

Inequidades socioeconómicas en relación a los  
espacios verdes y cobertura de copa: un  
enfoque espacial aplicado a la ciudad de Cali,  
Colombia

Juan Fernando Correa Caicedo

*Presentado para obtener al título de Maestría en SIG (M.Sc. UNIGIS)*

*Universidad de Salzburgo*

*2018-03-20*



# Índice general

<b>Presentación</b>	<b>13</b>
<b>Agradecimientos</b>	<b>15</b>
<b>Resúmen</b>	<b>17</b>
<b>Abstract</b>	<b>19</b>
<b>1. Introducción</b>	<b>21</b>
1.1. Antecedentes . . . . .	21
1.2. Objetivos . . . . .	24
1.2.1. Objetivo general . . . . .	24
1.2.2. Objetivos específicos . . . . .	24
1.3. Preguntas de investigación . . . . .	25
1.4. Hipótesis . . . . .	25
1.4.1. Beneficios del arbolado urbano . . . . .	25
1.4.2. Acceso a espacios verdes . . . . .	26
1.5. Justificación . . . . .	27
1.6. Alcance . . . . .	28
<b>2. Revisión de la literatura</b>	<b>29</b>
2.1. Servicios ecosistémicos y su valoración . . . . .	29
2.2. La perspectiva de la justicia ambiental . . . . .	35

2.3.	Modelamiento y análisis espacial de variables ambientales y sociales . . . . .	38
2.3.1.	Enfoque estadístico . . . . .	47
2.3.2.	Enfoque geoestadístico . . . . .	49
2.3.3.	Otras técnicas . . . . .	51
<b>3.</b>	<b>Metodología</b>	<b>53</b>
3.1.	Área de estudio . . . . .	54
3.2.	Datos . . . . .	55
3.2.1.	Datos de registros oficiales del municipio . . . . .	55
3.2.2.	El censo arbóreo . . . . .	56
3.2.3.	El censo de población . . . . .	60
3.3.	Métodos y técnicas . . . . .	62
3.3.1.	Procesamiento de datos . . . . .	65
3.3.2.	Cálculo de métricas de acceso a servicios ambientales . . . . .	66
3.3.3.	Cálculo de métricas de sobre la población . . . . .	68
3.3.4.	Criterios y selección de sectores censales . . . . .	73
3.3.5.	Selección de variables dependientes y regresiones lineales . . . . .	74
3.3.6.	Análisis geoestadísticos . . . . .	77
<b>4.</b>	<b>Resultados</b>	<b>85</b>
4.1.	Modelando la cobertura de copa . . . . .	85
4.1.1.	Correlaciones y gráficos de dispersión bivariados . . . . .	85
4.1.2.	Modelos de regresión lineal . . . . .	93
4.1.3.	Modelado espacial AU . . . . .	103
4.2.	Acceso a espacios verdes . . . . .	135
4.2.1.	Correlaciones y distribuciones bivariadas . . . . .	138
4.2.2.	Modelos de regresión lineal EV . . . . .	142
4.2.3.	Modelado espacial de espacios verdes . . . . .	153
<b>5.</b>	<b>Discusión</b>	<b>181</b>
5.1.	Sobre la cobertura de copa . . . . .	181

5.2. Sobre los espacios verdes . . . . .	189
<b>6. Conclusiones</b>	<b>197</b>



# Índice de cuadros

2.1.	Clasificación de servicios ecosistémicos importantes en zonas urbanas y funciones y componentes subyacentes del ecosistema . . . . .	30
2.2.	Métricas para caracterizar servicios del AU . . . . .	40
2.3.	Métricas para caracterizar servicios del EV . . . . .	44
2.4.	Métricas para caracterizar aspectos socioeconómicos de la población y estructura de las unidades espaciales del EV . . . . .	45
3.1.	Variables para caracterizar el AU . . . . .	56
3.2.	Muestra de los datos del censo arbóreo . . . . .	57
3.3.	Variables sobre la población . . . . .	61
3.4.	Variables sobre las viviendas . . . . .	61
3.5.	Total de población en la ciudad de Cali . . . . .	71
4.1.	Resumen métricas de ajuste OLS para el área de copa (AC) y cobertura de copa (CC) . . . . .	100
4.2.	Metricas de ajuste para los modelos de área de copa . . . . .	127
4.3.	Metricas de ajuste para los modelos de porcentaje de área de copa . . . . .	135
4.4.	Resumen métricas de ajuste OLS Indice contenedor y de acceso área-distancia . . . . .	150
4.5.	Metricas de ajuste para los modelos de porcentaje de área de EV . . . . .	168
4.6.	Metricas de ajuste para los modelos de áreas-distancia de EV . . . . .	178
5.1.	Coeficientes del modelo SD de área de copa . . . . .	183

