

Modelo Lineal Generalizado con respuesta Beta para estudiar el Índice de Escasez Hídrica en el Caribe colombiano incluyendo efectos espaciales

Generalized Linear Model with Beta Answer to Study the Index of Water Shortage in the Colombian Caribbean Including Spatial Effects

Miguel Alejandro Gonzalez Ruiz¹
Luis Eduardo Gómez Daza²
Luis Fernando Santa Guzmán³

Para citar este artículo, utilice el nombre completo así:

Gonzalez, M. A., Gómez, L. E. & Santa, L. F. (2015). Modelo Lineal Generalizado con respuesta Beta para estudiar el Índice de Escasez Hídrica en el Caribe colombiano incluyendo efectos espaciales. *Perspectiva Geográfica*, 20(1), 175-202.

Resumen

El análisis de datos espaciales contempla la inspección, selección y transformación de datos, con el fin de mostrar información útil. Este artículo busca explicar, por medio del análisis de datos espaciales, el comportamiento y la distribución del Índice de Escasez Hídrico en la región del Caribe

1 Ing. Catastral y Geodesta en formación. Universidad Distrital Francisco José de Caldas. miagonzalezr@correo.udistrital.edu.co

2 Ing. Catastral y Geodesta en formación. Universidad Distrital Francisco José de Caldas. luegomezd@correo.udistrital.edu.co

3 Ing. Catastral y Geodesta, Estadístico y Magíster en Geomática. Integrante del Grupo de Investigación Estadística Espacial. Profesor asistente de la Universidad Distrital Francisco José de Caldas lfsantag@unal.edu.co

colombiano. Se hace una selección de variables: distribución del recurso, características sociodemográficas y distribución del recurso en términos de la cantidad de precipitación; se emplean técnicas para evaluar la presencia de dependencia espacial, las cuales demuestran que es necesario, luego de aplicar un análisis exploratorio de datos, incluir efectos espaciales en la estimación de los parámetros de los modelos estadísticos.

Palabras clave: efectos espaciales, escasez hídrica, estadística espacial, índice de escasez, Modelo Lineal Generalizado (GLM).

Abstract

The analysis of spatial data is responsible for inspecting, selecting, and transformation of data in order to display useful information. This article seeks to explain by analyzing spatial data behavior and distribution of water scarcity index in the Colombian Caribbean region. A selection of variables which take into account the distribution of the resource, the sociodemographic characteristics and resource distribution in terms of the amount of precipitation is done; techniques are used to assess the presence of spatial dependence which reveals a need after applying an exploratory data analysis, including spatial effects in the estimation of the parameters of statistical models.

Keywords: spatial effects, water scarcity, spatial statistics, scarcity index, Generalized lineal Models (GLM).

Introducción

Desde el punto de vista práctico del índice de escasez hídrico, en Colombia, el trabajo sobre el particular, se limita a realizar estudios del agua cada cuatro años pero, más allá de cuantificar o dar estadísticas generales en el ámbito nacional, no se tiene explicación del por qué está ocurriendo este fenómeno y cuáles son las variaciones, tanto en el espacio como en el tiempo.

La creciente necesidad de entender las causas del comportamiento espacial de fenómenos sucedidos en la superficie terrestre y que afectan la supervivencia de las especies ha hecho surgir modernas técnicas de análisis de datos que se ajustan, de manera más coherente, a la realidad analizada, y permiten ofrecer una mejor explicación de cómo ocurren estos comportamientos y, así, generar posibles escenarios para apoyar la toma de decisiones. Es en este punto donde se plantea el objetivo principal de este artículo, pues busca establecer, a través de un modelo de regresión tipo *beta*, el cual es adecuado para describir porcentajes, proporciones, fracciones y razones (Ferrari & Cribari-Neto, 2004), el problema de la escasez hídrica en la Costa Caribe

colombiana y, además, incorporar métodos como la regresión geográficamente ponderada para corregir inadecuaciones en los modelos estimados, debido a la presencia de efectos espaciales.

A pesar de que Colombia cuenta con una disponibilidad generosa de agua y poco se ha preocupado por la misma, este recurso no es infinito ni está distribuido homogéneamente en el territorio (Costa Posada, et al., 2005). Además, el crecimiento demográfico, el desarrollo económico, la urbanización y la contaminación están ejerciendo una presión sin precedentes sobre los recursos hídricos renovables; en especial, sobre regiones donde su oferta es escasa, como lo es el norte del país (FAO, 2013).

Durante muchos años, la disponibilidad del recurso hídrico de los países del mundo ha sido evaluada de acuerdo con la cantidad de agua del territorio. Estimaciones de este tipo se pueden destacar en: el Programa para la Evaluación Mundial de los Recursos Hídricos (Lvovitch, 1986) y en el Programa Hidrológico Internacional (UNESCO, 2015). Del mismo modo, son muchas las aplicaciones en las cuales no solo se cuantifica el recurso

sino que, además, se buscan explicar las causas que modifican su distribución a través del tiempo y el espacio. Son varios los estudios asociados con cambio climático, sequía y modelos hidrológicos, entre otros.

En 2003, en México, se desarrolló un modelo econométrico con datos obtenidos a través de una encuesta que realizó el Instituto Nacional de Ecología en diferentes regiones del país, en donde se tuvo en cuenta variables, como el consumo en domicilios y características físicas y socioeconómicas de los hogares (Jaramillo-Mosqueira, 2005). En España, se realizó un análisis econométrico para optimizar el recurso hídrico en la producción agrícola (Portero & Agulló, 2002). En Bangladesh, se investigó el manejo de la escasez hídrica por parte de agricultores de arroz en un clima semiárido; el análisis mostró que los individuos con más experiencia en agricultura, mejor educación, con derechos de tenencia de la tierra más seguros, acceso a la electricidad e instalaciones institucionales y conciencia de los efectos climáticos son más propensos a adoptar mejores estrategias de adaptación a la falta del recurso (Alam, 2015).

En otra investigación, en la Cuenca del Río Bukhan, en Corea, se desarrolló un nuevo marco para cuantificar la vulnerabilidad espacial para la gestión de los recursos hídricos sostenibles a través de cuatro índices de vulnerabilidad hidrológicos; estos se modificaron para cuantificar los daños por inundaciones, sequías, deterioro de la calidad del agua, y riesgo general de cuencas; se consideró el impacto del cambio climático (Jun, et al., 2011). En España, más específicamente en una cuenca árida y semiárida en el sur del país, se realizó un modelo hidro-económico que vincula una forma reducida del componente hidrológico, con componentes económicos y ambientales. Los resultados indicaron que los eventos de sequía tienen grandes impactos en el bienestar social, con los principales ajustes sufridos por el riego, el medio ambiente y demás (Kahil, et al., 2015).

De igual manera se destaca, en el territorio colombiano, la evaluación de la importancia del índice de escasez, como indicador para el manejo adecuado del recurso hídrico, basados en el estudio al río Pamplonita y de los municipios que se abastecen de esta cuenca hídrica. Se tuvo en cuenta, principalmente, variables como la oferta

y la demanda; se realizó un análisis de las falencias y soluciones que se puede concebir en este contexto (Romero & Ortiz, 2008).

La regresión de tipo *beta* fue introducida por Ferrari & Cribari-Neto (2004) mediante una reparametrización en función de la media y la varianza. El modelo propuesto es útil para situaciones en las que la variable de interés es continua y restringida al intervalo (0, 1) y se relaciona con otras variables a través de una estructura de regresión. Los parámetros del modelo son interpretables en términos de la media de la respuesta. La estimación se realiza por máxima verosimilitud y las pruebas de hipótesis utilizan aproximaciones obtenidas con la normalidad asintótica del estimador de máxima verosimilitud.

Aplicaciones de la regresión *beta* pueden encontrarse en Branscum et al. (2007), Eskelson et al. (2011) y Knoche et al. (2014). En el primer caso se estudiaron factores, tales como el tiempo y la distancia geográfica variables que podrían influir en la dirección de la evolución de la fiebre aftosa, por medio de un modelo de regresión jerárquico bayesiano para el análisis estadístico de los datos continuos

con espacio muestral restringido al intervalo (0, 1). Los datos se modelaron mediante distribuciones *beta* con medias que dependen de covariables a través de una función de enlace. En el segundo caso, comunidades de vegetación del sotobosque son componentes críticos de los ecosistemas forestales, para lo cual se modeló las medidas de abundancia, como la cobertura de arbustos que toman valores entre 0 y 1, exhibe heterocedasticidad y son, a menudo, objeto de dependencia espacial. Estas características de la distribución tienden a ser ignoradas cuando se analizan los datos. En el último caso, se muestra una alternativa para realizar inventarios forestales mediante la utilización de correlaciones entre el porcentaje de cobertura del dosel y variables forestales más fáciles de medir, especialmente el área basal con análisis de regresión *beta*; los resultados mostraron un buen desempeño en áreas dominadas por coníferas.

