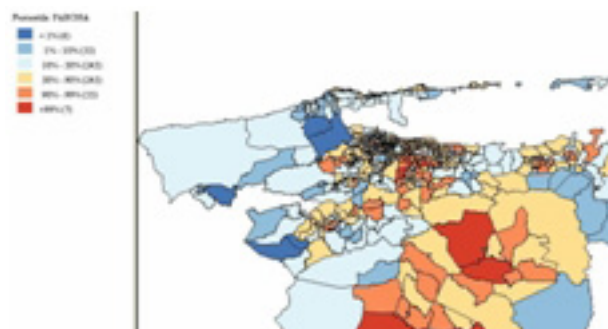


Figura 6. Mapa temático con Percentiles asociados al Población ABC (año base 2008)

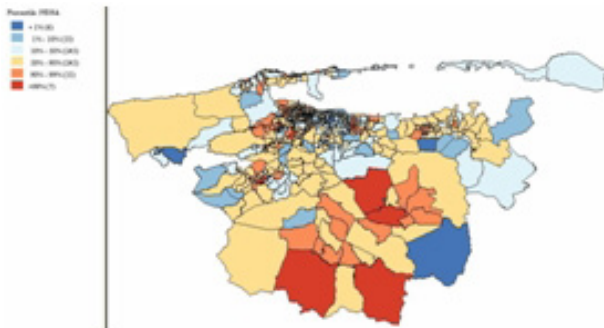


Figura 7. Mapa temático con Percentiles asociados con la Población D (año base 2008)

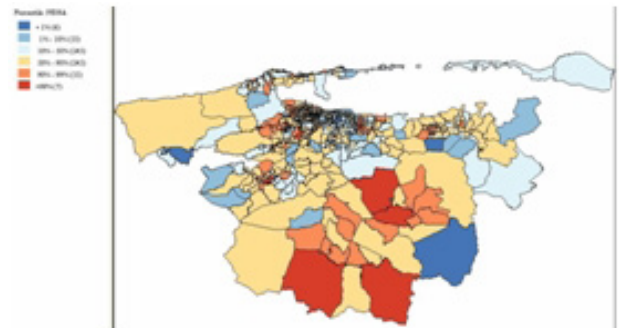


Figura 8. Mapa temático con Percentiles asociados con la Población E (año base 2008)

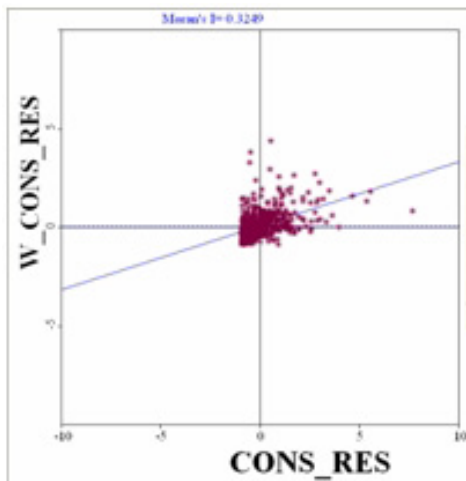


Figura 9. Diagrama de Moran e indicador de autocorrelación para consumo Residencial (año base 2008)

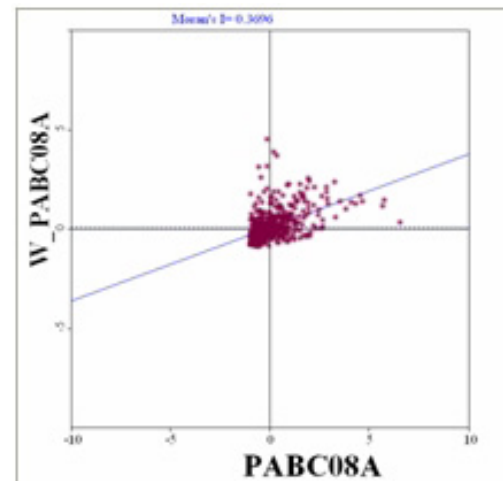


Figura 10. Diagrama de Moran e indicador de autocorrelación para Población ABC (año base 2008)