5.2. Coeficiente de autocorrelación modelo SD de área de copa . . . . .	183
5.3. Coeficientes del modelo SAR de porcentaje de área de copa . . . . .	183
5.4. Coeficiente de autocorrelación modelo SAR de área de copa . . . . .	184
5.5. Coeficientes del modelo SEM de porcentaje de área de espacio verde . . . . .	190
5.6. Coeficiente de autocorrelación modelo SEM de porcentaje de área de espacio verde . . . . .	190
5.7. Coeficientes del modelo SD del indicador de acceso área-distancia . . . . .	191
5.8. Coeficiente de autocorrelación modelo SD del indicador de acceso área-distancia . . . . .	191

# Índice de figuras

3.1. Área de estudio . . . . .	54
3.2. Capas usadas para el procesamiento de los espacios verdes y las características de las manzanas . . . . .	58
3.3. Small multiples de los individuos arbóreos por emplazamiento . . . . .	59
3.4. División en barrios y sectores urbanos de Santiago de Cali . . . . .	63
3.5. Diagrama de flujo de la metodología . . . . .	64
3.6. Sectores urbanos de las variables dependientes sobre cobertura de copa . .	67
3.7. Espacio verdes y rango de 1 km desde centriodes de SU . . . . .	69
3.8. Mapas de las variables de población seleccionadas (en deciles) . . . . .	70
3.9. Mapas de las variables de población seleccionadas como porcentajes (en deciles) . . . . .	72
3.10. Mapas de las variables sobre el tipo de uso de los predios como porcentaje de la cantidad de predios (escala continua) . . . . .	73
3.11. Mapa de los sectores excluidos . . . . .	75
3.12. Grafo de vecindad entre todos los SU de la ciudad de Cali . . . . .	79
3.13. Grafo de vecindad entre los SU seleccionados para el análisis del AU . . .	80
4.1. Comparación por pares entre predictores de población . . . . .	86
4.2. Comparación por pares entre predictores de población porcentuales . . . .	88
4.3. Coeficiente Pearson entre variables de población . . . . .	89
4.4. Coeficiente Spearman entre variables de población . . . . .	90
4.5. Coeficiente Pearson entre cobertura de copa y variables de población . .	91

4.6. Coeficiente Spearman entre cobertura de copa y variables de población . . . . .	92
4.7. Coeficiente Pearson entre variables de uso de los predios y espacios verdes en los sectores urbanos . . . . .	94
4.8. Coeficiente Pearson entre variables de uso de los predios y espacios verdes en los sectores urbanos . . . . .	95
4.9. Coeficiente Pearson entre coberturas de copa y variables de predios y EV .	96
4.10. Coeficiente Pearson entre coberturas de copa y variables de predios y EV .	96
4.11. Gráficas diagnósticas para el análisis de regresión lineal de área de copa con los nuevos términos . . . . .	101
4.12. Gráficas diagnósticas para el análisis de regresión lineal de porcentaje de cobertura de copa . . . . .	102
4.13. Variable dependiente, valor ajustado por el modelo y residuos del OLS para log.area copa normalizada . . . . .	103
4.14. Variable dependiente, valor ajustado por el modelo y residuos del OLS para cobertura copa.ap normalizada . . . . .	104
4.15. Matrices de vecindad del análisis espacial . . . . .	105
4.16. Gráfico de Moran para los residuos del modelo lineal para del área de cobertura de copa usando $W_q$ . . . . .	107
4.17. Mapas LISA para la matriz $W_q$ de ‘log.area copa’ . . . . .	108
4.18. Mapas LISA para la matriz $W_d$ de ‘log.area copa’ . . . . .	109
4.19. Gráfico de Moran para los residuos del modelo lineal para el porcentajes cobertura de copa usando $W_q$ . . . . .	110
4.20. Mapas LISA para la matriz $W_q$ de ‘cobertura copa.ap’ . . . . .	110
4.21. Mapas LISA para la matriz $W_d$ de ‘cobertura copa.ap’ . . . . .	111
4.22. Gráfico de Moran para los residuos del modelo lineal para del área de cobertura de copa usando $W_q$ y $W_d$ . . . . .	112
4.23. Mapas LISA para la matriz $W_q$ de los residuos del modelo lineal para del área de copa . . . . .	113