Por estas razones, este artículo tiene como propósito encontrar algunos factores que expliquen el índice de escasez, como una herramienta para establecer posibles factores de riesgo de desabastecimiento del recurso hídrico. En el modelo *beta* no

se tiene en cuenta la dimensión espacial del fenómeno. La dimensión espacial en los modelos generalmente está enfocada en dos aspectos: el problema de la dependencia espacial y la heterogeneidad espacial (Anselin, 1988). La dependencia espacial es la similitud de valores entre regiones vecinas y el problema de la heterogeneidad espacial hace referencia al hecho de que los valores cambian muy fuertemente de un lugar a otro sin ser constante, a través del espacio, la variabilidad del fenómeno.

El ignorar la presencia de efectos espaciales, tales como la dependencia y la heterogeneidad, puede conllevar problemas de especificación en la estimación de un modelo estadístico; por tanto, el riesgo de hacer estimaciones no eficientes es muy alto en cuanto la representación de tal relación entre las variables; se requiere someter los residuos a pruebas específicas (Moran y Geary), para mostrar la aleatoriedad espacial (García Palomares, et al., 2012). En este sentido, los datos del índice de escasez hídrico por la heterogeneidad de los municipios del país y los cambios en la distribución del recurso agua, muy posiblemente, tendrán problemas de heterogeneidad espacial, por lo cual

es necesario incluir, en el modelo de regresión beta, el efecto espacial para modelar la heterogeneidad; y la técnica más usada es la Regresión Geográficamente Ponderada, consistente en generar un modelo para cada región en el espacio. Es decir, si se tienen 20 unidades en el espacio, se tendrán 20 modelos correspondientes a cada una de las regiones. La estimación de los parámetros del modelo se hace a través de regresiones por ventanas, las cuales permiten reducir el efecto de la variabilidad de los datos en toda la región de estudio.

Este proceso permite moverse desde una perspectiva global a un análisis local del problema, para obtener más detalle y precisión (Lloyd & Shuttleworth, 2005). En la gran mayoría de los casos esta clase de regresión genera menor error de estimación en relación con los modelos tradicionales, además de anular o reducir el problema de la autocorrelación espacial (Hadayeghi, et al., 2010). Al proveer resultados específicos para cada localización, estos pueden ser usados como evidencias para apoyar políticas o tomas de decisiones locales; por eso, con frecuencia, estas técnicas son llamadas “basadas en el lugar” (García Palomares, et al., 2012).

Planteamiento del problema

Precisamente, como no se tienen estudios, el país se está exponiendo a situaciones, como la que lo está afectando ahora: un fenómeno del niño muy fuerte que disminuye los volúmenes de los cuerpos de agua tanto para producción de energía, como para consumo humano lo que genera, a largo plazo, problemas en la calidad de vida; es decir, un problema mal estudiado desde el punto de vista espacial se convierte en un problema que afecta a la sociedad de manera directa.

La complejidad en la cuantificación del estado de algunos hechos particulares, en muchas ocasiones, implica el uso de variables proxy o indicadores como una alternativa para aproximarse a la medición de cantidades no observables de interés (Lewis-Beck, et al., 2003), tales como: la inteligencia, la pobreza, el desarrollo económico, el desarrollo humano y la huella ecológica, entre otros. Estos indicadores son, generalmente, construidos con técnicas estadísticas y son el resumen o la síntesis de variables que son directamente observables sobre los objetos de estudio.

Una de las principales características del uso de variables latentes es la dificultad de su actualización, lo que restringe la posibilidad de contar con información en tiempo real de la situación estudiada, ya que esto implica un alto consumo de recursos en la medición de las variables observables que las definen y, por tanto, limitan los análisis y las decisiones que puedan llegar a ser tomados con base en ellas.

Como consecuencia de esto, es común observar que se empleen representaciones cartográficas de los valores de los indicadores asociados a entidades regionales, tales como veredas, municipios, departamentos, etc. Ello que restringe el estudio de los fenómenos a la simple descripción del estado actual de la situación y no trasciende la posibilidad de la construcción de modelos para explicar el comportamiento de los fenómenos, así como a la identificación de factores subyacentes asociados a la variación espacial y temporal de los mismos.

Un ejemplo particular de este hecho es desarrollado en el Estudio Nacional del Agua (ENA) (IDEAM, 2010) en Colombia, el cual proporciona valores del índice de escasez hídrico, definidos como

el uso del recurso agua en función de condiciones de oferta y demanda bajo ciertos períodos de tiempo para cada municipio del país. Dichos valores son representados en formato cartográfico sin proporcionar una explicación del comportamiento del recurso. Es decir, tan solo se aproxima a la medición de la escasez, pero no se estudian los posibles factores detonantes que permitan cuantificar el riesgo, pues el índice refleja qué tan expuesta está la población. Por tanto, surge la necesidad de desarrollar modelos de análisis de datos espaciales para explicar el comportamiento del índice de escasez del recurso hídrico (Regresión *Beta* y Regresión Geográficamente Ponderada), de tal manera que se identifiquen los principales factores determinantes de la situación, y que esto pueda servir de base para la toma de decisiones por parte de las entidades territoriales.

Descripción del lugar

El estudio se realizó para los departamentos del Atlántico, Cesar, La Guajira y Magdalena, localizados en la Región Caribe de Colombia, que comprenden un total de 81 municipios (ver Figura 1). Se estimó que la población para el año 2012

en los departamentos mencionados previamente, fue de 5'463.541 habitantes (DANE, 2005).

En el departamento del Atlántico se encuentran los siguientes humedales: Ciénaga de Mallorquín, Ciénaga Los Manatíes, Embalse del Guajaro, Ciénaga de Tocagua, Ciénaga Convento, Ciénaga Sanaguare, Ciénaga La Vieja, Ciénaga de Lurua-co, Ciénaga La Bahía, Ciénaga de Balboa, entre otras, incluyendo la cuenca del río Magdalena. De las 22.000 hectáreas aproximadas de humedales se estima 16.000 Ha. en el Embalse el Guájaro, por lo que es fácil evidenciar la distribución poco equitativa de estas reservas hídricas (Plan de Acción, 2012). Solo se ha realizado estudios relacionados con aguas subterráneas en esta zona, en alrededor de 5% del área total del territorio; se desconoce, así, el caudal captado por sectores económicos. La Costa Caribe es una de la regiones con más alto Índice de Pobreza Multidimensional junto con la Costa Pacífica, con valores de 34.6 para el año 2014 (Moreno & Estrada, 2013).

Luego, para identificar la problemática, es necesario contextualizar que para departamentos como Atlántico se tiene una cobertura de

abastecimiento hídrico para el 62% de uso resulta ser muy alto (Plan de Acción, 2012).
 de la población y que, en relación a la oferta y la demanda, el indicador