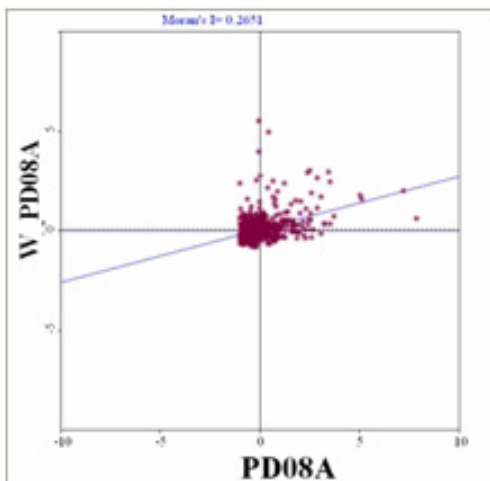


Figura 11. Diagrama de Moran e indicador de autocorrelación para Población D (año base 2008)

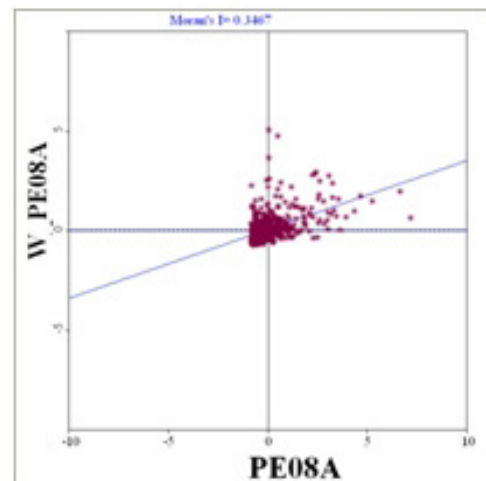


Figura 12. Diagrama de Moran e indicador de autocorrelación para Población E (año base 2008)

El análisis de estos diagramas revela la existencia de un fenómeno de autocorrelación positiva significativa. Es importante resaltar que, si bien a primera vista pudiera pensarse que los valores del Indicador I son bajos, los índices están basados en 613 zonas y todos los valores del indicador I son estadísticamente diferentes de cero (p-value

inferior a 0,01).

Si bien la existencia de autocorrelación espacial puede responder a razones muy particulares, en este artículo no se analizan estas razones puesto que el objeto principal es verificar la posible existencia de un fenómeno de

autocorrelación, que invalide uno de los supuestos básicos del modelo clásico de regresión.

A continuación se presentan los mapas LISA para las variables seleccionadas. Como se mencionó anteriormente, en un mapa LISA se representan las zonas con valores significativos en determinados indicadores estadísticos de asociación espacial local.

Los gráficos LISA se basan en el estadístico I de Moran y ponen de manifiesto la presencia de puntos calientes (“hot spots”) o atípicos espaciales, características esenciales para determinar la existencia de heterogeneidad espacial.

Las Figuras 13 a 16 muestran los mapas LISA para las variables: Consumo Residencial, Población ABC, Población D, Población E y Población Total.

El patrón observado en todas estas figuras es básicamente

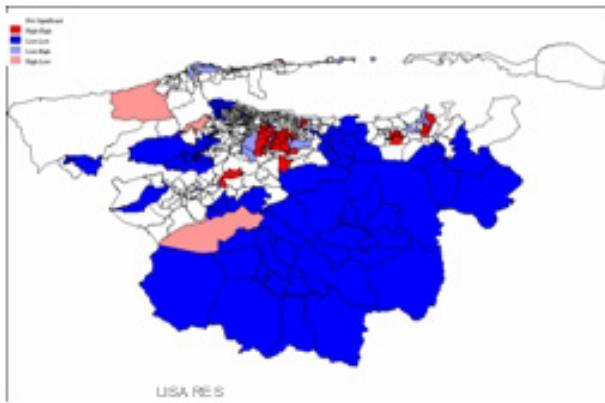


Figura 13. Mapa LISA Consumo residencial (año base 2008)