4.24. Mapas LISA para la matriz $W_d$ dende los residuos del modelo lineal para del área de copa . . . . .	114
4.25. Gráfico de Moran para los residuos del modelo lineal para el porcentaje de cobertura de copa usando $W_q$ y $W_d$ . . . . .	116
4.26. Mapas LISA para la matriz $W_q$ de los residuos del modelo lineal para el porcentaje de cobertura copa . . . . .	117
4.27. Mapas LISA para la matriz $W_d$ dede los residuos del modelo lineal para del área de copa . . . . .	118
4.28. Gráfico de Moran para los residuos de los modelos espaciales de área de copa usando $W_q$ . . . . .	128
4.29. Diagnóstico comparativo entre modelos . . . . .	129
4.30. Gráfico de Moran para los residuos de los modelos espaciales del porcentaje de área de copa usando $W_q$ . . . . .	136
4.31. Diagnóstico comparativo entre modelos de porcentaje de copa . . . . .	137
4.32. Metricas de acceso a espacio verdes . . . . .	138
4.33. Metricas de acceso a espacio verdes seleccionadas . . . . .	139
4.34. Coeficiente Pearson entre acceso a EV y variables de población . . . . .	140
4.35. Coeficiente Spearman entre acceso a EV y variables de población . . . . .	141
4.36. Coeficiente Pearson entre acceso a EV y variables de uso de los predios . . . . .	141
4.37. Coeficiente Spearman entre acceso a EV y variables de uso de los predios . . . . .	142
4.38. Coeficiente Pearson entre acceso a EV y variables sobre aspectos físicos de las manzanas y SU . . . . .	143
4.39. Coeficiente Spearman entre acceso a EV y variables sobre aspectos físicos de las manzanas y SU . . . . .	143
4.40. Gráficas diagnósticas para el análisis de regresión lineal de porcentaje de area de espacio verde . . . . .	146
4.41. Variable dependiente, valor ajustado por el modelo y residuos del OLS para 'area ep.porcentaje' normalizada . . . . .	147
4.42. Gráficas diagnósticas para el análisis de regresión lineal de 'ia.areas.dist' . .	151

4.43. Variable dependiente, valor ajustado por el modelo y residuos del OLS para ‘ia.Areas.dist’ normalizada . . . . .	152
4.44. Matrices de vecindad del análisis espacial de espacios verdes . . . . .	154
4.45. Mapas LISA para la matriz $W_q$ de ambos indicadores de acceso a EV . . . . .	158
4.46. Gráfico de Moran para los residuos del modelo lineal para el porcentaje de área de espacio verde . . . . .	159
4.47. Gráfico de Moran para los residuos del modelo lineal para el indicador ‘ia.Areas.dist’ . . . . .	161
4.48. Mapas LISA para la matriz $W_q$ de los residuos de los modelos lineales para los indicadores de acceso a EV . . . . .	162
4.49. Gráfico de Moran para los residuos de los modelos espaciales de porcentaje de área de EV $W_q$ . . . . .	169
4.50. Diagnóstico comparativo entre modelos . . . . .	170
4.51. Gráfico de Moran para los residuos de los modelos espaciales de acceso área-distancia $W_q$ . . . . .	179
4.52. Diagnóstico comparativo entre modelos . . . . .	180
5.1. Mapas LISA - Porcentaje población Afro . . . . .	184
5.2. Mapas LISA - Porcentaje población sin estudios . . . . .	185
5.3. Mapas LISA - Porcentaje población con estudios superiores . . . . .	186
5.4. Mapas LISA - Densidad de Población . . . . .	187
5.5. Mapas LISA - Porcentaje de viviendas tipo cuarto . . . . .	188
5.6. Distribucion de lo indicadores de acceso local a EV . . . . .	192
5.7. Distribucion de lo indicadores de acceso a EV área-distancia . . . . .	194
5.8. Mapas LISA - Porcentaje de espacios verdes y área media de manzana . . . . .	195
5.9. Mapas LISA - Porcentaje de unidades economicas y personas con alguna limitación . . . . .	196