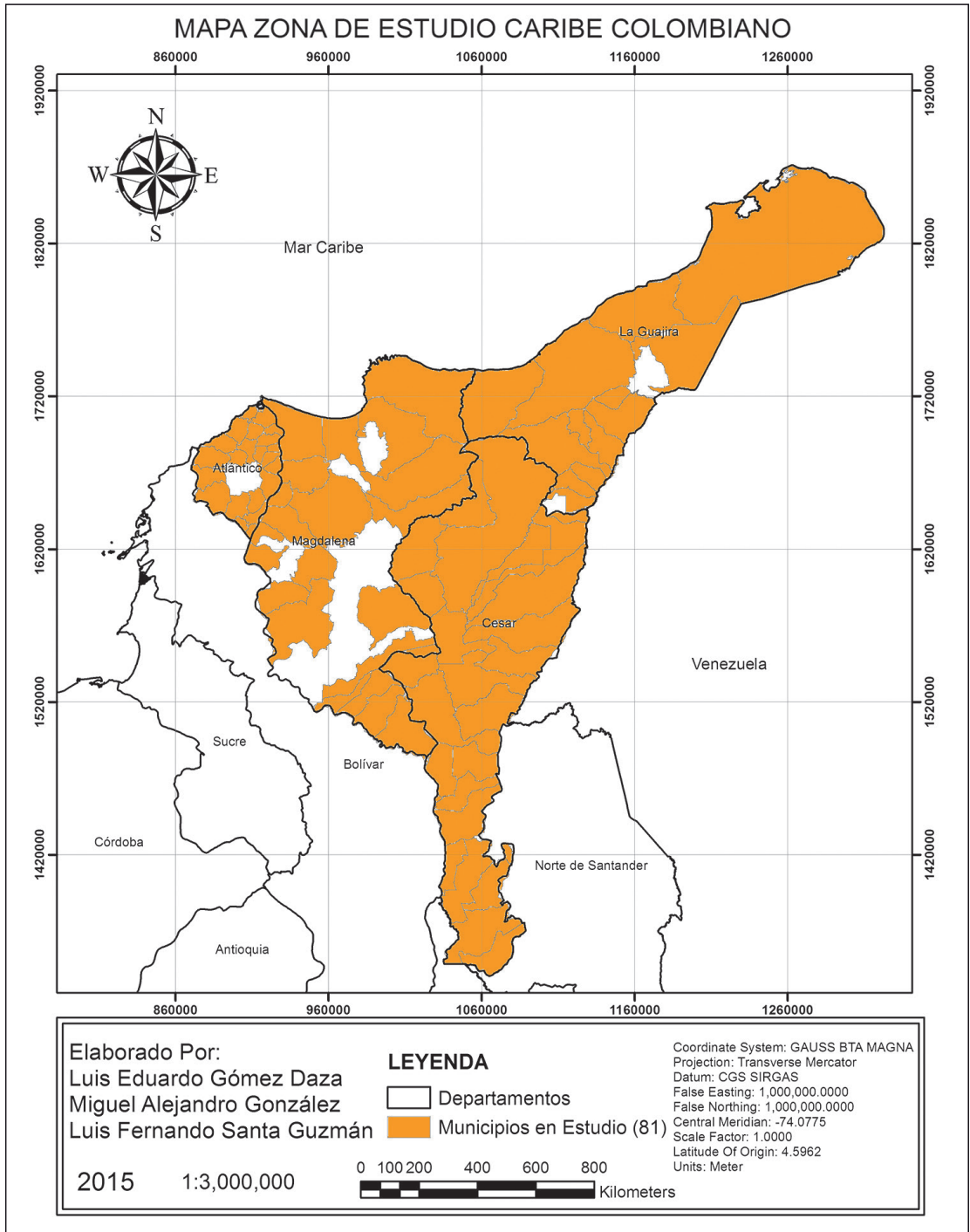


Figura 1. Delimitación de la zona de estudio en el Caribe Colombiano.

Fuente: Elaboración propia.

Para la zona de estudio se encuentra problemáticas permanentes por parte de la comunidad, en general, hacia los recursos naturales existentes, sin contar la indiferencia por parte de los entes gubernamentales. Se tiene caza y tala indiscriminada, explotación minera sin controles debidos. En cuanto saneamiento básico y residuos sólidos de esta zona, existen empresas que manejan el agua potable y alcantarillado (DNP, 2013), entre estas: la Estación de Tratamiento de Agua Potable, Sociedad de Acueducto, Alcantarillado y Aseo, KAGUA S.A. ESP; Empresa de Servicios Públicos de Agua y Aseo de Luruaco, Administración Pública Cooperada, entre otras (SUI, 2014). En municipios como Luruaco se registran residuos sólidos esparcidos en todo el casco urbano, con concentración especial en los arroyos; las aguas se derraman en pozos sépticos de las viviendas. Además el sistema de acueducto no cuenta con total potabilidad, debido a fallas técnicas en el sistema de distribución y en el de tratamiento; como consecuencia de lo anterior, se da una desaparición lenta y paulatina de las riquezas hídricas, de la fauna y flora, contaminación excesiva tanto del aire como del suelo y agua, lo cual causa proliferación

de enfermedades de todo tipo (UMATA, 2012). Por lo anterior, se hace útil y necesario la obtención del índice de escasez del recurso hídrico en el caribe colombiano, debido al gran número de problemáticas encontradas en la zona.

Metodología

Para la construcción de los modelos, se tomó como base la información sobre el índice de escasez de agua para el año 2008 (IDEAM, 2010), junto con los datos que recopilaban los valores referentes a las demás variables obtenidas a través del Sistema de Información Geográfica para la Planeación y el Ordenamiento Territorial en el mismo período de tiempo (IGAC, 2009); esto para los 81 municipios comprendidos en la zona de estudio (en MS Excel); con esto, se creó una base de datos (BD). Teniendo la BD y el archivo en formato *shapefile* de los municipios de la zona de estudio, se realizó una intersección tomando como atributo identificador el nombre del municipio; también se utilizó las coordenadas de los centros poblados de cada municipio (Sarmiento & Murcia, 2015). De tal forma, se obtuvo el *shapefile* de los municipios el cual

se encuentra con todas las variables anexas; el mismo se cargó en R y se procedió a formular, modelar y ejecutar el proceso con efectos espaciales. El software utilizado para la realización de todo el proceso fue *R versión 3.2.1* y sus correspondientes paquetes (R Core Team, 2015).

Se obtienen tres grupos de variables; las primeras son las de la distribución del recurso: Índice de escasez hídrico, Capacidad de regulación hídrica, Demanda hídrica, Oferta año seco, Oferta hídrica, Índice de vulnerabilidad hídrica,

Distancia de planta de agua al municipio; luego, las características sociodemográficas: Ciudad (variable dummy), Proporción de personas en NBI (Necesidades básicas insatisfechas), Proporción de personas en Miseria, el Componente de vivienda, el Componente de servicios públicos, el Componente de hacinamiento y el Índice de cambio de la población 2005-2008. Y, por último, la distribución del recurso en el sentido de la lluvia: Precipitación promedio en cada municipio. En la Tabla 1 se encuentra la descripción de las variables, anteriormente nombradas.

Tabla 1. Descripción de las variables de la Base de Datos.

VARIABLE	DESCRIPCIÓN
Índice de Escasez Hídrico	Relación porcentual de la demanda y la oferta de agua.
Capacidad de Regulación Hídrica	Regulación hídrica del suelo, formación superficial y perfil geopedológico.
Demanda Hídrica	Demanda cuantificada de actividades que requieren el recurso hídrico, necesidades básicas o para el desarrollo de la sociedad.
Oferta año seco	El año seco se identificó como el de menor valor anual; se tomó como base los caudales.
Oferta Media	Corresponde al valor medio multianual de series históricas de caudales representativos.
Índice Vulnerabilidad Hídrica	Disponibilidad de agua en las fuentes. Indica el grado de fragilidad del sistema hídrico para el abastecimiento y la amenaza de sequía cuando se presentan condiciones hidroclimáticas medias y extremas.
Distancia de planta de agua al municipio	Distancia de la planta de agua abastecedora de un municipio más cercana.
Ciudad	Variable que identifica las ciudades presentes en la zona de estudio con 1. Las que no lo son con 0.

VARIABLE	DESCRIPCIÓN
Proporción de personas en NBI	Es el porcentaje de personas vs. Hogares sobre la población total vs. Total de hogares que tiene, por lo menos, una necesidad básica insatisfecha (NBI).
Proporción de personas en Miseria	Es el porcentaje de personas en condiciones de miseria.
Componente de Vivienda	Indicador municipal de vivienda habitacional.
Componente de Servicios Públicos	Indicador municipal de la existencia de servicios públicos.
Componente de Hacinamiento	Indicador municipal de hacinamiento.
Tasa de cambio de la población 2005-2008	Indicador de la relación existente entre la cantidad personas que viven en un territorio y la extensión de este, del año 2005 al 2008.
Precipitación promedio en cada municipio	Cualquier forma de hidrometeoro que cae de la atmósfera y llega a la superficie terrestre. Incluye lluvia, llovizna o granizo.

Fuente: Elaboración propia.

Luego de estructurar la base de datos se procede a realizar el modelo de regresión con respuesta tipo *Beta*, para explicar la variable respuesta Índice de escasez hídrica a partir del conjunto de variables explicativas consideradas; se usa como herramienta el paquete *beta-reg* dentro del mismo software R. Para la selección del modelo y su respectiva validación se prueba las distintas funciones de enlace disponibles en el software para realizar la estimación de los parámetros por máxima verosimilitud (Ferrari & Cribari-Neto, 2004).

Los pasos metodológicos en el desarrollo del trabajo son los siguientes:

Análisis exploratorio de datos espaciales: La variable respuesta se somete a la evaluación de la presencia del efecto de la dependencia espacial. Se emplea las pruebas de Mantel, la I de Moran y la C de Geary.