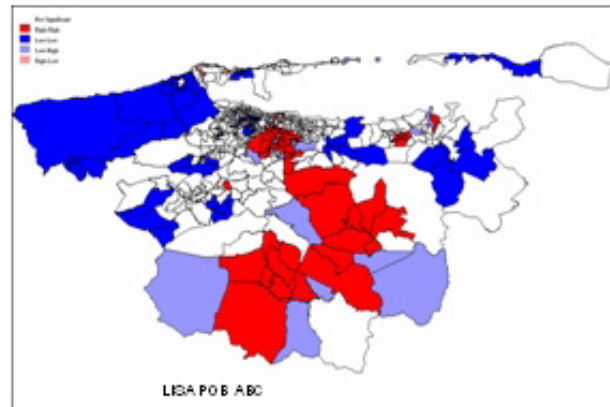


Figura 14. Mapa LISA población ABC (año base 2008)

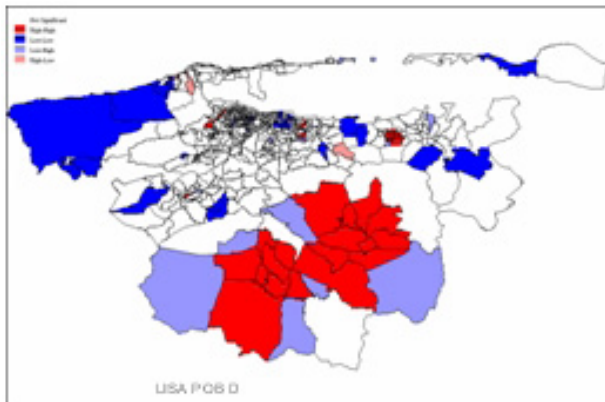


Figura 15. Mapa LISA población D (año base 2008)

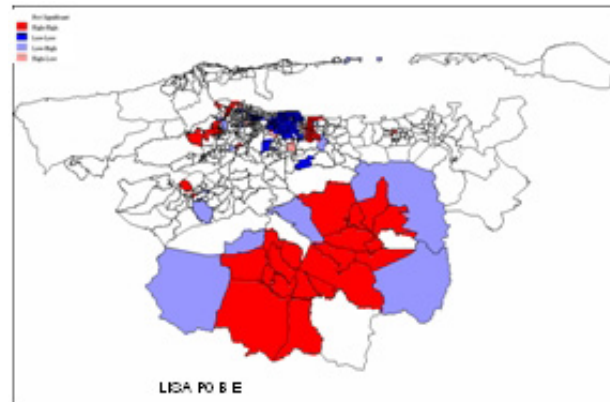


Figura 16. Mapa LISA población E (año base 2008)

el mismo, caracterizado por la existencia de atípicos espaciales, así como de zonas homogéneas particulares (por ejemplo zonas en color azul rodeadas de zonas en color azul) que, al igual que los mapas temáticos de percentiles, surgieren la existencia de una posible heterogeneidad espacial.

Ante la evidencia sugerida por el análisis espacial de datos de fenómenos de autocorrelación y/o heterogeneidad, se consideró el uso de GWR. La selección del mejor modelo se basó en las siguientes consideraciones:

- a) Signos esperados de los coeficientes del modelo.
- b) Se analizan modelos lin-lin, log-lin, log-log (el primer

indicador está referido a la variable dependiente mientras que el segundo a las variables independientes).