# Presentación

Borrador del texto. Tesis para obtener el título de master en Geographical Information Science & Systems (UNIGIS M.Sc.).



# Agradecimientos

A Elizabeth, Augusto, Ana María y Catalina. Este trabajo fue posible gracias a su ayuda y cariño.



# Resumen

Este proyecto hace uso del censo arbóreo urbano realizado en Santiago de Cali (Colombia) entre el año 2014 y 2015, de los datos del censo de población de 2005 y los datos de la estructura ecológica del municipio del 2014. El propósito es identificar si existen patrones espaciales estadísticamente significativos que den evidencia de sesgos de los indicadores de acceso a servicios ambientales del arbolado urbano (AU) y espacios verdes EV con indicadores socioeconómicos, con el fin de identificar espacialmente inequidades en el acceso a este tipos de servicios ecosistémicos.

Para lograrlo se hace uso de modelos de regresión espacial que capturar fenómenos de agrupamiento y dispersión en los patrones espaciales a través la inclusión de términos de autocorrelación espacial en la regresión lineal. Los modelos espaciales de regresión se prueban con dos tipos de matrices para observar el efecto de la topología de interacción entre las variables y los resultados del ajuste. Se hace uso de gráficos estadísticos, mapas de temáticos y de LISA para la identificación de zonas con niveles deficientes en el acceso.

Los resultados muestran que existen inequidades explicadas por variables de estatus como el acceso a educación superior, que están además negativamente correlacionadas con el porcentaje de afrocolombianos en un sector censal urbano (SU). En relación al acceso a espacios verdes (EV), no existe evidencia fuerte de que las variable de etnicidad o estatus sean buenos predictores del acceso. Sin embargo, sí se encontró una alta concentración del área del EV disponible en muy pocos sectores de la ciudad.



# Abstract

This project makes use of the urban arboreal census carried out in Santiago de Cali (Colombia) between 2014 and 2015, the data of the 2005 population census and the data of the ecological structure of the municipality of 2014. The purpose is to identify if there are spatially significant patterns that provide evidence of biases in the indicators of access to environmental services of urban trees (AU) and green spaces (EV) with socioeconomic indicators, in order to spatially identify inequities in access to these types of ecosystem services.

To achieve this, spatial regression models are used to capture grouping and dispersion phenomena in spatial patterns through the inclusion of terms of spatial autocorrelation in linear regression. The spatial models are tested with two types of matrices to observe the effect of the topology in the interaction between the variables and the results of the adjustment. Statistical graphs, thematic and LISA maps are used to identify areas with poor levels of access and discuss explanation related to the causes.

The results show that there are inequities explained by status variables such as access to higher education, which are also negatively correlated with the percentage of Afro-Colombians in an urban census sector (SU). In relation to access to green spaces (EV), there is no strong evidence that the variables of ethnicity or status are good predictors of access. However, there was a high concentration of the EV area available in very few sectors of the city.



# **Capítulo 1**

## **Introducción**

### **1.1. Antecedentes**

Los árboles son pieza clave de los ecosistemas donde la vida humana ha prosperado. Son hogar y fuente de alimento de muchas especies (Osorio and Molina, 2009); forman espacios con condiciones climáticas y funcionales que complejizan el paisaje y las posibles relaciones entre los animales (Chapman and Onderdonk, 1998). Se puede ubicar la década de los 70 el inicio de un pensamiento ambiental que empieza a ser relevante en el discurso económico mundial (Leff, 2012) y que se consolida con la publicación del Informe Brundtland en 1987. Sin embargo, las preocupaciones sobre la sostenibilidad y conservación de los ecosistemas que sustentan la vida en el planeta son relativamente nuevas en la economía mundial, si la comparamos con la simbiosis entre árboles y humanos que dan origen a la agricultura y lo convierten en un elemento simbólico de gran riqueza en el universo religioso y cultural de la humanidad (León Calle and others, 2011).

