

**ANÁLISIS ESPACIAL DEL PRECIO DEL SUELO CON
MODELOS DE REGRESION LINEAL MÚLTIPLE (MRLM) Y
SISTEMAS DE INFORMACIÓN GEOGRÁFICA (SIG),
RESISTENCIA (ARGENTINA)**

Romina Ortiz*, Federico Arias, Cristian Da Silva*** y Osvaldo Cardozo******

*Laboratorio de Tecnologías de la Información Geográfica (LabTIG). UNNE-CONICET (Argentina).

r.g.o_94@live.com.ar

**Laboratorio de Tecnologías de la Información Geográfica (LabTIG). UNNE-CONICET (Argentina).

fedearias-@hotmail.com

***Laboratorio de Tecnologías de la Información Geográfica (LabTIG). UNNE-CONICET (Argentina).

cjdasilva@hum.unne.edu.ar

**** Laboratorio de Tecnologías de la Información Geográfica (LabTIG). UNNE-CONICET (Argentina).

odcardozo@hum.unne.edu.ar

RESUMEN

En el presente trabajo aplicamos un modelo de regresión lineal múltiple (MRLM) con el propósito de analizar la relación entre el precio fiscal del suelo a nivel de barrio con variables socio-económicas y geográficas. Todas las variables fueron normalizadas en un entorno SIG y los test estadísticos estimados con paquetes estadísticos.

El modelo de regresión obtenido muestra la importancia de variables como Densidad de hogares de nivel socioeconómico alto, distancia al centroide comercial y distancia a espacios verdes a la hora de explicar el precio fiscal del suelo. El modelo es estadísticamente significativo (F de 93,436 y un p-valor de 0,0000), y presenta un R2 Ajustado que explica aproximadamente un 67 % de la variabilidad del precio fiscal del suelo por medio de las tres variables incluidas en el modelo.

Palabras clave: Precio Fiscal del Suelo, Modelo de Regresión Múltiple, Sistemas de Información Geográfica.

ABSTRACT

This paper analysed the relationship between land price and socio-economics and geographical variables. For this, we have applied multiple lineal regression model (MLRM) to analyse the variables based on statistics software. The GIS was used for normalized the variables.

The results show the importance of variables such as density of high socioeconomic status households, distance to commercial center and distance to green places in order to explain the land price. The model is statistical significant (F de 93.436 and p-value of 0.0000) and the R2 Adj. explains the 67% of the land price variability.

Key words: Land Price, Multiple Regression Model, Geographic Information Systems.

INTRODUCCIÓN

El rubro inmobiliario es una de las actividades más dinámicas de la economía, especialmente en las áreas urbana donde actúa como promotor del desarrollo urbano y desempeña un importante rol en la fijación del precio del suelo, por lo tanto, en la cantidad de suelo disponible para el crecimiento espacial de la ciudad. Pese a ello, su estudio ha despertado poco interés en el ámbito científico-académico, prueba de esto es que los análisis sistemáticos y objetivos sobre el tema son recientes.

Lo cierto es que conocer en detalle los factores influyentes en la formación del precio del suelo es de suma utilidad, y como lo expone Lever (2009) tienen una amplia gama de aplicaciones. Por su parte, Cano y Chica (2004) lo reconocen como fundamental en el negocio hipotecario ya que, por un lado el monto del préstamo se calcula como un porcentaje del valor de tasación, y por otro, garantiza a la entidad financiera la recuperación del préstamo en caso de impago.

Un aspecto poco considerado pero no menos importante es el rol desempeñado por el Estado, como elemento de equilibrio en el libre mercado del suelo. En este sentido, la rentabilidad que ha logrado el negocio inmobiliario en la última década, plantea un panorama especulativo considerable donde acceder a un inmueble es dificultoso en cualquier país, independientemente del grado de desarrollo. La consecuencia de una falta de interés estatal en políticas públicas de este tipo, nos recuerda como la “burbuja inmobiliaria” de España que estalló en 2008 y provocó importantes desordenes económicos a nivel europeo y mundial.

Enmarcar estos estudios dentro de un campo disciplinar es dificultoso y arriesgado, porque resulta claro advertir que el precio del suelo depende de un conjunto con variables socio-demográfica, económicas y geográficas. Las variaciones que afectan el precio pueden provenir tanto por las inversiones realizadas en el predio, normalmente a cargo del propietario, como por las condiciones del entorno y que por lo general incumben al Estado o las empresas promotoras; de esta última deriva lo que se conoce como precio fiscal del suelo.

Otro rasgo característico en mercados complejos como el inmobiliario, es la aparición de “precios sombra” que surgen de cada atributo del bien. Su conocimiento cobra gran interés por la alta heterogeneidad de dichos atributos, lo que puede

significar un alto valor relativo de los bienes. En este sentido, la teoría de precios hedónicos constituye un significativo avance metodológico para la modelación de mercados implícitos por atributos, mediante la aplicación de técnicas econométricas es capaz de medir el valor de un bien compuesto y analizar la forma en que se “mezclan” sus atributos (Lever, 2009).

ANTECEDENTES

La revisión literaria acerca del valor de la tierra en términos generales, permitió identificar una diversidad de enfoques y perspectivas de análisis aplicadas debido a sus múltiples implicancias. Así, desde el punto de vista económico, el clásico modelo agrario de Von Thünen de 1826 propone zonificar el espacio en torno a franjas de distancia para construir un esquema de usos del suelo y su correspondiente valor. Asimismo, Hurd en 1903 planteaba una relación entre valor del suelo urbano–renta y renta–localización, con lo cual se evidencia la estrecha relación entre variables socioeconómicas y geográficas en la valoración del suelo (Humarán, 2010).

Por otro lado, la dimensión social del precio del suelo es la menos atendida por la comunidad científica, aunque no por ello sea menos importante. Anderson (1962) realiza un estudio del valor del suelo, encontrando relaciones importantes entre esta y la renta, distribuido de manera diferente en el territorio, concluyendo que el valor se explica claramente en función de la estructura socio-económica del lugar. Asimismo, Harvey (1993) en sus estudios relacionados con las tendencias del urbanismo y desigualdades sociales, manifiesta que la problemática tiene amplia difusión espacial.

Más recientemente en el contexto de la planificación del crecimiento espacial urbano, el precio del suelo cobra una gran trascendencia debido a que actúa como variable que limita el crecimiento del uso residencial y aumenta por ello los problemas relacionados con el acceso a la vivienda. No obstante, muchos autores han considerado el precio del suelo conjuntamente con otras variables en sus análisis de crecimiento óptimo de ciudades, con un amplio uso de Sistemas de Información Geográfica (SIG) y Evaluación Multicriterio (EMC) como metodología de normalización, análisis y modelización (Barredo Cano, 1996; Wu & Webster, 1998; Barredo Cano y Bosque Sendra, 1999; Jiang & Eastman, 2000; Malczewski, 2002).

Pero sin dudas, la teoría de precios hedónicos es la aproximación más utilizada. Este enfoque del problema se remonta a mediados de los 60’ en Estados Unidos cuando Ridker y Henning analizaron el impacto de un conjunto de variables

ambientales en la conformación del predio de las casas en Saint Louis (Lever, 2009). La revisión de la literatura sobre todo latinoamericana permite encontrar cierta diversidad en las aplicaciones, siendo el factor común en todos ellos la presentación de ecuaciones hedónicas (Vecchione, 1989; Nuñez y Schovelin, 2002; Meloni y Ruiz, 2002; Cano y Chica, 2004; Núñez y Roca, 2007; Lever, 2009).