Selección de modelo: Para la selección del modelo se tiene en cuenta el criterio del logaritmo de la verosimilitud (“log-likelihood”). Se implementa un proceso de selección secuencial de variables explicativas, paso a paso, en donde el criterio de exclusión o inclusión de una variable está determinado por aquella que maximice el logaritmo de la verosimilitud obtenido

en modelos de regresión simples de cada variable explicativa con la variable respuesta.

Validación: Se usa la prueba del factor de inflación de la varianza (VIF) para evaluar problemas de multicolinealidad, y pruebas para identificar la dependencia espacial (Global y Local) sobre los residuales de *Pearson* (Ferrari & Cribari-Neto, 2004).

Regresión Geográficamente Ponderada (GWR): Finalmente, para corregir la posible presencia de efectos espaciales del fenómeno, se estima un modelo GWR. Este considera así, explícitamente, el componente espacial de los datos,

incorporando en su ecuación el valor de las coordenadas geográficas de las observaciones (García Palomares, et al., 2012).

Resultados y discusión

La base de datos, finalmente, está compuesta de los tres grupos de tipos de variables definidos anteriormente, en formato *shapefile* junto con las coordenadas de los centros poblados de los municipios, con 82 filas y 20 columnas.

El índice de escasez hídrico, se sometió a las pruebas de dependencia espacial. Se empleó los respectivos métodos y pruebas como se muestra a continuación:

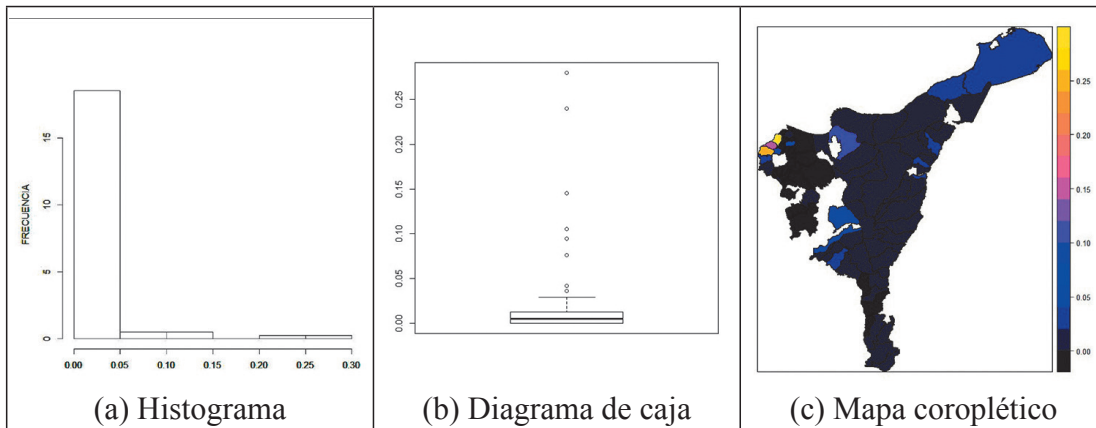


Figura 2. Análisis exploratorio del Índice de escasez hídrico.

Fuente: Elaboración propia.

En la Figura 2 se observa el histograma (a) y el diagrama de caja (b), los cuales evidencian la distri-

bución de los valores del índice de escasez hídrico; denotan una gran presencia en el intervalo 0.05; de la

misma forma, se observa que existen pocos datos que se acercan más al valor de 1. También, se observa el mapa coroplético (c), el cual muestra más detalladamente la distribución de los valores del índice de escasez en los municipios.

Se procede a realizar la selección de la matriz de pesos espaciales con el menor AIC del índice de escasez hídrico; esta selección demuestra que para este caso la mejor matriz

de contigüidad es la de 2-vecinos más cercanos ($AIC = -674.655$); con base en el procedimiento anterior, se obtiene el I de Moran (valor = 0.38 y $p_valor = 4.1 \times 10^{-06}$) y la C de Geary (valor = 0.65 y $p_valor = 0.04$) para evaluar la presencia de autocorrelación espacial global, lo que permite identificar que sí se da estructura de dependencia espacial de este tipo con respecto a la ubicación geográfica.

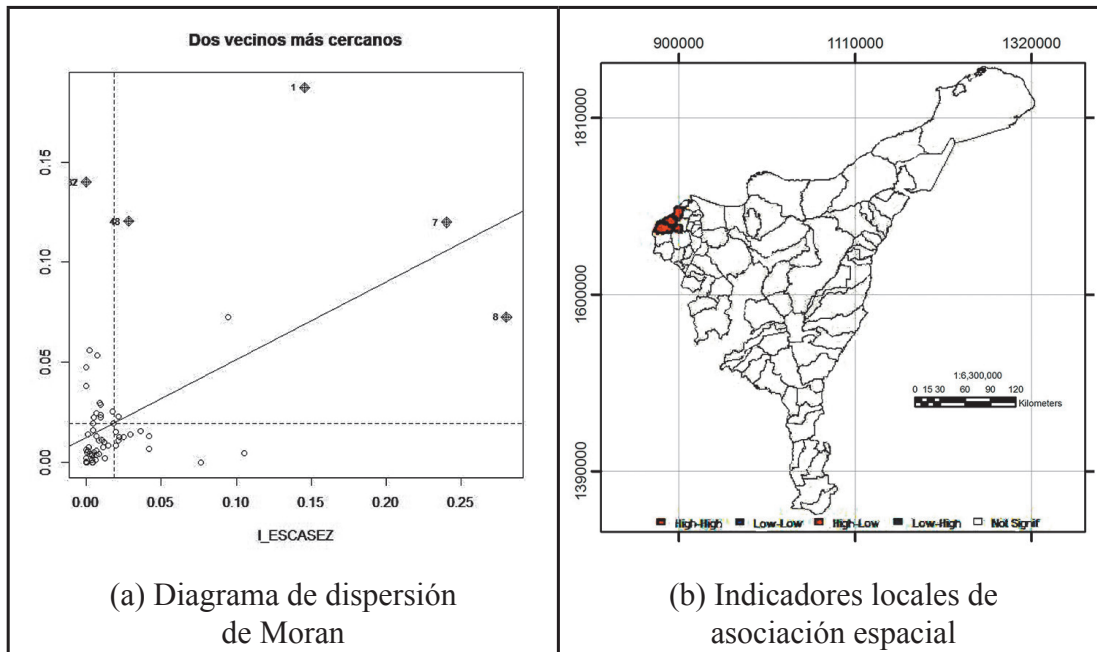


Figura 3. Estructura de dependencia espacial de tipo local para el índice de escasez hídrico.

Fuente: Elaboración propia.

Luego de realizar la estructura espacial global del índice de escasez hídrico es necesario obtener el análisis exploratorio local, el

mismo que se muestra en la Figura 3, con el diagrama de dispersión de Moran (a) y los indicadores locales de asociación espacial (LISA), los

cuales demuestran la existencia de valores atípicos, y, también, demuestran la concentración alto-alto en algunos municipios, lo cual indica que en esos municipios existe un esquema de autocorrelación de tipo local, con lo que se comprueba que el índice de escasez hídrico es alto en los municipios Piojo, Juan de Acosta, Tubara y Usiacuri, y en sus vecinos también.

Luego, tal como se indicó en la metodología, se utilizaron los enlaces *logit*, *log*, *loglog*, *cloglog*, *cauchit*, *probit*, y se seleccionó el enlace *loglog* que, según lo especificado, fue el que mejor maximizó el logaritmo de las verosimilitudes; mostró mejor pseudo R² ajustado y un pertinente indicador de p-valor, según lo que el parámetro de precisión *phi* mostró, como se puede evidenciar en la Tabla 2.

Tabla 2. Selección de la mejor función de enlace a partir de los criterios del máximo “*log-likelihood*”, R² Ajustado y p-valor para el parámetro de precisión *phi*.

Función de enlace	Log_ verosimilitud	R2 Ajustado	Estimación phi	p-valor
<i>logit</i>	368.2	0.48	35.35***	4.21 x 10 ⁻⁶
<i>log</i>	368.1	0.47	35.22***	4.34 x 10 ⁻⁶
<i>loglog</i>	368.4	0.56	39.84***	3.84 x 10 ⁻⁶
<i>cloglog</i>	367.1	0.48	35.29***	4.25 x 10 ⁻⁶
<i>probit</i>	368.2	0.51	32.26***	4.12 x 10 ⁻⁶
<i>cauchit</i>	266	0.27	34.26***	2.45 x 10 ⁻⁹

*** Parámetro significativo entre 0% y 0.1%.

** Parámetro significativo entre 0.1% y 1%.

*Parámetro significativo entre 1% y 5%.

. Parámetro significativo entre 5% y 10%.