El hombre ha materializado espacios urbanos con dimensiones que retan la imaginación y llevan al límite los sistemas de infraestructura, abastecimiento y gobernabilidad. En la empresa de consolidar antroposferas, las ciudades que construimos han desplazados muchos de los ecosistemas naturales de los territorios que fueron la razón de escoger justamente esos

sitios para el asentamiento, trasladándolos más allá de los límites de la ciudad, atenuando su presencia/visibilidad en el mundo de los ciudadanos. Son reemplazados por vías, zonas verdes, áreas industriales, comerciales y residenciales(Azócar et al., 2007).

Las preocupaciones sobre el crecimiento de la población mundial, su concentración en centros urbanos y las transformaciones que trae consigo el proceso de cambio climático, nos obligan a pensar en cómo maximizar los beneficios que nos brindan las zonas verdes y el bosque urbano como estrategia para mitigar los efectos negativos de estos procesos (Nesbitt and Meitner, 2016; Laredo and Mirtha, 2011). A este escenario se suma trabajos como (Nowak and Greenfield, 2012) que revelan patrones de decaimiento estadísticamente significativos del arbolado urbano en 17 de 20 ciudades norteamericanas o (Restrepo et al., 2015) que reporta la reducción de las condiciones de vitalidad del arbolado del Valle de Aburra en Colombia derivado de la interacción de causas naturales y antrópicas, las cuales afectan directa o indirectamente la fisiología y salud de los árboles en los espacios urbanos. Nowak and Greenfield (2012) se pregunta si los administradores locales conocen los cambios que presenta las coberturas arbóreas, puesto que esta es una representación simple pero confiable (y ampliamente aceptada) para tasar la extensión de los beneficios derivados de los bosque urbanos, dado que los servicio que proveen los árboles están relacionados con la salud y el funcionamiento de sus hojas.

El desarrollo urbano de Cali se fue dando en forma de barrios. Cali entra al siglo XX con unos de 25000 habitantes y una 100 hectáreas (Vásquez, 1990); en 2005 tiene una población 2200000 habitantes según el Censo de Población del 2005 y una área urbana 12.101.326 hectáreas. Durante todo este siglo experimenta un rápido crecimiento demográfico, rápida expansión y crecimiento de la vivienda por las migraciones, causadas en gran medida por los fenómeno de desplazamiento por conflictos en las áreas rurales y el fortalecimiento de la industria y el empleo en las ciudad (Vásquez, 1990). Este contexto de crecimiento vertiginoso y no planificado, y las transformaciones del paisaje e intervención de los ecosistemas coincide con la necesidades y preocupaciones para garantizar un ambiente saludable en la ciudad de Cali.

La Organización Mundial de la Salud (OMS) recomienda que una ciudad tenga un árbol por cada tres habitantes, incluyendo espacios públicos y privados. Cali debería tener una población arbórea cercana a los 800.000 árboles, pero de acuerdo con los resultados del censo arbóreo de 2014–2015, se han contabilizado cerca de 290.000 árboles solo en el espacio público. La siembra de 20.000 árboles en la zona urbana, en las comunas que muestran temperaturas por muy por encima de la media (las comunas 3, 4, 5, 8 y 13) propuesta en (CIAT, 2015b) contrasta con los 100.000 árboles que tiene como meta en Plan de Desarrollo Municipal 2016-2019; ambas cifras están lejos de impactar el déficit arbóreo de 500.000 individuos; urge la reforestación, “*sin embargo, la ubicación de los sitios de siembra ha sido uno de los principales inconvenientes del proyecto, ya que no existen zonas de espacio público adecuadas o disponibles para tal fin.*” (CIAT, 2015b, p. 19)