Básicamente, la técnica de precios hedónicos consiste en estimar económicamente ecuaciones que tienen como variable dependiente el precio del bien o servicio y, como regresoras a los atributos o características del mismo (Meloni y Ruiz, 2002). En Lever (2009) se presenta una interesante propuesta de clasificación de variables para una ecuación hedónica del tipo $P = f(I, V, U, Z, S, E, w)$ donde:

P = Precio del suelo x m².

I = Inmueble: superficie construida, superficie del terreno, aptitud de usos del suelo, aspectos arquitectónicos y de diseño interior, equipamiento interior, número de habitaciones, calidad de los materiales.

V = Vecindario: nivel socioeconómico, tipo de residentes, seguridad.

U = Ubicación: área residencial/industrial, distancia geográfica y accesibilidad a los centros de importancia.

Z = Zonning (características determinadas por la ubicación del inmueble en el plano regulador de la ciudad): densidad de construcción, tipos de actividades, usos del suelo permitidos.

S = Servicios, equipamiento exterior e infraestructura: agua, alcantarillado, electricidad, pavimento.

E = Externalidades (existentes en el entorno de la propiedad): actividades contaminantes, vertederos, áreas verdes.

w = conjunto de parámetros que acompañan a cada atributo y que constituyen los precios implícitos o "sombra" de cada atributo del inmueble.

En el área local son escasos los trabajos que analicen los usos del suelo. En este contexto, uno de los primeros trabajos fue un amplio diagnóstico del equipamiento e infraestructura en el Gran Resistencia, donde una de las variables analizadas fue el uso del suelo urbano (Scornik 1998), o el estudio integrado de los usos del suelo en el casco céntrico en relación a otros aspectos del entorno urbano, implementando una metodología de relevamiento in situ y una propuesta de clasificación (Lucca y Rey, 1999), mientras que Da Silva et. al, (2013) analizan distribución espacial de los usos del suelo para el casco céntrico de la ciudad, corroborando que el uso comercial muestra indicios de concentración en inmediaciones a la plaza central, y donde más del 40 % de las personas que compran utilizan transporte público -frente a un 10 % que lo hace en automóvil-, lo cual sugiere una buena correspondencia entre la población usuaria del transporte y la actividad comercial, pero en ninguno de los casos anteriores se analiza la cuestión del valor del suelo.

La excepción constituye un reciente trabajo de Arias et. al. (2013), en el que explora la relación existente entre el precio inmobiliario del suelo en conjunto con la distribución de los usos del suelo asociados al transporte.

Sobre la base de estos antecedentes, se busca básicamente probar una metodología objetiva para medir las valoraciones inmobiliarias, específicamente se busca identificar el peso o incidencia de algunas variables socio-ambientales en la conformación del precio fiscal en la ciudad de Resistencia.

En base a lo expresado se han formulado los siguientes objetivos específicos:

- Caracterizar la distribución espacial del fenómeno en estudio.
- Obtener un modelo explicativo de la variación del precio del suelo.

La hipótesis de trabajo (H_1) sostiene que las variables ambientales o del entorno inciden significativamente en el precio del suelo a nivel de barrio, y se esperan los precios más elevados en el centro de la ciudad. Esta afirmación se apoya en el supuesto de que, el área de mayor cobertura de servicios y accesibilidad corresponde al centro de la ciudad.

Entonces, se buscará contrastar $H_0: \beta_i = 0$ versus $H_1: \beta_i \neq 0$; es decir al menos unos de los coeficientes β_i estimados debe ser distinto de 0.

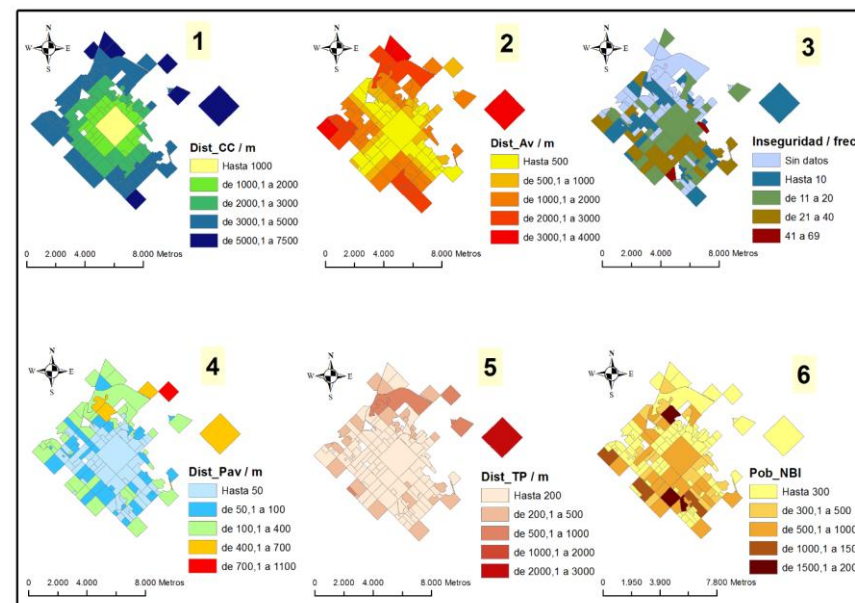
MATERIALES Y MÉTODOS

Los datos

La base cartográfica de barrios, red vial, límites políticos y el precio fiscal del suelo provienen de la Dirección de Catastro y Cartografía de la Municipalidad de Resistencia, en formato ESRI shapefile actualizada al año 2013. Esta constituye la variable dependiente y representa el valor en pesos por m².

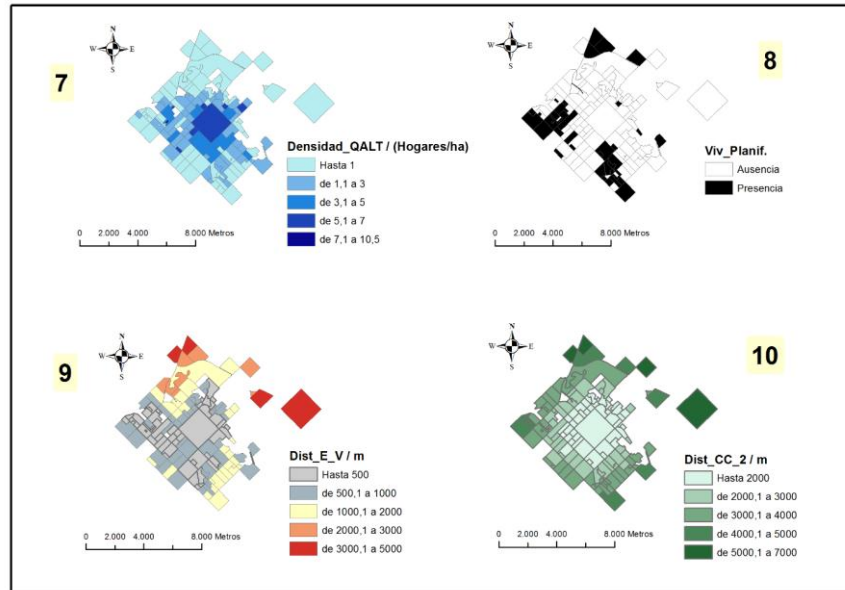
Las variables poblacionales se han extraído del Censo Nacional de Población, Hogares y Viviendas 2001 a nivel de Radio y Fracción censal, los cuales fueron normalizados a nivel de Barrio (Población con NBI). Además, se ha empleado datos de niveles socioeconómicos (Estrato Socioeconómico Alto) normalizados y calculados a partir de datos generados en el marco del proyecto *Georreferenciación de niveles socioeconómicos a nivel de manzana en el Gran Resistencia y la ciudad de Corrientes* (Odrozola y Falcón, 2007). Otra variable que se incluyó en los análisis debido a que fue recogida en varios trabajos de referencia, fue el nivel de inseguridad a nivel de barrio (Pyszczek, 2012).