Fuente: Elaboración propia.

Se evidenció en el parámetro *phi* el cual define un grado de precisión que entre más alta es la varianza de la estimación, la variable respuesta (índice de escasez hídrico) es menor, para este caso fue el enlace

loglog quien arrojó el mayor valor del estimador, donde el p-valor de la Prueba de Wald Parcial para coeficientes es menor a 0.05 el cual validó el nivel de significancia del coeficiente estimado.

A continuación, según el resultado de la selección paso a paso, se obtuvo que para cada variable explicativa, los valores más altos de “log-likelihood” se registraron en cinco variables: Capacidad de Regulación, Índice de Vulnerabilidad,

Componente de Vivienda, Componente de Hacinamiento y Tasa de cambio de la Población. El resumen de los parámetros estimados con el modelo seleccionado se encuentra en la Tabla 3.

Tabla 3. Resumen de parámetros estimados para el modelo resultante de la selección paso a paso acompañado de las medidas diagnóstico del modelo y de las pruebas de autocorrelación espacial

VARIABLES INDEPENDIENTES	ESTIMACIÓN	VIF	LOG_VEROSIMILITUD	R2 AJUSTADO	I. MORAN	C. GEARY
Intercepto	0.98***	NO APLICA	368.4 p_valor= 3.84x10-6	0.56	0.23 p_valor = 0.051 ***	0.65 p_valor = 0.091 ***
Capacidad de Regulación [BAJA]	0.003	1.45				
Capacidad de Regulación [MODERADA]	0.13	1.45				
Capacidad de Regulación [MUY BAJA]	-0.75***	1.45				
Índice de vulnerabilidad [BAJA]	-0.96***	1.73				
Índice de vulnerabilidad [MEDIA]	-0.67***	1.73				
Condición de vivienda	-0.003	1.51				
Condición de hacinamiento	0.01***	1.59				
Tasa de Población [2005-2008]	-0.91	1.27				

*** Parámetro significativo entre 0% y 0.1%.

** Parámetro significativo entre 0.1% y 1%.

* Parámetro significativo entre 1% y 5%.

. Parámetro significativo entre 5% y 10%.

Fuente: Elaboración propia.

La función de enlace seleccionada, reflejó que las estimaciones de parámetros es vista como las respuestas de la media para cada una de las variables explicativas (Ferrari & Cribari-Neto, 2004) lo que, en términos generales, implica que los valores negativos del estimador disminuyen el valor en la estimación del índice de escasez, mientras que valores positivos lo aumentan.

En el comportamiento del fenómeno, la identificación de variables como la 'Condición de hacinamiento' [0.01***] permitió inferir que, probablemente en los hogares con sobrepoblación, se puede presentar una alta demanda del recurso hídrico. Luego, municipios que presentan esta condición, posiblemente, van a tener problemas de escasez hídrica. Además, se evidenció que, en el caso de la 'Tasa de cambio de población' [-0.91], una estimación negativa en la particularidad del modelo, indicó que si se da aumentos de población disminuye el pronóstico del índice de escasez, algo que, aunque no va con la lógica del proceso, si se observa de manera más detallada, puede deberse a procesos migratorios por problemáticas de violencia o al mismo problema de escasez hídrica.

Por tanto, la tipificación particular de estas variables se puede considerar un punto de partida para que los administradores municipales las tengan en cuenta en el momento de hacer la gestión del riesgo y planeación del territorio y, en especial, la administración de recursos hídricos.

Para la validación del modelo se presentó la prueba del Factor de Inflación de la Varianza (VIF) para evaluar multicolinealidad entre las variables explicativas. Se demostró que las variables no presentan este problema, ya que el valor de la raíz cuadrada del VIF fue menor a 2 en todas las variables resultantes de la selección paso a paso. Ver Tabla 3. Luego de realizar la regresión y obtener la validación del modelo, se evaluó la presencia de la dependencia espacial mediante la aplicación de la estructura de tipo local con los residuales de Pearson.

En la Figura 4, se observó el comportamiento del análisis exploratorio de los residuales de Pearson; se muestra una concentración de valores negativos de los residuales en los tres diagramas; adicional a esto, se observa en el mapa coroplético (c), que tan solo algunos municipios tienen valores positivos; esto

demuestra que la concentración de datos está en valores muy bajos, y tan solo unos pocos se salen de la tendencia.

En cuanto la selección de la matriz de pesos espaciales con el menor AIC de los residuales de Pearson, se observó que se tiene un menor AIC en 1-vecino más cercano con

AIC= -31.29 en el coeficiente. Se obtuvo el I de Moran (valor = 0.23 y p_valor = 0.051) y la C de Geary (valor = 0.65 y p_valor = 0.091) para evaluar la presencia de autocorrelación espacial global, lo que permite identificar que no existe estructura de dependencia espacial de tipo global en cuanto los residuales de Pearson.

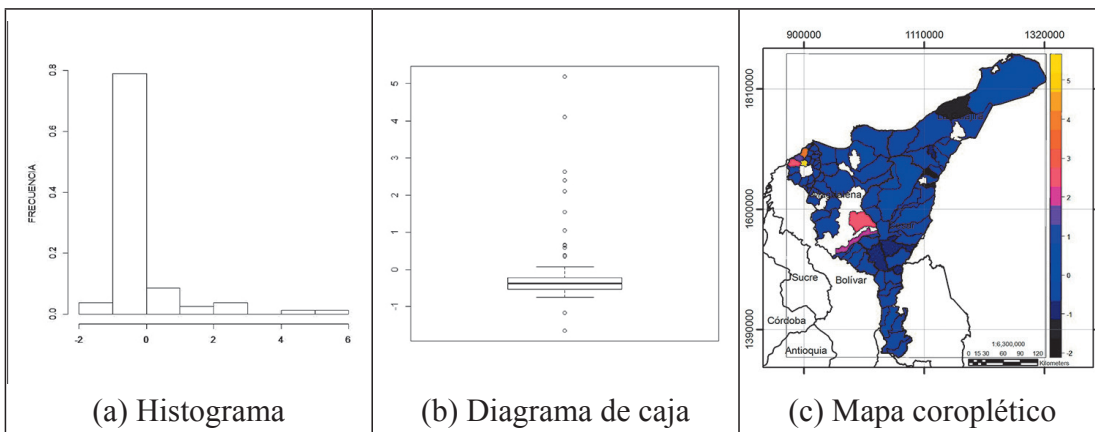


Figura 4. Análisis exploratorio de los residuales de Pearson.

Fuente: Elaboración propia.

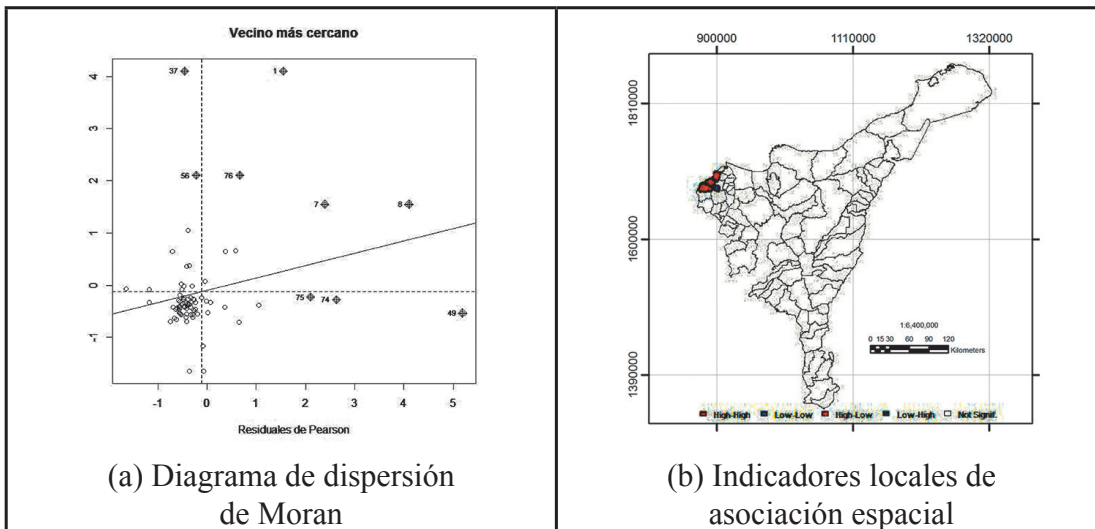


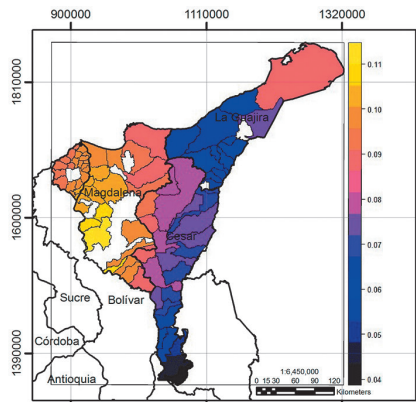
Figura 5. Estructura de dependencia espacial de tipo local para los residuales de Pearson.