En las agendas municipales, a nivel mundial, ha crecido la importancia de las relaciones y patrones de distribución de los beneficios de áreas verdes en comparación con la distribución espacial de variables sociales y económicas como el ingreso, acceso al trabajo, la etnicidad o el género, con miras a reducir las desigualdades entre los ciudadanos en el acceso y disfrute de los servicios ambientales. La definición y valoración de estos beneficios hace uso de medidas como la abundancia, la cobertura de las copas de los árboles, índices de vegetación, diferentes distancias a las zonas verdes dentro de un marco alineado con conceptos como la justicia ambiental, equidad y la sostenibilidad. Estos indicadores son calculados con datos de censos de población, encuesta de calidad de vida, censos arbóreos, imágenes satelitales, cartografías y bases de datos de entidades oficiales y académicas. Los servicios ambientales tienen cargas y costos de mantenimiento para la administración y gestión de los recursos y servicios ambientales, lo que exige que se identifique las zonas, condiciones de los recursos y de la población para la ejecución de acciones eficaces y eficientes por parte de los gobiernos y autoridades ambientales.

La distribución espacial equitativa de los beneficios que proveen el arbolado urbano y las zonas verdes de espacio público, que constituyen un bien común, financiado y de respon-

sabilidad de las administraciones municipales<sup>1</sup>, es un componente cuantificable por medio métodos estadístico y técnicas de análisis espacial con miras a construir evidencia sobre el disfrute y acceso a los beneficios ambientales en espacios urbanos y su relación con las condiciones de vida de la población que habita ese territorio.

## **1.2. Objetivos**

### **1.2.1. Objetivo general**

Identificar y analizar espacialmente la existencia de inequidades en el acceso a dos servicios ambientales urbanos: el acceso a espacios verdes y beneficios del arbolado, en relación con variables socioeconómicas de la población y aspectos relacionados con el uso y estructura física de las unidades geográficas censales en la zona urbana de Santiago de Cali, Colombia.

### **1.2.2. Objetivos específicos**

- Generar métricas de acceso a espacios verdes y del nivel de beneficios del arbolado urbano.
- Identificar y caracterizar las variables sociales, económicas y estructurales para ser relacionadas con acceso a espacios verdes y beneficios del arbolado.
- Modelar y evaluar las relaciones entre los diferentes indicadores ambientales, sociales, económicos y estructurales.

---

<sup>1</sup> Así está expresado en las leyes ambientales que dan forma al Sistema Nacional Ambiental, SINA, (LEY 99 DE 1993) y reglamentan los planes de desarrollo y de ordenamiento territorial (LEY 388 DE 1997), así como la creación de organismos en la estructura municipal (Acuerdo 01 de 1996).

## **1.3. Preguntas de investigación**

Algunas de las preguntas a las que se enfrenta esta investigación son: ¿Cuáles son las zonas que muestran mayores correlaciones negativas entre las variables sociales y la cobertura de copas o el acceso a zonas verdes? ¿Es igual tener acceso a un parque pequeño que a uno grande? ¿Es el acceso a servicios del AU y EV una característica local de los sectores geográficos o se extienden esos beneficios a agrupaciones de sectores urbanos vecinos? ¿Existen aspectos de tipo urbanístico y de la estructura de los sectores urbanos que condicionan el desarrollo del arbolado? ¿Están estos factores de estructura y uso relacionados con las condiciones de vida de la población? ¿Qué tipo de modelos son los más apropiados para capturar la dependencia espacial en los datos, si es que esta existe?

## **1.4. Hipótesis**

### **1.4.1. Beneficios del arbolado urbano**

#### **1.4.1.1. Hipótesis nula**

La distribución espacial de indicadores socioeconómicos y estructurales es uniforme con respecto a la distribución de beneficios del arbolado urbano (AU) en Santiago de Cali.

#### **1.4.1.2. Hipótesis alternas**

La distribución espacial de indicadores socioeconómicos y estructurales son un predictor de la distribución de servicios del arbolado urbano (AU) en Santiago de Cali

#### **1.4.1.3. Predicciones**

- El patrón espacial de acceso a la educación de la población se correlaciona con el acceso a servicios ambientales del AU.
- El patrón espacial de indicadores de discapacidades en la población se correlaciona con el acceso a servicios ambientales del AU.
- El patrón espacial de etnicidad se correlaciona con el acceso a servicios ambientales del AU.
- El patrón espacial de uso de predios se correlaciona con el acceso a servicios ambientales del AU.
- El patrón espacial de características físicas de los predios se correlaciona con el acceso a servicios ambientales del AU.