Figura 1: Distribución de las Variables Empleadas. Parte 1 (1 - Distancia al Centroides Comercial; 2 - Distancia a las Av. Principales; 3 - Inseguridad; 4 - Distancia al Pavimento; 5 - Distancia a Paradas del Transporte Público; 6 - Población con NBI).



[Fuente: Elaboración Propia]

Figura 2: Distribución de las Variables Empleadas. Parte 2 (7 - Densidad de Estrato Socioeconómico Alto; 8 - Presencia de Viviendas Planificadas; 9 - Distancia a Espacios Verdes; 10 - Distancia a Calles Comerciales).



[Fuente: Elaboración Propia]

Por último, dado que los datos están georreferenciados se han calculado por medio de herramientas SIG las siguientes variables: Densidad de Estrato Socioeconómico Alto, Distancia al Centroide Comercial, Distancia a Calles Comerciales, Distancia a las Av. Principales, Distancia al Pavimento, Distancia a Paradas del Transporte Público, Presencia de Viviendas Planificadas y Distancia a Espacios Verdes (Ver Figura 1 y Figura 2).

Modelo de Regresión Lineal Múltiple (MRLM)

Una definición clásica del análisis regresión acuñada por Francis Galton la define como una técnica estadística que permite predecir el valor promedio de una variable, tomando como base los valores conocida de otra/s.

Asimismo, en Estadística son muchas las maneras de nombrar a las variables empleadas, en este caso adoptamos la terminología de Variables dependientes y Variables independientes. A su vez, existe una diversidad de expresiones utilizadas para representar la relación lineal entre las variable/s dependiente e independiente/s, por ello empleamos la notación comúnmente utilizada, quedando definida como:

Dónde: y es la variable dependiente; β_0 , el punto de intersección en el eje de ordenadas; $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ constantes numéricas que se deben estimar utilizando para ello los datos disponibles: $x_1, x_2, x_3, \dots, x_k$ son variables independientes utilizadas para realizar las estimaciones; ϵ es el termino aleatorio o error.

Emplear modelos de inferencia estadística habitualmente está destinado a obtener buenos ajustes en la explicación del término dependiente, sin embargo esto último no constituye todo lo necesario, más bien, se deben garantizar niveles aceptables de explicación y sobre todo abstenerse de violar los supuestos de comportamiento -homocedasticidad, linealidad, no autocorrelación, normalidad, no colinealidad, entre otros- (Da Silva et. al., 2014).

Supuestos en los que se apoya el modelo de regresión

Empleados tanto para los modelos de regresión lineal simple y múltiple, la consideración de estos a la hora de llevar a cabo el análisis, permite obtener interpretaciones de los resultados de manera adecuada, a lo que se adiciona la ventaja de ampliar las conclusiones obtenidas (Estebanez Alvarez y Bradshaw, 1978). Por ello, se describen sintéticamente los supuestos básicos del modelo de regresión:

- Linealidad: establece que la relación entre la variable dependiente y la/s independiente/s debe ser lineal, debido a que de lo contrario el modelo puede llevar a resultados erróneos. El incumplimiento del supuesto de linealidad suele denominarse error de especificación. Algunos ejemplos son: omisión de variables independientes importantes, inclusión de variables independientes irrelevantes, no linealidad (la relación entre las variables independientes y dependientes no es lineal), parámetros cambiantes (los parámetros no permanecen constantes durante el tiempo que duran la recogida de datos), no aditividad (el efecto de alguna variable independiente es sensible a los niveles de alguna otra variable independiente).

- Ausencia de errores de medición: establece que las variables independientes son consideradas fijas y no sujetas a errores de medición, debido a que no existen ningún "test" simple para medir el error. En el caso de la variable dependiente, puede tomarse una muestra y por lo tanto, puede estar sujeta a errores de medición, lo cual en cuyo caso, los errores forman una distribución normal.

- Normalidad de los residuales: para cada valor de la variable independiente (o combinación de valores de las variables independiente), los residuos se distribuyen normalmente. En otras palabras, si el error de medida es el único factor que produce variación en los valores de y (para cualquier valor de x), entonces las variaciones formaran una distribución normal.

- La Media de los residuales debe de ser igual a 0: este supuesto implica que la media de los residuales se encuentra en la línea de regresión. El cumplimiento de este supuesto establece que la línea de regresión es la forma funcional y proporciona un buen ajuste a los datos estudiados.

- Homocedasticidad: establece que la varianza de las distribuciones deben tener todos los mismos valores. La presencia de heterocedasticidad puede influir en los test que se apliquen al modelo, y por ello producir error en las estimaciones.

- No Colinealidad entre las variables independientes: establece que no existe relación lineal entre las variables independientes. El incumplimiento de este supuesto da origen a colinealidad o multicolinealidad. Además, la violación de este supuesto puede dificultar la evaluación del peso de las variables independientes en la explicación de la variación de la variable dependiente.

- El número de observaciones ha de ser mayor que el número de variables independientes: este supuesto establece que se deben contar con un número de datos (n) mayor que la cantidad de variables independientes empleadas.

Además, el supuesto de No autocorrelación es otro de los considerados al momento de aplicar un modelo de regresión lineal, pero ello merece una atención especial, debido a que estamos empleando en el análisis variables geográficas, lo cual comentamos en la próxima sección.

Autocorrelación

Este supuesto establece que los residuales no deben presentar autocorrelación temporal ni espacial. En el primer caso, los residuales no se dispersan aleatoriamente en torno a la recta de regresión en tanto que en el segundo, manifiestan agrupamientos definidos por proximidad con valores positivos y/o negativos. La importancia de evaluar la autocorrelación radica en que, de ser positiva su detección no es recomendable emplear la ecuación lineal, y sería conveniente considerar el empleo de alguna forma curvilínea que mejore el ajuste de los datos.

Contraste de Durbin – Watson (DW)

Es una de las pruebas mayormente empleadas para detectar presencia de autocorrelación serial en los modelos de regresión. Este test se basa en el contraste de la hipótesis de no autocorrelación en los residuales, lo cual es caso de no suceder, la presencia de autocorrelación en los residuos es la alternativa vigente.

El Estadístico se define como:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

A partir de este se puede interpretar que:

- Si hay autocorrelación positiva las diferencias entre los residuos toma un valor muy próximo a 0.
- Por el contrario, si existe autocorrelación negativa tomará un valor muy próximo a 4.
- En caso de no existencia de autocorrelación, el DW tomará un valor muy próximo a 2.

Asimismo, se debe calcular los valores críticos para realizar el contraste debido a que la distribución varía según las variables empleadas, y por lo tanto los límites críticos varían. Calcular estos límites dL (Límite inferior) y dU (Límite superior) permite evaluar si el contraste es concluyente o no, por ello se espera que el DW calculado no caiga entre las dos cotas mencionadas (Gallego Gómez, 2008).

Autocorrelación Espacial

El principio de la autocorrelación espacial se basa en el análisis simultáneo de las localizaciones (distancia entre observaciones) y los valores (atributos asociados) que adquiere una variable en el territorio. Es decir, evalúa si el patrón de distribución espacial muestra signos de estar agrupado, disperso o simplemente es aleatorio (de Smith et. al, 2013).