- c) Valores del coeficiente de determinación ajustado y del índice de AIC. Una variación de 3 unidades en el AIC es considerada como un modelo de aporte estadísticamente significativo (Fotheringham *et al.* 2002).
- d) Pruebas t para los parámetros del modelo
- e) Pruebas F para la validación global del modelo
- f) Diagrama de Moran e índice I de los residuos del modelo
- g) Mapas LISA de residuos

Las regresiones ponderadas geográficamente son evaluadas

mediante el paquete GWR (Charlton . 2010).

En todas las formas funcionales analizadas, la variable asociada con población del estrato E (PE08A) resultó ser estadísticamente no significativa. Una explicación desde el punto de vista eléctrico puede sugerir que, al ser ésta la población de menores recursos, sus viviendas no deben estar registradas como suscriptores (no poseen medidores) y por ende su consumo no es contabilizado.

La Tabla 3 muestra los resultados del mejor modelo ajustado, correspondiente a la especificación funcional lin-lin, mediante el modelo de regresión clásico global.

Se puede observar que el modelo presenta un valor relativamente bajo del R^2 ajustado (0,60), un valor de F altamente significativo e índice AIC de 20290 (como se mencionó, el AIC se usará para comparar distintos modelos). Todos los coeficientes son significativamente distintos de cero. Sin embargo el signo de PD08A es contrario al esperado.

El valor para el coeficiente PABC08A de 2363 kWh/año promedio por suscriptor, supone un consumo promedio mensual de aproximadamente 200 kWh/mes.

La Figura 17 muestra el diagrama de Moran con índice I de 0,2392, significativamente distinto de cero, que confirma la existencia de autocorrelación espacial positiva en los residuos del modelo global.

Como corroboración adicional, la Figura 18 muestra el mapa LISA de los residuos para el caso global: se evidencia claramente el efecto de heterogeneidad espacial.

Tabla 3. Resultados para el Modelo lin-lin asociado con Consumo residencial, mediante Mínimos Cuadrados Ordinarios (salida de GWR)

Parameter	3	
AIC	20290,29	
R^2	0,600369	
R^2 -adj	0,597740	
Parameter	Estimate	t
Intercept	560975	2,02
PABC08A	2719,27	24,31
PD08A	-540,81	-7,57

La Tabla 4 muestra los resultados de GWR. Se utiliza kernel tipo “Bi-Square” adaptativo, cuyo ancho de banda óptimo se determinó minimizando el AIC.

Tabla 4. Resultados de GWR para el Modelo lin-lin asociado a Consumo residencial

Parameter	3		
X-coord	XOO		
Y-coord	YOO		
Kernel	Adaptive		
Kernel Type	Bi-Square		
Bandwith selection	AIC minimisation		
AIC	19314,63		
R^2	0,963788		
R^2-adj	0,9489501		
ANOVA			
Source	SS	d.f	F
OLS residuals	8734,80 E+12	4.00	
GWR Improvement	7615,92 E+12	173.88	
GWR Residuals	7588,72 E+12	435.12	25,11

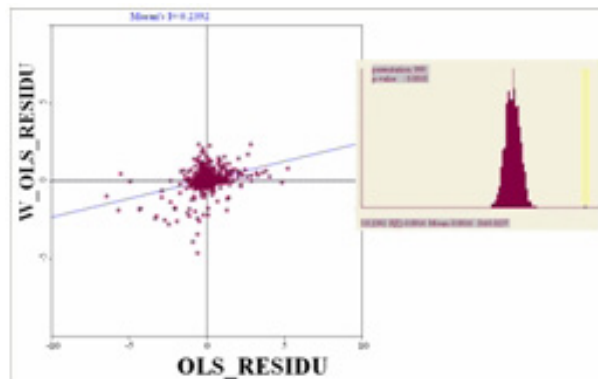


Figura 17. Diagrama de Moran e indicador de autocorrelación para residuos del modelo GLOBAL para Consumo Residencial (año base 2008)