### **1.4.2. Acceso a espacios verdes**

#### **1.4.2.1. Hipótesis nula**

La distribución espacial de indicadores socioeconómicos y estructurales es uniforme con respecto a la distribución del acceso a espacios verdes (EV) en Santiago de Cali.

#### **1.4.2.2. Hipótesis alternas**

La distribución espacial de indicadores socioeconómicos y estructurales son un predictor de la distribución del acceso a espacios verdes (EV) en Santiago de Cali.

#### **1.4.2.3. Predicciones**

- El patrón espacial de acceso a la educación de la población se correlaciona con el acceso a EV

- El patrón espacial de indicadores de discapacidades en la población se correlaciona con el acceso a EV
- El patrón espacial de etnicidad se correlaciona con el acceso a EV
- El patrón espacial de uso de predios se correlaciona con el acceso a EV
- El patrón espacial de características físicas de los predios se correlaciona con el acceso a EV.

## 1.5. Justificación

La inclusión de los componentes ambientales en el ámbito de la planificación urbana en los planes de ordenamiento territorial que exige la legislación colombiana<sup>2</sup> necesita de la creación de medidas y la elaboración de análisis sobre su relación con las condiciones de vida de la población. Las herramientas para establecer políticas públicas y el seguimiento a las acciones realizadas por las administraciones municipales deben estar asociadas a características medibles y objetivas para su implementación. Se espera que los objetivos y proyectos estén sustentados en estudios científicos que identifiquen brechas y oportunidades para la intervención y mejoramiento de los servicios ambientales de los cuales es responsable el gobierno local. En esta medida este estudio contribuye a la identificación de relaciones de inequidad en la distribución de los beneficios que provee el arbolado urbano a través del análisis espacial de la cobertura arbórea, el acceso a zonas verdes y la distribución de las variables sociales y económicas de la población. El estudio promete ser un punto de partida para la identificación de zonas de intervención del arbolado con el fin de cerrar brechas relacionadas con el desarrollo sostenible y la justicia ambiental.

Contar con los datos del censo arbóreo de Santiago de Cali permite hacer análisis de estos beneficios ambientales para la población usando datos con alta resolución espacial y construir estadísticas a escalas apropiadas para la intervención y aprovechamiento de los recursos naturales de la ciudad explotando el potencial que ofrece la información censal y

---

<sup>2</sup>LEY 388 DE 1997

los conjuntos de datos espaciales de los que dispone la administración municipal (Schwarz et al., 2015). Los resultados de esta investigación buscan aportar al debate académico y enriquecer el proceso de la toma de decisiones y la planificación de la ciudad, sentando bases técnicas y resultados concretos que el desarrollo de políticas, proyectos e instrumentos que potencien al árbol y el acceso a espacio verdes como estrategia para la mejora de la calidad de vida de los caleños.

## 1.6. Alcance

Este trabajo se concentra en describir los patrones espaciales y establecer la correlación entre métricas para representar los beneficios del arbolado urbano(AU) y los espacios verdes (EV) con las variables sociales y económicas de la población e incluye variables de uso de los predios y características que incluyen factores del contexto urbanístico de la ciudad buscando estimar la importancia relativa de las relaciones entre los indicadores ambientales y sociales. El valor explicativo de la posición en el plano geográfico de las métricas ambientales y socioeconómicas permite seleccionar los modelos de regresión apropiados para cuantificar el grado de correlación que existe (Fotheringham et al., 1998). El problema comprende la exploración de las variables sociales, económicas, estructurales y ambientales, el cálculo de indicadores, la cuantificación de la correlación, el modelado de la relaciones entre variables y la identificación de las zonas con acumulación de desigualdades.