Ante la utilización de variables geográficas en modelos de regresión lineal, la autocorrelación espacial se presenta como uno de los supuestos más importantes a evaluar. La necesidad de expresar el patrón de distribución existente en los residuales resultantes permite caracterizar el modelo en términos de bondad y por lo tanto, especificar áreas donde el comportamiento espacial de las variables tiende a ser la principal fuente de problemas.

Test de autocorrelación espacial Global (I de Moran)

Este test evalúa el patrón que presenta la variable analizada en todo el área de estudio, pudiendo resultar tres situaciones: agrupada, dispersa o aleatoria.

La ventaja de utilizar este tipo de test estadístico se sustenta en la incorporación de una matriz de pesos espaciales calculada a partir de algún tipo de conceptualización de relación espacial. Esto permite evaluar la distribución de la variable analizada en virtud de su localización geográfica y el entorno.

Formalmente, el test queda definido como:

$$I = \frac{n}{s} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2}$$

Dónde: n es el número de observaciones; w_{ij} es una matriz de pesos espaciales (calculada en función a la distancia); s es la suma de todos los elementos de la matriz w; z_i y z_j son valores estandarizados de la variable analizada en i y j.

Este test estadístico opera con el propósito de contrastar las hipótesis que sostienen:

- H_0 : la distribución de las observaciones es producto de procesos aleatorios.
- H_1 : la distribución de las observaciones presenta algún grado de asociación espacial.

Calculado el I de Moran, se obtiene el valor esperado del índice E (I) suponiendo que no existiera autocorrelación espacial. Además, se calcula una puntuación z y un p-valor como medidas que indican el rechazo o no de la hipótesis nula. Cuando la puntuación z y el p-valor indican que el I de Moran es estadísticamente significativo, es decir, puntuaciones z muy altas (positivas) o muy bajas (negativas) asociadas con p-valores pequeños, se encuentran en las colas de la distribución normal e indican que es poco probable que el patrón espacial observado sea resultado de procesos aleatorios, por lo tanto, se puede rechazar la hipótesis nula (Da Silva et. al., 2013).

Local Indicators of Spatial Association (LISA)

Anselin (1995) caracterizó de dos maneras al LISA, en primera instancia puede ser utilizado para detectar “hot spot” o “puntos calientes” lo cual permite ver la asociación espacial de la variable analizada, y por otro lado, permite identificar “outliers” en estudios de regresión.

Considerada como versión local del índice I de Moran, mide el grado de autocorrelación espacial en el entorno próximo de cada observación. Pese a que todas las observaciones tienen una influencia sobre las demás observaciones, después de un umbral de distancia la influencia puede ser desestimada, por lo tanto, en la matriz w se requiere imponer una estructura que limite el número de vecinos a tener en cuenta. Formalmente, la ecuación se expresa como:

$$I_i = z_i \sum_j w_{ij} z_j$$

Dónde: I_i es el índice de Moran en la observación i ; w_{ij} es una matriz de pesos espaciales (calculada en función a la distancia); z_i y z_j son valores estandarizados de la variable analizada en las localizaciones i y j .

Al igual que en la interpretación del I de Moran global, un valor alto de LISA significa una agrupación de valores similares (tanto altos como bajos), mientras que un valor bajo de LISA indica una agrupación de valores diferentes. Asimismo, la utilidad de este indicador radica en la capacidad de identificar la localización espacial de los valores atípicos en los residuales y su comportamiento en relación al entorno.

Aplicación de un Modelo de Regresión Lineal Múltiple (MRLM)

Una vez concebido todos los supuestos y normalizadas las variables a emplear, se procedió a evaluar las variables independientes a incluir en el modelo de regresión propuesto. Para ello se ha empleado el método denominado “paso a paso” (stepwise) que permite elegir el subconjunto de variables regresoras con mejor performance para el modelo. La lógica procedimental del paso a paso está dada por introducir las variables independientes que no se encuentren en la ecuación y que tengan la probabilidad F más pequeña. Por otro lado, las variables ya introducidas en la ecuación se eliminan si su probabilidad F llega a ser suficientemente grande, concluyendo el método de selección cuando ya no existan variables candidatas a ser incluidas o eliminadas. Una vez concluido con la selección de las variables, se debe optar por la utilización de medidas de bondad de ajuste y verificar las hipótesis estructurales para seleccionar el modelo de regresión más adecuado, de modo tal que los resultados obtenidos sean consistentes y estadísticamente significativos.

Los datos fueron transformados con el propósito de lograr en ellos un comportamiento lineal con respecto a la variable dependiente. Asimismo, los datos obtenidos y calculados mediante SIG son descriptos en la siguiente tabla:

Tabla 1: Descripción de las variables utilizadas en el modelo de regresión.

Tipo	Denominación	Descripción	Unidad de Medida
Dependiente	PRECIO_SUE	Precio fiscal del suelo definido por la municipalidad de Resistencia.	\$/m ²
Independiente	DENSIDAD_QALT	Densidad de hogares con estrato socioeconómico alto calculada en el barrio.	Hogares/ha
	DIST_AV	Indica la distancia euclidiana a las avenidas principales (9 de Julio, 25 de Mayo, Sarmiento y Alberdi).	m
	INSEGURIDAD	Frecuencia de opinión acerca de la estigmatización como barrio inseguro	Frecuencia
	DIST_PAV	Representa la distancia euclidiana a la red vial.	m
	DIST_TP	Indica la distancia euclidiana a las paradas de Transporte Público.	m
	POB_NBI	Población con Necesidades Básicas Insatisfechas.	Total por unidad administrativa
	VIV_PLANIF	Presencia de viviendas planificadas.	-
DIST_E_V	Distancia euclidiana a los espacios verdes presentes en el área de estudio.	m	

	DIST_CC_2	Distancia euclidiana a calles comerciales.	m
	LN_DIST_CC	Logaritmo natural de la distancia euclidiana al Centroide Comercial.	(Logaritmo Natural) m

[Fuente: Elaboración Propia]

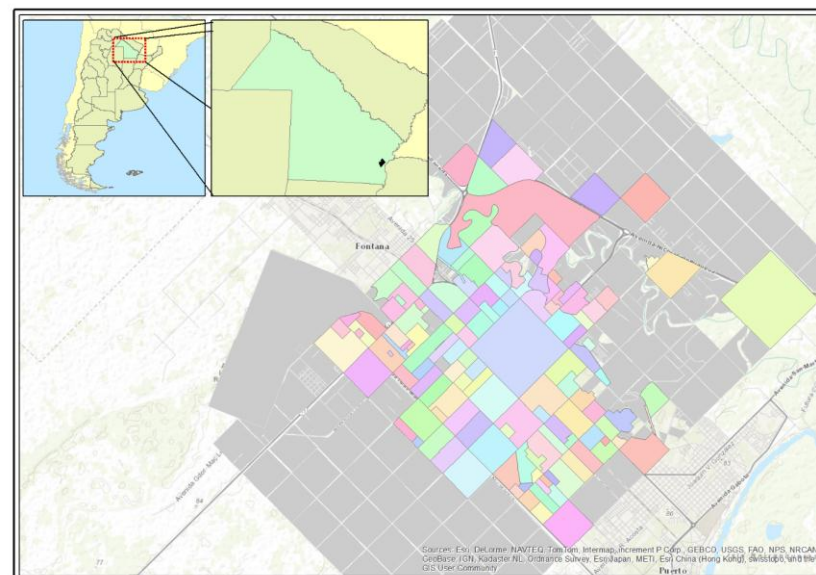
Área de Estudio

La ciudad de Resistencia localizada en el Departamento San Fernando, y capital de la Provincia del Chaco, fue fundada en el año 1878 en un asentamiento forestal, sumándose luego un contingente de inmigrantes italianos que permitió luego la colonización total del territorio.