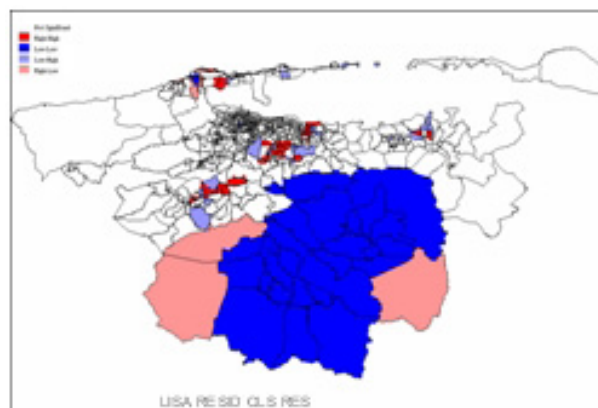


Figura 18. Mapa LISA residuos del modelo GLOBAL para Consumo Residencial (año base 2008)

En primer lugar resalta el valor de R^2 ajustado, cercano a 0,95 y el AIC muestra un valor de aproximadamente 19314, casi 900 unidades inferior al modelo clásico de regresión, indicando una contribución estadísticamente significativa, al observar que la prueba F muestra un valor superior a 25.

Una de las características fundamentales del modelo GWR es que permite determinar los coeficientes para las variables independientes en cada zona (note que en este caso es como si se ajustaran 613 modelos clásicos de regresión, uno para cada zona).

Esta característica permite analizar en detalle cada zona y determinar comportamientos particulares, tales como que el consumo de un cierto tipo de población en una zona sea distinto o igual a otra o que por el contrario existan comportamientos anómalos.

Las Figuras 19 y 20 muestran la distribución espacial asociadas a los coeficientes de las variables PABC08A y PD08A, indicando que los valores por zonas son, en principio, diferentes.

La Tabla 5 muestra el resumen de la distribución de los coeficientes asociados a cada variable. Por ejemplo, los coeficientes asociados a Población estrato ABC, presentan un valor mínimo de -3908 pero valores positivos a partir del cuartil inferior. Algo parecido sucede con los valores asociados a Población estrato D.

Si bien, los valores negativos violan el supuesto del signo acorde con la teoría, el análisis detallado de los valores asociados con PABC08A, revela que este comportamiento se observa sólo en ocho zonas del estudio, que corresponden a zonas con muy poca presencia de población ABC, siendo mayoritariamente la presencia de población de los estratos E.

Como se mencionó anteriormente, en estas zonas existen pocos suscriptores formales, por lo que el modelo trata de compensar con valores de coeficientes negativos para la variable PABC08A.

Tabla 5. Resultados de GWR para el Modelo lin-lin asociado con Consumo residencial, resumen distribución de los parámetros

Parameter	Minimum	Low Quartile	Median	Upr Quartile	Maximum
PABC08A	-3908,15	1900,87	2576,38	3314,37	8767,81
PD08A	-2159,30	-41,50	248,84	1541,61	3476,90

Para los coeficientes asociados con PD08A, los valores negativos ocurren mayoritariamente en zonas de bajos recursos económicos. En estas zonas, si bien existen muchos suscriptores formales, es posible que parte de la población estrato D no tenga medidor de consumo asociado, por lo que el modelo trata de compensar con valores de coeficientes negativos para la variable PD08A. Evidentemente estas zonas requieren recopilación de información adicional.

La Figura 21 presenta la relación entre valores observados (datos) y los valores estimados por el modelo para cada zona analizada, sugiriendo un ajuste muy bueno.

Para corroborar el ajuste del modelo GWR, la Figura 22 presenta el diagrama de Moran de los residuos: claramente se observa que el índice I es cercano al valor esperado en caso de no autocorrelación. Este hecho sugiere que los residuos obtenidos no presentan evidencia de autocorrelación espacial y que el modelo GWR la ha modelado adecuadamente.

La Figura 23 muestra el diagrama LISA de los residuos asociados al modelo GWR ajustado: se nota la existencia de solamente un pequeño número de zonas atípicas, que sugieren que el modelo GWR es adecuado.