Fuente: Elaboración propia.

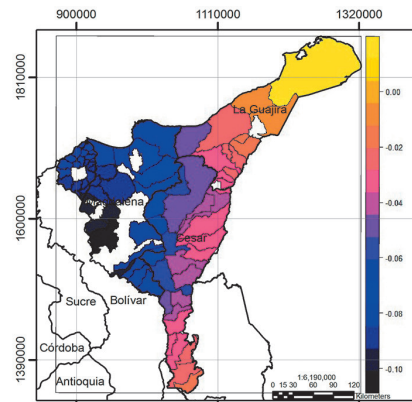
En el análisis exploratorio local de los residuales de Pearson se observa un cambio en cuanto a un municipio; ya no se observa valores de alto-alto para tal municipio; hay presencia de bajo-alto, aunque la presencia de autocorrelación de tipo local se mantiene.

Como se observa que no hay autocorrelación de tipo global, pero sí hay presencia de autocorrelación de tipo local, se dice que existe presencia de inestabilidad estructural

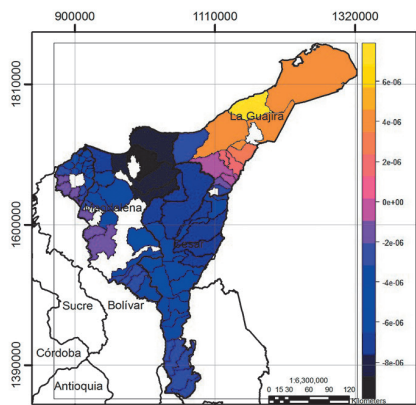
asociada a una fuerte variabilidad espacial, por lo que se procede a corregir dicho problema con la estimación de una regresión geográficamente ponderada. La implementación de esta técnica facilita la elaboración de una amplia variedad de mapas con los resultados generados: variables dependientes e independientes, R^2 locales, coeficientes locales de los predictores (elasticidades), valores t, residuos estandarizados (Mennis, 2006), como se observa a continuación:



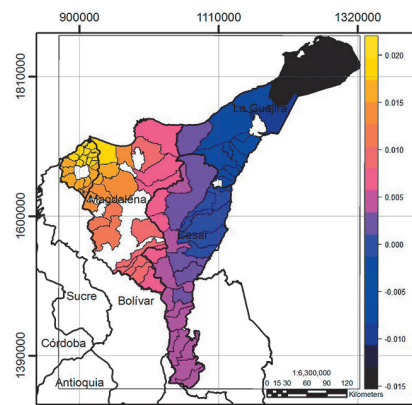
(a) Intercepto



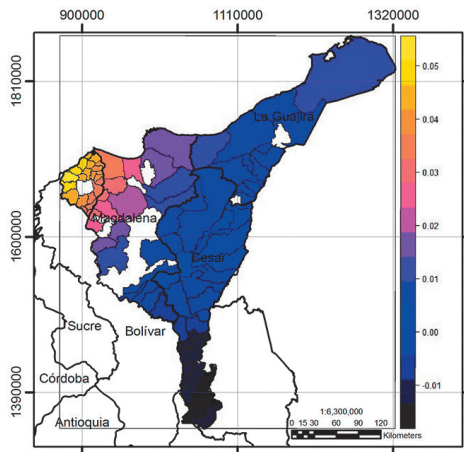
(b) Vulnerabilidad Baja



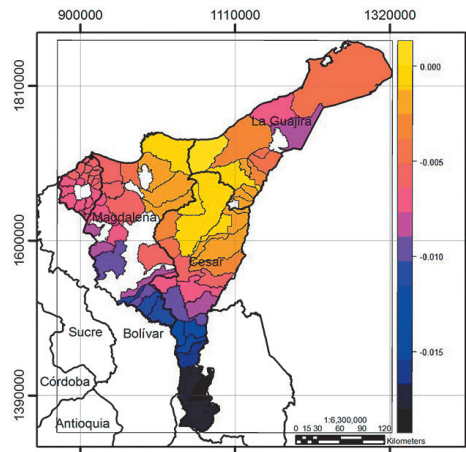
(c) Precipitación promedio



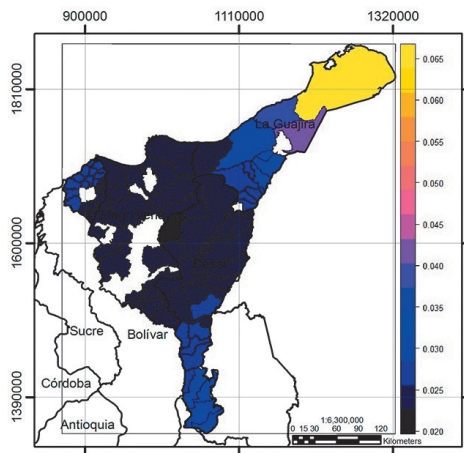
(d) Ciudad



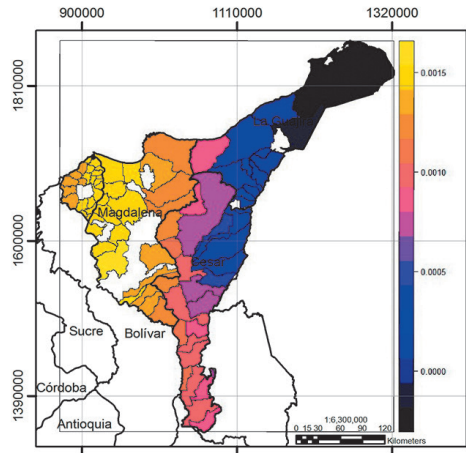
(e) Capacidad de regulación baja



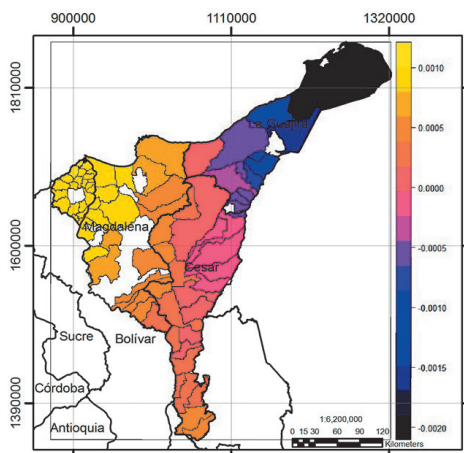
(f) Capacidad de regulación moderada



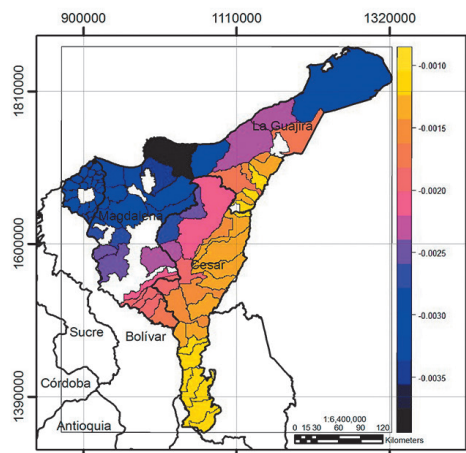
(g) Capacidad de regulación muy baja



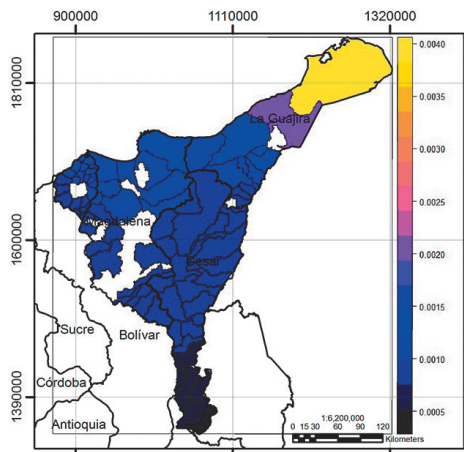
(h) Distancia a planta



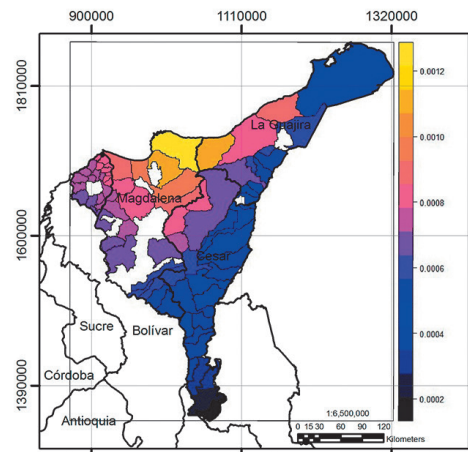
(i) Proporción de personas en NBI



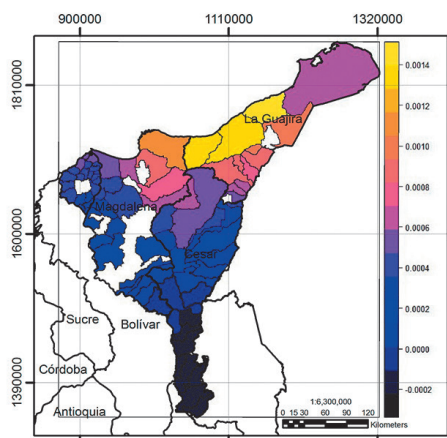
(j) Proporción de personas en Miseria



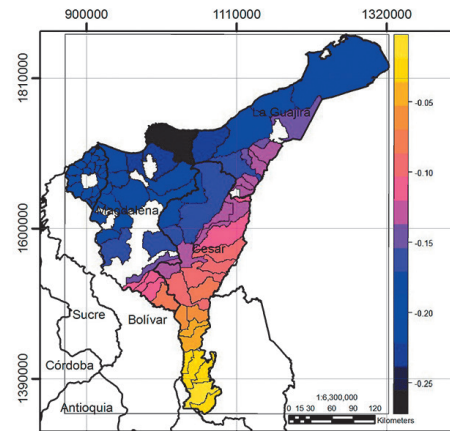
(k) Componente de vivienda



(l) Componente de servicios públicos



(m) Componente de hacinamiento



(n) Índice de cambio de la población 2005-2008

Figura 6. Comportamiento de los betas que acompañan a cada variable explicativa del índice de escasez hídrica, según la Regresión Geográficamente Ponderada.

Fuente: Elaboración propia.

Se realizó la regresión geográficamente ponderada y se obtuvieron los resultados de la Figura 6, los cuales muestran la magnitud de los parámetros del modelo en cada uno de los municipios; es decir, se realizó la desagregación del coeficiente de determinación (R^2) global en coeficientes locales y el análisis

de su distribución geográfica, lo que permitió reconocer en dónde las variables independientes tienen un mayor o peor poder explicativo (Fotheringham, et al, 2001); es decir que estas variables representan valores altos en color amarillo, y valores bajos en color azul.

El intercepto (a) es la media de la variable respuesta (índice de escasez) cuando todas las demás valen cero; en este caso, colores amarillos indican que el promedio del índice de escasez es alto respecto de los que están en azul; el color amarillo se observa en los municipios sur-occidentales; es decir, en los departamentos del Atlántico y el occidente del Magdalena, mientras que se da un registro de valores bajos en los departamentos de la Guajira y Cesar.

Los mapas del (b) al (n) representan qué tanto peso aporta la variable en la región de estudio al modelo, pues son las magnitudes de los betas del modelo; es decir, valores cercanos al color amarillo en la zona muestran que la variable tiene más peso que valores cercanos al color azul, en donde la variable no aporta tanto peso al modelo que explica el índice de escasez hídrico. Por ejemplo, encontramos que en la vulnerabilidad baja (b) y en la precipitación promedio (c) existen altos valores de coeficientes en los betas en la zona nororiental de la región, por lo que la variable tiene más peso ahí, y bajos valores en la zona suroccidental, lo que muestra que la variable no aporta tanto peso en esa zona; todo lo

contrario se encuentra en la variable ciudad (d) y en la proporción de personas en NBI (i), las cuales demuestran tener valores altos en la zona suroccidental y bajos en la zona nororiental. En la capacidad de regulación moderada (f) se encuentra que los valores altos están en todo el centro del Caribe. En la proporción de personas en miseria (j) y la tasa de cambio de la población 2005-2008 (n) se encuentran valores altos en el sur y valores altos en el norte; de modo contrario a lo que sucede en el componente de servicios públicos (l) y el componente de hacinamiento (m), los cuales tienen valores altos en el norte del Caribe y bajos en el sur del mismo.

Se encontró, usando el método de GWR, que el $R^2 = 0.88$, lo cual muestra una mejora importante en el modelo referente al GLM con respuesta Beta. La distribución espacial de los ajustes locales producidos con la GWR permite conocer la variación espacial del poder explicativo del modelo (García Palomares, et al., 2012). Por tanto, en la Figura 7 se observa las zonas del nivel de ajuste del modelo local, pues valores cercanos al color amarillo, como en el municipio de Puerto

Colombia, representan un mejor ajuste referente a colores cercanos al negro, como se observa en el municipio de Maicao.

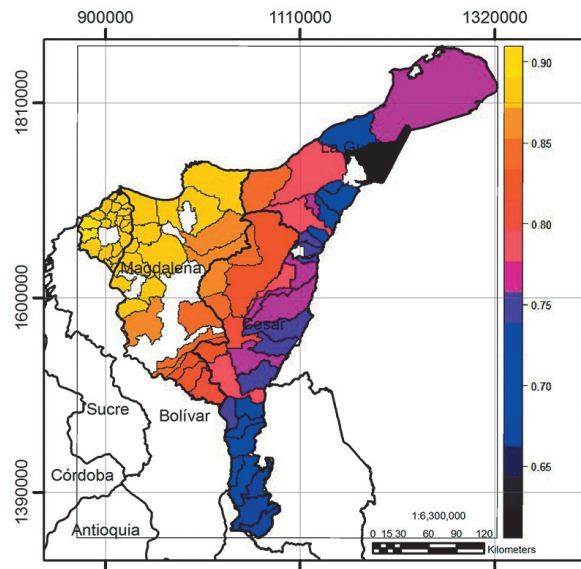


Figura 7. Magnitudes del R2 en el modelo local.

Fuente: Elaboración propia.

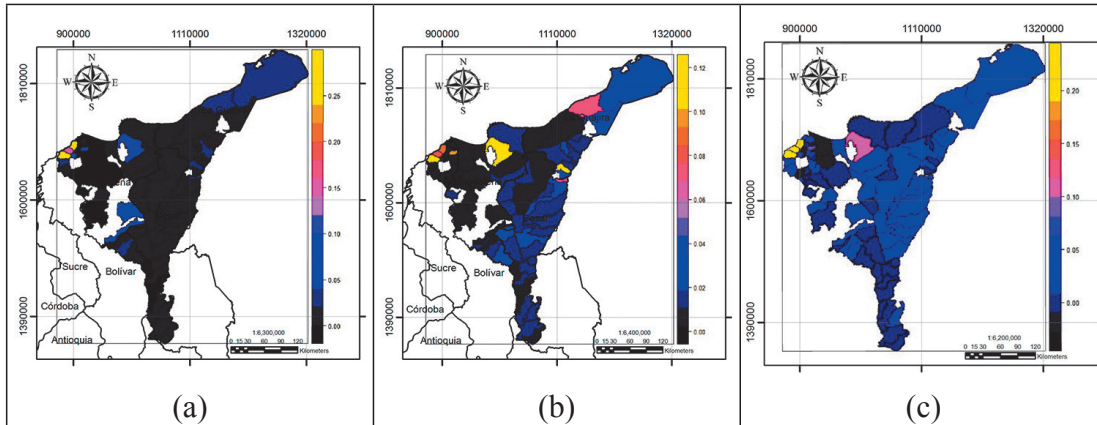


Figura 8. Mapas comparativos de los resultados. (a) Índice de escasez, (b) Modelo de regresión *Beta* y (c) Regresión Geográficamente Ponderada.

Fuente: Elaboración propia.

Finalmente, se observa en la Figura 8, la comparación de los resultados de los métodos exploratorios espaciales aplicados; tal figura indica, claramente, en cuanto resultados, que siempre existieron tres municipios en la zona sur occidental, en el departamento del Atlántico, cuyos

valores en relación con los demás resaltaron mostrando una autocorrelación de tipo local muy significativa. Estos tres municipios son: Piojo, Juan de Acosta y Tubara. Sin importar el modelo, se observa que los signos de los estimadores tienden a ser iguales, lo que indica que existe una disminución en el riesgo del fenómeno.