Con una población de 390.874 habitantes según el Censo Nacional de Población, Hogares y Viviendas del año 2010, es la ciudad más poblada de la región del NEA (Nordeste Argentino). Asimismo, se encuentra situada en el valle aluvial del Río Paraná, y atravesada por el Río Negro al norte y por el riacho Arazá al Sur, el cual fue parcialmente entubado causando importantes problemas en el natural escurrimiento de las aguas en la zona.

Para un adecuado análisis estadístico se han empleado las unidades administrativas barriales debido a que resumen significativamente, las variaciones locales del precio fiscal del suelo y reduce los tiempos computacionales de procesamiento, tanto del software SIG como estadísticos empleados.

Figura 3: Distribución de barrios en la ciudad de Resistencia.



[Fuente: Elaboración Propia]

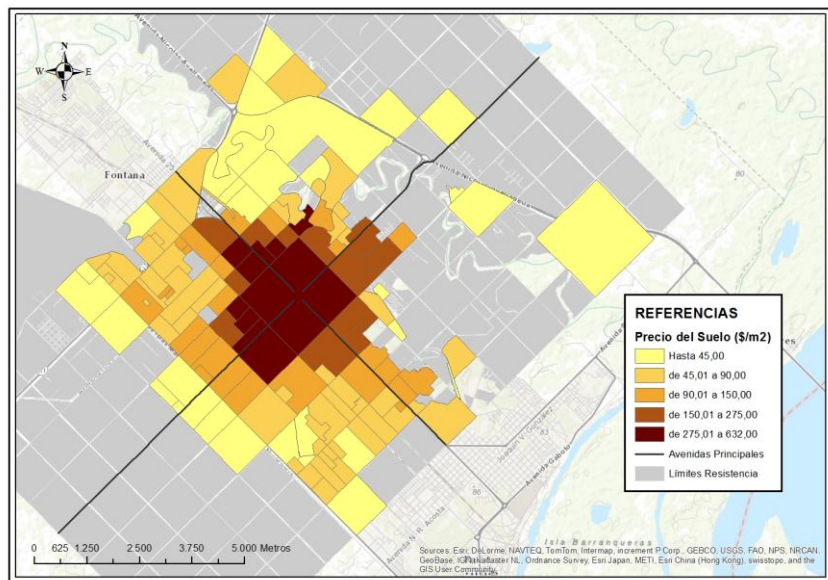
RESULTADOS

Distribución del Precio fiscal del Suelo

Uno de los primeros pasos en el análisis explicativo del precio del suelo consiste en conocer la distribución del mismo a los efectos de tener una visión espacial de su comportamiento.

En la Figura 4 vemos la distribución espacial del precio fiscal del suelo, donde el centro de la ciudad es el espacio geográfico que alberga los intervalos de precios más elevados (150,01 a 632,00 \$/m²), disminuyendo paulatinamente hacia la periferia.

Figura 4: Distribución del Precio fiscal del Suelo por barrio.



[Fuente: Elaboración Propia]

Análisis de Correlación de las variables

Los resultados obtenidos fueron sometidos al cumplimiento de los supuestos de modo tal que el ajuste obtenido sea estadísticamente significativo. En este contexto, uno de los primeros pasos fue elaborar una tabla de correlaciones bivariadas para corroborar la relación lineal entre todas las variables. Para ello se ha utilizado el coeficiente de correlación de Pearson (r) a niveles de significación de 0,01 y 0,05.

La Tabla 2 muestra los coeficientes de correlación obtenidos para cada una de las variables independientes y para cada independiente con respecto a la dependiente. Puede observarse que las correlaciones entre las variables independientes entre sí no alcanzan al 0,7, siendo la más alta la de DIST_CC_2 con DIST_AV con un coeficiente de 0,656 a un nivel de significación de 0,05.

Tabla 2: Matriz de correlaciones bivariadas entre las variables.

	PRECIO_SUELO	DENSIDAD_QA LT	DIST_AV	INSEGURIDAD	DIST_PAV	DIST_TP	POB_NBI	VIV_PLANIF	DIST_E_V	DIST_CC_2	LN_DIST_CC
PRECIO_SUELO	1										
DENSIDAD_QALT	,642**	1									
DIST_AV	-,431**	-,281**	1								
INSEGURIDAD	-,093	,034	-,036	1							
DIST_PAV	-,364**	-,365**	,186*	-,140	1						
DIST_TP	-,210*	-,230**	,225**	-,075	,572**	1					
POB_NBI	-,139	-,159	-,026	,345**	,023	-,070	1				
VIV_PLANIF	-,267**	,059	,498**	,095	-,093	-,131	-,026	1			
DIST_E_V	-,250**	-,328**	,156	-,175*	,528**	,619**	-,077	-,109	1		
DIST_CC_2	-,647**	-,492**	,656**	,016	,507**	,453**	-,034	,284**	,566**	1	
LN_DIST_CC	-,761**	-,540**	,603**	,060	,458**	,367**	,010	,300**	,468**	,961**	1

[Fuente: Elaboración Propia]

** correlación significativa en el nivel 0,01 (2 colas); * correlación es significativa en el nivel 0,05 (2 colas).

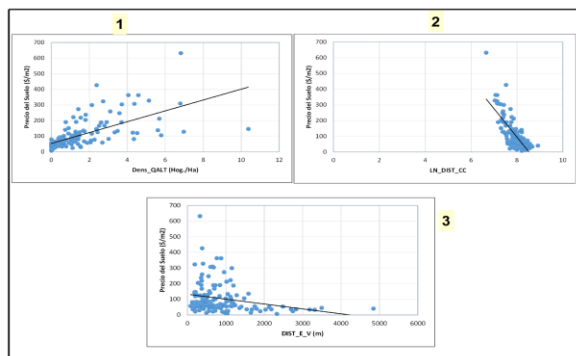
Sin embargo, si observamos la correlación entre LN_DIST_CC y DIST_CC_2, el coeficiente obtenido es de 0,961 con lo cual se podría estar violando el supuesto de No Colinealidad entre las variables independientes, por lo tanto se debe considerar emplear una u otra variable en el modelo. La naturaleza de esta característica radica en que ambas variables comparten la unidad espacial de observación, aunque DIST_CC_2 a diferencia de LN_DIST_CC incorpora no solamente el centroide comercial, sino también otras calles de presencia comercial marcada.

Resultados del modelo de regresión aplicado

Posteriormente a la normalización de las variables y la exploración descriptiva de asociación, se procedió a implementar el método de paso a paso para seleccionar las variables que mayor significancia tengan en el modelo final.

De todas la variables empleadas, las que han resultado significativas son LN_DIST_CC, DENSIDAD_QALT y DIST_E_V, dado sus coeficientes de significancia obtenidos (ver Tabla 4). Asimismo, a los efectos de observar la asociación de dichas variables independientes con la dependiente, se ha elaborado gráficos de dispersión que sintetizan dicho comportamiento, ellos son:

Figura 5: Gráfico de Correlación entre (1= PRECIO_SUELO - DENS_QALT); (2= PRECIO_SUELO -LN_DIST_CC); (3= PRECIO_SUELO - DIST_E_V).



[Fuente: Elaboración Propia]

Con los gráficos de dispersión (Figura 5) se puede observar la linealidad entre las variables independientes y la dependiente, con lo cual se satisface el supuesto de Linealidad con ciertas limitaciones en la relación de Precio del Suelo y DIST_E_V que presenta una tendencia más a curvilínea debido a la distribución de los espacios verdes en la ciudad.