Finalmente, el modelo GWR también reporta los valores de R^2 ajustado para cada zona. La Figura 24 muestra la distribución asociada a dichos valores, que van desde 0,72 hasta valores cercanos a 0,99, siendo la moda cercana a 0,90.

CONCLUSIONES

Al hablar de datos espaciales es necesario tomar en cuenta que los mismos son georeferenciados, multidireccionales y multidimensionales. Estas características determinan que su empleo pueda provocar la aparición de los denominados Efectos Espaciales: la Dependencia o Autocorrelación Espacial (AE) y la Heterogeneidad Espacial (HE).

La AE es la ausencia de independencia entre las observaciones en un espacio, mientras que la HE se refiere a la falta de uniformidad en las relaciones entre las variables a través de las zonas. El AEDE permite identificar estos patrones en el comportamiento de variables georeferenciadas. El análisis de estos efectos resulta importante puesto que permite hacer más reales los modelos, al introducir, de forma explícita, las diferencias entre zonas en el análisis.

El análisis de la región de estudio (formada por 613 zonas) permitió visualizar, de manera clara y sencilla, la

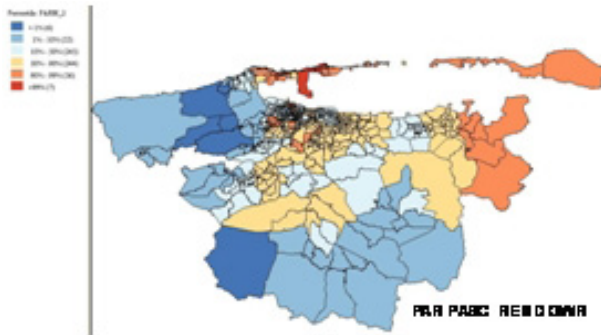


Figura 19. Coeficientes asociados a PABC

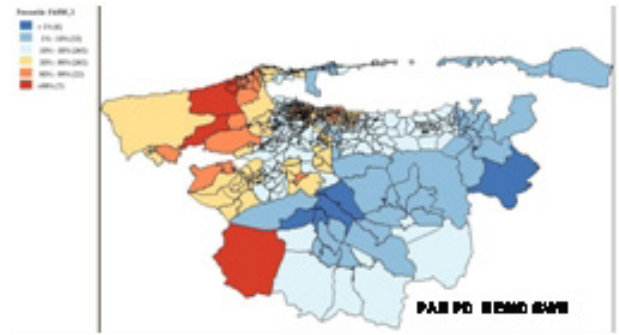


Figura 20. Coeficientes asociados a PD

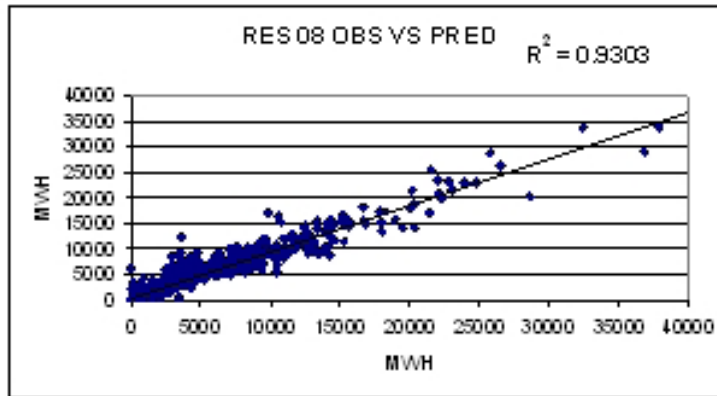


Figura 21. Comparación de valores observados y estimados, modelo lin-lin, GWR para Consumo Residencial (año base 2008)