Conclusiones

Durante el desarrollo pragmático de lo planteado, en primera instancia, reconocer y definir variables a partir de procesos lógicos y metodologías que producen una consecución de pasos y límites definidos permitió, fundamentalmente, deducir que, según la elección de funciones de enlace para modelos con respuesta *Beta* y adicionalmente el criterio de selección paso a paso para todas las variables exógenas descritas, la escogencia de aquellas variables que mejor explican el índice de escasez, fue acorde con un sistema de validación correcto y consecuente con la lógica del proceso, dado que para este tipo de variables, la relación espacial y social definida en principio resalta el porqué del comportamiento de una situación, que genera una consecuencia sobre otra semejante.

Para los modelos de regresión y la estimación de sus parámetros, se obtuvo que el modelo de regresión *Beta*, con las variables explicativas ya seleccionadas, presentó altos niveles de significancia; que por efecto en el análisis de residuales se dejó ver incidencia de autocorrelación espacial de tipo local y global, que indujo a desarrollar el modelo de regresión geográficamente ponderado. El análisis de multicolinealidad, reveló que, para este tipo de estimaciones, la heterocedasticidad en el proceso es inherente.

Las pruebas de autocorrelación para los residuos del modelo *Beta* sometió a un proceso de incorporación de efectos espaciales por medio de regresión geográficamente ponderada, donde esencialmente para cada variable regresora existe una evaluación de los pesos representativos de cada variable sobre la estimación del modelo. Esto permitió deducir, en un nivel general, que la incorporación de efectos espaciales en modelos donde la incidencia de los mismos afecta la bondad del ajuste sobre los residuales, sugiere el uso no tan global de modelos asociados a fenómenos regionales de tipo local, como lo es el índice de escasez.

Incorporar técnicas con efectos espaciales (autocorrelación espacial y GWR) muestra claras ventajas sobre el uso de modelos estadísticos convencionales. El modelo GWR se comporta de forma más eficiente porque considera las variaciones espaciales de las relaciones entre las variables y obtiene parámetros locales (García Palomares, et al., 2012), los cuales reflejan más exactamente su influencia en la determinación del comportamiento

del índice de escasez hídrico en la costa caribe colombiana; el GWR, además de tener mejor poder explicativo que un modelo convencional profundiza en el análisis local, ya que se dispone de una ecuación de regresión para cada observación ajustada localmente, pues incluye las coordenadas de cada sitio; por tanto, es posible observar los lugares en donde el modelo tiene un mayor o menor ajuste (R^2) y así tomar buenas decisiones.

Referencias

- Alam, K. (2015). Farmers adaptation to water scarcity in drought-prone environments: A case study of Rajshahi District, Bangladesh. *Agricultural Water Management*, 148, pp. 196-206.
- Anselin, T. (1988). *Spatial econometrics: methods and models*. Springer Science & Business Media, 4.
- Branscum, A. J., Johnson, W. O., & Thurmond, M. C. (2007). Bayesian beta regression: applications to household expenditure data and genetic distance between foot-and-mouth disease viruses. *Journal of Australian & New Zealand*, pp. 287-301.
- Costa, C., Domínguez, E., Gonzalo, H., & Vanegas, R. (2005). El índice de escasez de agua ¿Un indicador de crisis o una alerta para orientar la gestión del recurso hídrico? *Revista de Ingeniería*, (22), 104-111.
- DANE – Departamento Administrativo Nacional de Estadística. (2005). *Censo general 2005*. Recuperado de: <http://www.dane.gov.co/index.php/poblacion-y-registros-vitales/censos/censo-2005>
- DNP – Departamento Nacional de Planeación. (2013). *Gestión del Sector de Planeación Nacional 2013-2014*. Recuperado de: <https://colaboracion.dnp.gov.co/CDT/DNP/Gesti%C3%B3n%20Sector%20Planeaci%C3%B3n%20Nacional%202013-2014.pdf>
- Eskelson, B. N., Madsen, L., Hagar, J. C., & Temesgen, H. (2011). Estimating riparian understory vegetation cover with beta regression and copula models. *Forest Science*, 57,(3) 212-221.

- FAO. (2013). Afrontar la escasez de agua. *Organización de las naciones unidas para la alimentación y la agricultura*. Recuperado de: <http://www.fao.org/3/a-i3015s.pdf>
- Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799-815.
- Fotheringham, S., Charlton, M., & Brundson, C. (2001). Spatial variations in school performance: a local analysis using geographically weighted regression. *Geographical & Environmental Modelling*, 5(1), 43-66.
- García, J. C., Gutiérrez, J., & Cardozo, D. (2012). Regresión Geográficamente Ponderada (GWR) y estimación de la demanda de las estaciones del Metro de Madrid. *XV Congreso Nacional de Tecnologías de la Información Geográfica*.
- Hadayeghi, A., Shalaby, A. S., & Persaud, B. N. (2010). Development of planning level transportation safety tools using geographically weighted poisson regression. *Accident Analysis and Prevention*, pp. 676-688.
- IDEAM. (2010). *Estudio Nacional del Agua*. Recuperado de: http://documentacion.ideam.gov.co/openbiblio/bvirtual/023080/ENA_2014.pdf
- IGAC – Instituto Geográfico Agustín Codazzi. (2009). *Sistema de Información Geográfica para la Planeación y el Ordenamiento Territorial (Infraestructura Colombiana de Datos Espaciales)*. Recuperado de: <http://sigotn.igac.gov.co/sigotn/>
- Jaramillo-Mosqueira, L. A. (2005). Evaluación econométrica de la demanda de agua de uso residencial en México. *El Trimestre Económico*, LXXII(2), 367-390.
- Jun, K. S., Chung, E., Sung, J. Y., & Lee, K. S. (2011). Development of spatial water resources vulnerability index considering climate change impacts. *Science of the total environment*, 409(24), 5228–5242.
- Kahil, M. T., Dinar, A., & Albiac, J. (2015). Modeling water scarcity and droughts for policy adaptation to climate change in arid and semiarid regions. *Journal of Hydrology*, 522, 95-109.
- Knoche, M., Fischer, C., Pohl, E., Krause, P., & Merz, R. (2014). Combined uncertainty of hydrological model complexity and satellite – based forcing data evaluated in two data-scarce semi-arid catchments in ethiopia. *Journal of hidrology*, 519, 2049–2066.
- Lewis-Beck, M., Bryman, A., & Liao, T. (2003). *The SAGE Encyclopedia of Social Science Research Methods*. SAGE Publications.
- Lloyd, C. D., & Shuttleworth, I. (2005). Analysing commuting using local regression techniques: scale, sensitivity, and geographical patterning. *Environment and Planning*, 37(1), pp. 81-103.

- Lvovitch, M. (1986). *Water and life*. Moscow: Pub House “Mysl”.
- Mennis, J. (2006). Mapping the results of geographically weighted regression. *Cartographic Journal*, 43, pp. 171-179.
- Moreno, S., & Estrada, L. (2013). Análisis Espacial de la Pobreza Multidimensional en Colombia a Partir del Censo de Población de 2005. *CANDANE, Centro Andino de altos Estudios*. Recuperado de: http://www.dane.gov.co/candane_2014/files/Analisis%20Espacial%20de%20Pobreza.pdf
- Plan de Acción. (2012). Síntesis Ambiental II. *Desarrollo con Sostenibilidad Ambiental*.
- Portero, I. S., & Agulló, F. A. (2002). Agua y producción agrícola: un análisis econométrico del caso de Murcia. *Revista española de estudios agrosociales y pesqueros*, (197), 129-158.
- R Core Team. (2015). R: The R Project for Statistical Computing. *R versión 3.2.1*.
- Romero, H. A., & Ortiz, L. F. (2008). Ajuste metodológico al índice de escasez de agua propuesto por el IDEAM en el plan de ordenación y manejo de la cuenca del río Pamplonita, Norte de Santander, Colombia. *Revista Colombia Foresta*, 11, 165-174.
- Sarmiento & Murcia (2015). *Modelos bayesianos para el comportamiento del cáncer gástrico en Colombia para el período 2005-2012* (Tesis de pregrado). Universidad Distrital Francisco José de Caldas.
- SUI, S. Ú. (2014). Registros del Sistema único de Información de Servicios Públicos (SUI) del Gobierno Nacional. Recuperado de: <http://www.sui.gov.co/SUIAuth/portada.jsp?servicioPortada=1>
- UMATA, U. (2012). Informe de Gestión Ambiental 2012. Luruaco, Atlántico, Colombia.
- UNESCO. (2015). *The United Nations World Water Development Report 2015: Water for a Sustainable World*. United Nations Educational, Scientific and Cultural Organization.