Atendiendo los coeficientes obtenidos (Ver Tabla 3) en la aplicación del modelo de regresión lineal múltiple, podemos observar un R² de 0,677 y un R² Ajustado de 0,669, lo cual expresados en porcentajes indican que un 66,9% de la variabilidad de los datos de precio fiscal del suelo, se encuentran explicados por las variables independientes empleadas en este modelo.

Tabla 3: Resumen del modelo.

R	R ²	R ² ajustado	Error estándar de estimación	Durbin-Watson
0,823 ^a	0,677	0,669	54,99282088	1,685

[Fuete: Elaboración Propia]

Tabla 4: Tabla ANOVA del modelo de regresión.

Modelo	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Regresión	847707,969	3	282569,323	93,436	0,000
Residuo	405244,187	134	3024,210		
Total	1252952,156	137			

[Fuete: Elaboración Propia]

Asimismo, la Tabla 4 muestra la significación global del modelo en el que el test F resulta estadísticamente significativo dado su valor de 93,436 que supera el límite crítico establecido para un número de observaciones (n) de 138 y una significación del 0,05.

Ahora bien, si atendemos los coeficiente obtenidos para cada una de las variables independiente, vemos que son estadísticamente significativas al 0,05 con una valor de “t” superior a 2 o inferior a -2 en todos los casos, lo cual revela la que los coeficientes obtenidos en este modelo no tienen problemas de significación.

Tabla 5: Coeficientes obtenidos en el modelo de regresión.

Modelo	Coeficientes			t	Sig.
	Estandarizado		No Estandariz.		
	B	Error estándar	Beta		
(Constante)	1291,379	118,447		10,903	0,000
DENSIDAD_Q	18,621	3,174	0,344	5,866	0,000
DIST_ESP_V	0,021	0,007	0,169	3,028	0,003
LN_DIST_CC	-155,069	14,869	-0,654	-10,429	0,000

[Fuente: Elaboración Propia]

Análisis de los Residuales

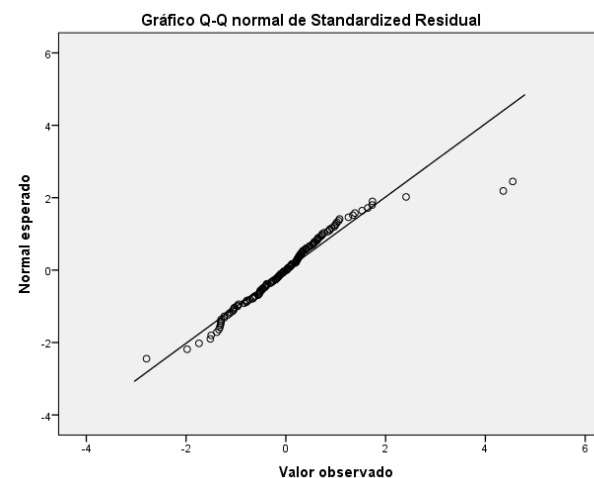
Normalidad de los residuales

El análisis de los residuales se realizó considerando los supuestos de la regresión, de manera tal que se pueda corroborar la distribución normal de los mismos.

En primera instancia se llevó a cabo un análisis de Normalidad de los residuales empleando a tal fin el gráfico Q-Q normal, con el propósito de evaluar la distribución de los residuales sobre una curva normal teórica.

La Figura 6 revela que el comportamiento de las observaciones tiende a concentrarse próximas a la recta teórica. Sin embargo, se observa la presencia de valores atípicos en los extremos del conjunto de observaciones.

Figura 6: Q-Q Plot Normal.



[Fuente: Elaboración Propia]

Asimismo, es importante contrastar con otro estadístico de normalidad más riguroso como el test de Kolmogorov-Smirnov, dado que la percepción visual puede obviar algunos detalles. Los resultados obtenidos de la prueba son:

Tabla 6: Prueba de Normalidad de Kolmogorov-Smirnov.

Estadístico	gl	Sig.
0,078	138	0,040

[Fuente: Elaboración Propia]

Dado que el estadístico resultante fue de 0,04, trabajando a un nivel de confianza del 95%, no podemos rechazar la H0, y por lo tanto, de acuerdo a este test analítico, observamos problemas en la normalidad de los residuales.

Supuesto de Media Cero y Homocedasticidad

Para los residuos estandarizados se han obtenidos las estadísticas descriptivas de modo tal que se pueda corroborar el cumplimiento de los supuestos, las cuales se detallan en la siguiente tabla:

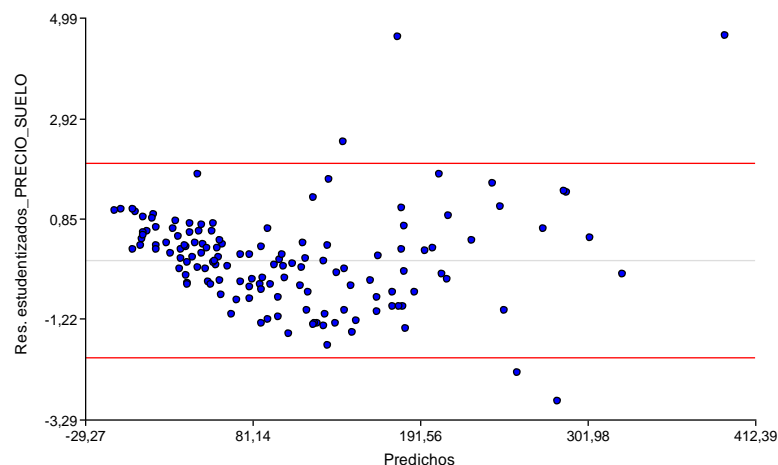
Tabla 7: Estadísticos descriptivos de los residuales.

N	Media	Desviación Estándar	Varianza
138	0,000	0,9889	0,978

[Fuente: Elaboración Propia]

Con ello se puede corroborar el cumplimiento del supuesto de media 0 y suma de los desvíos igual a 1, con una pequeña diferencia en el valor de la variación de los residuales producidos por las tendencias que presentan (Ver Figura 7) y por ello, el comportamiento reflejado en los estadísticos.

Figura 7: Residuos Estudentizados vs Predichos.



[Fuente: Elaboración Propia]

Análisis de Autocorrelación

El estadístico Durbin-Watson es una prueba de autocorrelación que permite observar comportamiento en los residuales de autocorrelación positiva o negativa o bien la no autocorrelación.

Por ello, dado un número de observaciones (N) de 138 y 3 variable independientes (k), se ha calculado los Límites inferior (Li) y superior (Ls), para llevar a cabo el análisis. Los resultados obtenidos para cada uno de los parámetros son:

Tabla 8: Estadísticos Durbin-Watson.

N	K	Li	Ls
138	3	1,69262	1,75138

[Fuente: Elaboración Propia]

Dado que el DW obtenido 1,685 vemos que se encuentra por debajo de los límites calculados, por ello concluimos que los residuales presentan una leve tendencia hacia la Autocorrelación Positiva, lo cual implica la violación de este supuesto y la necesidad de aplicar alguna regla de transformación de los datos o analizar más detenidamente la naturaleza de ellos.

Análisis de Autocorrelación Espacial

I de Moran

Para obtener una visión global de la autocorrelación y considerar la componente espacial en la distribución de los residuales del modelo de regresión, se ha considerado aplicar el I de Moran por medio de la conceptualización de relación espacial denominada *Zona de Indiferencia*.

Al analizar el resultado del test, se constata que el valor del I de Moran estimado (-0,025496) no es estadísticamente diferente respecto al esperado (-0,007299), por lo tanto el patrón de distribución observado es aleatorio.