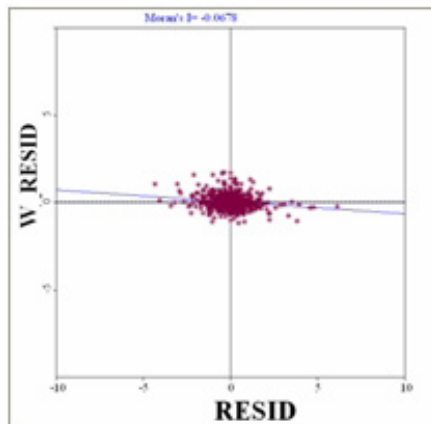


Figura 22. Diagrama de Moran para residuos del modelo lin-lin GWR para Consumo Residencial (año base 2008)

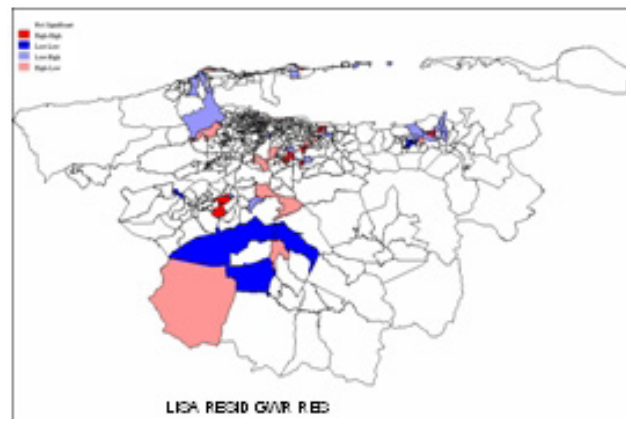


Figura 23. Mapa LISA residuos del modelo lin-lin GWR para Consumo Residencial (año base 2008)

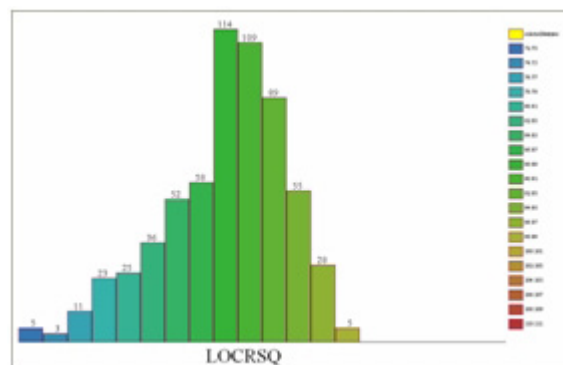


Figura 24. Distribución local de los valores del coeficiente de determinación, Modelo lin-lin para Consumo Residencia (año base 2008), GWR

presencia de efectos espaciales de AE y HE. Se analizaron los resultados obtenidos con un modelo global, que no considera efectos espaciales y el modelo GWR, que si los considera.

Los resultados del modelo GWR son mejores desde el punto de vista estructural así como desde el punto de vista de bondad de ajuste (R^2 superior a 0,90), permitiendo cuantificar las diferencias entre las variables independientes en cada zona.

El siguiente paso es estimar los consumos para años posteriores. Para esto, es necesario contar con las estimaciones de cada una de las variables independientes definidas en el modelo, lo que permitirá generar las estimaciones de consumo tanto de energía como de potencia, para cada zona. Esta información diferenciada por zonas, está en concordancia con los requerimientos definidos por Willis (Willis, 2002) para una planificación adecuada de la expansión de un sistema eléctrico de potencia, que requiere conocer con anticipación: ¿Dónde ocurre la demanda? ¿Cuál es su valor? ¿Cuándo ocurre?

AGRADECIMIENTO

Parte de este trabajo fue financiado bajo el Proyecto de Investigación CDCH-UCV-PI 08-7619-2009.

REFERENCIAS

- ANSELIN, L. (2005). Exploring Spatial Data with Geoda: A Workbook. University of Illinois.
- ANSELIN, L. (1999). Spatial Econometrics. University of Texas at Dallas, pp.23
- ASTORKIZA, I. (2006). El precio de la tierra rústica en un espacio natural protegido: análisis econométrico en el caso de la reserva de la biosfera de Urdaibai. Universidad del País Vasco.
- CHARLTON, M., BRUNSDON, CH., FOTHERINGHAM, S. (2010). Software for geographically weighted regression. Release 3.0.1, University of Newcastle upon Tyne.
- CHASCO, C. (2005). Análisis exploratorio de datos espaciales al servicio del geomarketing. Instituto Lawrence R. Klein, Universidad Autónoma de Madrid.
- CHASCO, C. (2003a). Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales. Comunidad de Madrid.
- CHASCO, C. (2003b). Modelos de heterogeneidad espacial. Universidad Autónoma de Madrid.
- CHASCO, C. (2007a) Modeling spatial variations in household disposable income with geographically weighted regression. Universidad Autónoma de Madrid.
- CHASCO, C. (2007b). Modelizando la no estacionariedad espacial. Universidad Autónoma de Madrid.
- FOTHERINGHAM, S., BRUNSDON, CH., CHARLTON, M. (2002). Geographically Weighted Regression. John Wiley & Sons.
- FRANCISCO, E. DE R., FAGUNDES, E., PONCHIO, M.C., ZAMBALDI, F. (2010). Development of an indicator of propensity to energy commercial losses using geospatial statistical techniques and socio-economic data: the case of AES Eletropaulo. RAM, Rev. Adm. Mackenzie (Online), vol.11, n.4.
- FRANCISCO, E. DE R. (2006). Relação entre o consumo de energia elétrica, a renda e a caracterização econômica de famílias de baixa renda do município de São Paulo. Master Thesis (Mestrado em Administração de Empresas)–Escola de Administração de Empresas de São Paulo da Fundação Getulio Vargas, São Paulo, Brazil.
- FRANCISCO, E. DE R. & FAGUNDES, E. B. (2006). Geostatistical study between social-economical variables and power losses. In: CONGRESS OF OPERATIONAL EXCELLENCE AND INNOVATION OF AES LATIN AMERICA, 1., 2006, São Paulo, Brazil. Anais. São Paulo: AES Latin America e Africa.
- GÓMEZ, M. (1999). Econometría espacial: algunos aspectos generales. Universidad Complutense de Madrid.
- INSURBECA, (2011). Estimación de la demanda eléctrica a largo plazo para el Distrito Capital y los estados Miranda, Vargas y Aragua” (Instituto asociado a la Facultad de Arquitectura y Urbanismo, UCV).
- LESAGE, J. (1999). The theory and practice of spatial econometrics. University of Toledo.
- MONTEIRO, C. & RAMIREZ-ROSADO, I. (2001) Applications of Geographical Information Systems for the Optimal Locations of Wind Electric Energy Generation”, MIC200, Modelling and Identification and Control.

- MONTEIRO C., MIRANDA V., RAMIREZ-ROSADO, I. (2001). Spatial Decision support systems for site permitting of distributed generation facilities, Proceedings of IEEE Porto Power Tech.
- MONTEIRO C. (2003). Fuzzy Spatial Forecasting, Tesis Doctoral, Universidad de Porto, Portugal.
- MORENO, R. (2000). Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial. Edicions de la Universitat de Barcelona.
- MORENO, R. (2002). Econometría espacial: nuevas técnicas para el análisis regional. Una aplicación a las regiones europeas. Asociación española de ciencia regional.
- MUR, J. (2006). Modelos de corte transversal. Universidad de Zaragoza.
- SÁNCHEZ, L. (2006). Métodos para el análisis espacial. Una aplicación al estudio de la geografía de la pobreza. Universidad de Wisconsin-Madison.
- WILLIS, H.L. (2002). Power Distribution Planning Reference Book, Second Edition, Revised and Expanded”, Marcel Dekker, New York.