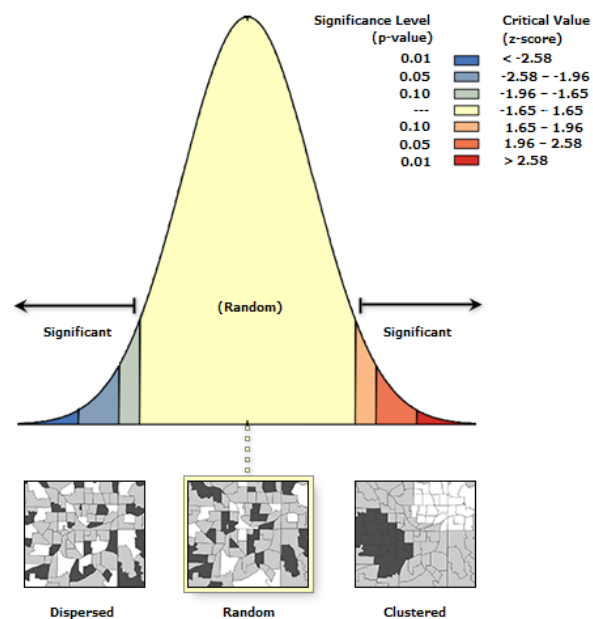
Tabla 9: Estadísticas del I de Moran.

I de Moran	I Esperado	Varianza	z-score	p-valor
-0,025496	-0,007299	0,000213	-1,248194	0,211960

[Fuente: *Elaboración Propia*]

Asimismo, el z-score obtenido arroja un valor muy bajo acompañado de un p-valor de 0,21, por lo tanto, los resultados obtenidos a partir de los estadísticos no cumplen los requisitos que conlleven rechazar la hipótesis nula y por ello se observa un patrón global aleatorio en los residuales del modelo de regresión aplicado.

Figura 8: Patrón observado a partir del índice de Moran.



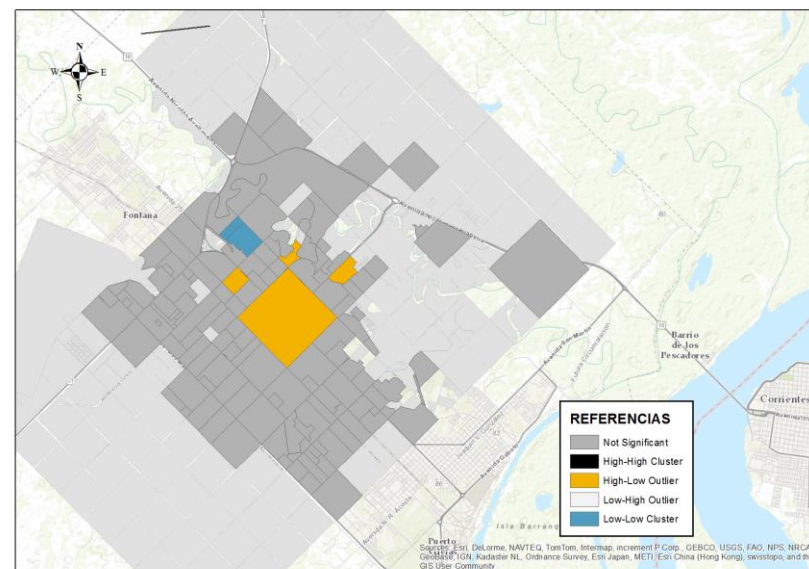
[Fuente: *ESRI (Environment System Research Institute)*]

Local Indicators of Spatial Association

Conociendo el patrón de distribución aleatorio de los residuales por medio del I de Moran, se procedió a evaluar los mismos en base a un criterio local, de manera que las observaciones adquirieran una independencia y no sean globalmente generalizadas. Por ello, el LISA resulta una manera más apropiada de analizar la distribución espacial de los residuales y por tanto, identificar los valores atípicos presentes.

El LISA resultante muestra la distribución espacial de los residuales con comportamiento atípico. En este sentido, los barrios que mayores problemas presentan en ajustarse al modelo son el *Centro*, *Villa Altabe*, *Villa Avalos* y *Villa 25 de Mayo* y *Pezzini* (Ver Figura 9), es decir que esos barrios presentan valores altos y bajos próximos (intervalo de High-Low Outlier).

Figura 9: Distribución espacial del LISA.



[Fuente: *Elaboración Propia*]

DISCUSIÓN Y CONCLUSIONES

Desde el punto de vista metodológico, el enfoque de los precios hedónicos se muestra adecuado para el análisis del valor del suelo, debido a que una ecuación hedónica puede resumir de manera objetiva la influencia de un conjunto de variables significativas en la conformación del precio del suelo. Pero el principal aporte es este sentido, ha sido su integración con los Sistemas de Información Geográfica (SIG) para la generación y cálculo de las variables que alimentaron el modelo de regresión múltiple.

El modelo obtenido presenta un ajuste ($R^2=0,67$) y significación conjunta ($F=93,436$ y $p<000$) aceptables. Por su parte, las variables introducidas muestran coeficientes con signos correctos, valores de coeficientes $\neq 0$ y estadísticamente significativos al 0,05. Esto nos permite afirmar que la densidad de estratos socioeconómicos altos, y las distancias a espacios verdes y el centro comercial de Resistencia, influyen efectivamente en la estructura del precio fiscal del suelo; de esta forma podemos aceptar la hipótesis de trabajo (H_1) formulada oportunamente. Sin embargo dado la complejidad de la temática, la implementación de un modelo que incluye solamente tres variables para explicar el precio del suelo, puede parecer demasiado sencillo, aunque en conjunto las variables explican el 66,9% la variabilidad y el modelo es estadísticamente robusto.

Asimismo, considerando la elasticidad de las variables explicativas observamos que la DENSIDAD_Q presenta una relación positiva con el precio del suelo, de modo tal que el aumento en una unidad la DENSIDAD_Q, conllevaría un aumento de \$18,6/m² en el valor del precio fiscal del suelo. Por otro lado, considerando el LN_DIST_CC, aumentar en una unidad de logaritmo natural la distancia al centro comercial, produciría una caída del precio fiscal del suelo en el orden de \$155/m².

Vemos entonces que las variables Distancia a Centroide Comercial (convertida a logaritmo natural) y Distancia a Espacios verdes presentan importancia considerables en la distribución de los precios del suelo. En la primer variable, la estructura de la ciudad influye notablemente, debido a que los servicios se concentran mayormente en áreas a la calle Peatonal, donde los usos del suelo predominante son el comercial y el administrativo (Da Silva et. al, 2013; Lucca y Rey, 1999).

Los hogares con estratos socioeconómicos altos, se concentran mayormente en el centro de la ciudad o próximos a barrios como Avalos y Los Troncos (Odrizola et. al., 2007). Ello influye notablemente en los precios del suelo, de manera tal que, algunos barrios adquieren valores que determinan comportamiento de outliers en la aplicación del modelo.

La autocorrelación por sí misma, debe atenderse cuidadosamente al emplear variables con características geográficas, debido a que el territorio puede influir notablemente en la bondad del modelo. En el modelo planteado, el estadístico Durbin-Watson y de Kolmogorov-Smirnov, han arrojado problemas de autocorrelación y problemas de normalidad en los residuales. Sin embargo, al evaluar los residuales por medio del I de Moran y el LISA empleando la Zona de Indiferencia como conceptualización de las relaciones espaciales, se ha podido advertir una distribución aleatoria en los residuales y la identificación local de los outliers. Consideramos entonces la importancia de la aplicación de test espaciales a la hora de evaluar variables de carácter geográfico, para mejorar la performance de los modelos incorporando la distribución espacial como componente propia del fenómeno.

Vecchione de Ochoa (2007) por medio de un modelo de regresión con variables como el ingreso familiar, área de construcción residencial y distancia al centro, obtiene ajuste de R^2 de 0,74 aunque valores en el estadístico Durbin-Watson, ponen en evidencia problemas de autocorrelación en los residuos. Por esta razón, recurre a la implementación de un modelo logarítmico con el objeto de mejorar el ajuste, logrando su modelo con R^2 de 0,83 aunque el valor de 1,76 en Durbin-Watson permanece encima del límite crítico. En comparación con el modelo propuesto en el presente trabajo, la incorporación del análisis de los residuales por medios de estadísticos que conceptualizan las relaciones espaciales entre las variables empleadas, ha podido resolver los problemas de autocorrelación en los residuales, y obtener un modelo que si bien explica $2/3$ de la variación del precio del suelo, cumple los supuestos del modelo de regresión.

Es oportuno por tanto, explorar la aplicación de modelos de regresión espaciales como los del tipo GWR, Spatial Lag, Spatial Error, etc para considerar el efecto geográfico en el comportamiento de las variables. Por otra parte, se debería complementar este trabajo con la modelización del precio del suelo en el mercado inmobiliario, a los efectos de tener una visión más completa de fenómeno, como

también evaluar su comportamiento o peso al momento de explicar otros procesos más complejos, tales como la expansión urbana de la ciudad.

BIBLIOGRAFÍA

- ANDERSON, T.R. (1962). Social and Economic Factors affecting the Location of Residential Neighborhoods. Papers and Proceedings, Regional Science Association, (9), pp.161-170.
- ANSELIN, LUC. (1995). Local Indicators of Spatial Association – LISA. Geographical Analysis, 5, 2. pp. 93-116.
- ARIAS, F., ORTIZ, R., CARDOZO, O., y DA SILVA, C. (2013). Análisis de la Distribución Espacial de Actividades Económicas en la Ciudad de Resistencia, Argentina. V Seminario Internacional Sobre Políticas Urbanas, Gestión Territorial y Ambiental para el Desarrollo Local. Resistencia: Instituto de Planeamiento Urbano y Regional - Facultad de Arquitectura y Urbanismo – UNNE.
- BARREDO C, J. I. y BOSQUE SENDRA, J. (1999). Multicriteria evaluation methods for ordinal data in a GIS environment, Geographical Systems, 5, pp. 313-327.
- BARREDO CANO, J. I. (1996). Sistemas de información geográfica y evaluación multicriterio en la ordenación del territorio. Ed. RA-MA. Madrid, pp. 264.
- CANO GUERVÓS, R. A Y CHICA OLMO, J. M. (2004). Una Metodología Objetiva para las Valoraciones Inmobiliarias. Capítulo 7. Departamento de Métodos Cuantitativos para la Economía y la Empresa. ISBN 84-338-3108-9. Editorial: Universidad de Granada. España, pp. 121-148.
- DA SILVA, C. J., CARDOZO, O. D., ODRIOZOLA, J. G., Y BONDAR, C. E. (2013). Usos del Suelo: Distribución, Análisis y Clasificación con Sistemas de Información Geográfica (SIG). Geografía y Sistemas de Información Geográfica. (GESIG-UNLU, Luján)(5), pp.142-152.
- DA SILVA, C. J., CARDOZO, O. D., ODRIOZOLA, J. G., BONDAR, C. E., Y BRILLADA, D. S. (2014). Uso Comercial del Suelo y Transporte Público en el Centro de la Ciudad de Resistencia, Argentina. XVIII Congreso Panamericano de Ingeniería del Tránsito, Transporte y Logística. Santander. pp. 1-18.
- DE SMITH, M. J; GOODCHILD, M. F. & LONGLEY, P. A. (2013). Geospatial Analysis. A Comprehensive Guide to Principles, Techniques and Software Tools. The Winchelsea Press, Winchelsea, UK.
- ESTEBANEZ ALVAREZ, J., & BRADSHAW, R. P. (1978). Técnicas de Cuantificación en Geografía. Ed. TEBAR FLORES. Madrid.
- GALLEGO GÓMEZ, J. L. (2008). Apuntes de Econometría. LADE y LE. Departamento de Economía. Universidad de Cantabria.
- GÓMEZ DELGADO, M., Y BARREDO CANO, J. I. (2005). Sistemas de Información Geográfica y evaluación Multicriterio en la ordenación del territorio. Ed. Ra-Ma Madrid.
- HARVEY. D (2009). Social Justice and the City. University of Georgia Press. Revised edition. ISBN: 978-0-8203-3604-6, pp.355.
- HUMARÁN NAHED, I., Y ROCA CLADERA, J. (2010). Hacia una medida integrada del Factor de Localización en la valoración residencial: el caso de Mazatlán. ACE. Architecture, City and Environment, 13, pp. 185-218.
- JIANG, H. & EASTMAN, J. R. (2000). Application of fuzzy measures in multi-criteria evaluation in GIS. International Journal of Geographical Information Science, 14, 2, pp. 173-184.
- LEVER G. D. (2009). El Modelo de Precios Hedónicos. Asociación de Arquitectos Tasadores de Chile (Asatch), pp. 13
- MALCZEWSKI, J. (2002). Fuzzy screening for land suitability analysis. Geographical & Environmental Modelling, 6, 1, pp. 27-39.
- MELONI, O. Y RUIZ NUÑEZ. F. (2002). El precio de los terrenos y el valor de sus atributos. Un enfoque de precios hedónicos. Económica, Vol. XLVIII (1-2) Facultad de Ciencias Económicas. Universidad Nacional de La Plata. ISSN: 1852-1649, pp. 69-88.
- NÚÑEZ CERDA, F. Y SCHOVELIN SURHOFF, R. (2002). Modelos de Precio de Suelo Urbano en Gran Concepción. Revista Ingeniería Industrial - Año 1, N° 1. Concepción, Chile. ISSN: 0717-9103.
- NÚÑEZ CERDA, F.J. Y ROCA CLADERA, J. (2007). Especificación y estimación de modelos de formación de precios del suelo urbano en el Gran Concepción, Chile. ACE. Architecture, City and Environment, 2, 4. ISSN-e 1886-4805, pp. 554-584.
- ODRIOZOLA, J. G. Y FALCÓN, V. L. (2007). Geo-referenciación de los niveles socioeconómicos del Gran Resistencia. Facultad de Ciencias Económicas Facultad de Humanidades. Asociación de la Producción, Industria y Comercio de Corrientes, pp. 20.
- PYSZCZEK, O. L. (2012). La estructura formal y funcional de la Seguridad Pública, y su relación con los Espacios del miedo, en la ciudad de Resistencia, Chaco. XIV Encuentro de Profesores en Geografía del Nordeste Argentino. Resistencia. pp 1-11.
- REY, W., Y LUCCA, A. (1999). Aplicación de SIG a Aspectos Socio-Ambientales del Centro de la ciudad de Resistencia 1999. Serie Geográfica N°8, pp.137-151.

SCORNIK, C. O. (1998). Diagnóstico Expositivo del Área Metropolitana del Gran Resistencia. Informe General. Ministerio del Interior. Resistencia, Chaco.

VECCHIONE DE OCHOA, L. (2007). Un modelo del valor del suelo urbano en el área metropolitana de Mérida. Instituto de Investigaciones Económicas y Sociales. Universidad de los Andes, pp. 189-211.

WU, F. & WEBSTER, C. J. (1998). Simulation of land development through the integration of cellular automata and multicriteria evaluation. Environment and Planning B: Planning and Design, 25, pp. 103-126.