El trabajo hace uso del censo arbóreo urbano realizado en Santiago de Cali (Colombia) entre el año 2014 y 2015, de datos del censo de población de 2005 y los datos de la estructura ecológica del municipio. Para ellos se apoya en los aportes de tipo metodológico y estadístico de la literatura especializada sobre modelos de regresión lineal y modelos de regresión espacial. Las unidades de análisis espacial son los sectores censales urbanos (SU) que mantienen una relación geométrica y de escala similar a la de los barrios, unidad básica del crecimiento y desarrollo urbano de la ciudad de Cali.

# **Capítulo 2**

## **Revisión de la literatura**

### **2.1. Servicios ecosistémicos y su valoración**

Podemos definir un servicio ambiental o ecosistémico como los beneficios para la población humana que se derivan directa o indirectamente de los ecosistemas (Bolund and Hunhammar, 1999). Los servicios dependen entonces de los tipos de los ecosistemas con los que cuente el entorno urbano. Bolund and Hunhammar (1999) distinguen 7 tipos de ecosistemas urbanos: árboles de calle, zonas verdes, bosques urbanos, tierras cultivadas, humedales, lagos/lagunas, y ríos/arroyos. Todos ellos en conjunto benefician a la población, y muchos estudios se han encargado de cuantificar el impacto de estos beneficios, en particular se cuentan los que se relacionan con los ecosistemas de árboles de calle y las zonas verdes, como reducción de las temperaturas (Ripoll et al., 2010), reducción de la polución en el aire (Durán Rivera and Alzate Guarín, 2009), secuestro de carbono (Nowak and Crane, 2002; McPherson et al., 2013), mitigación de los efectos de calentamiento por gases de efecto invernadero (Laredo and Mirtha, 2011) y mantenimiento del agua en los ecosistemas o proveyendo alimento como en el caso de los árboles frutales (Konijnendijk et al., 2005; Nolazco, 2012), reducción de los niveles de ruido (Bolund and Hunhammar, 1999). Otros estudios argumentan que los beneficios ambientales de los ecosistemas urba-

nos pueden medirse directamente en la salud de la población (Bolund and Hunhammar, 1999; Gómez-Bagethun and Barton, 2013). Una forma de clasificar todos estos beneficios e indagar sobre las medidas usadas las resume Gómez-Bagethun and Barton (2013) en una tabla que reproduczo a continuación (ver tabla 2.1).

Cuadro 2.1: Clasificación de servicios ecosistémicos importantes en zonas urbanas y funciones y componentes subyacentes del ecosistema

Funciones y componentes	Servicio Ecosistémico	Ejemplo	Ejemplo de indicadores/proxies
Conversión de energía en plantas comestibles a través de la fotosíntesis	Suministro de alimentos	Hortalizas producidas por lotes urbanos y áreas periurbanas	Producción de alimentos [toneladas/año]
Percolación y regulación de la escorrentía y la descarga del río	Regulación del caudal de agua y mitigación de escorrentía	El suelo y la vegetación percolan el agua durante eventos de precipitación intensa y / o prolongada	Capacidad de infiltración del suelo; [%] Sellado con respecto a la superficie permeable [ha]

Funciones y componentes	Servicio Ecosistémico	Ejemplo	Ejemplo de indicadores/proxies
Fotosíntesis, sombreado y evapotranspiración	Regulación urbana de la temperatura	Los árboles y otra vegetación urbana proporcionan sombra, crean humedad y bloquean el viento	Índice de área foliar; Disminución de la temperatura [° C] por cobertura arbórea[m <sup>2</sup> ] en parcelas cubierta de árboles
Absorción de ondas sonoras por la vegetación y el agua	Reducción de ruido	Absorción de ondas sonoras por barreras vegetales, especialmente vegetación espesa	Superficie de la hoja [m <sup>2</sup> ] y distancia a las carreteras [m]; Reducción de ruido [dBA] / unidad de vegetación [m]
Filtración y fijación de gases y partículas	Purificación de aire	Eliminación y fijación de contaminantes por la vegetación urbana en hojas, tallos y raíces	O <sub>3</sub> , SO <sub>2</sub> , NO <sub>2</sub> , CO y PM10 m removido en [toneladas/año] multiplicado por la cobertura arbórea (m <sup>2</sup> )

Funciones y componentes	Servicio Ecosistémico	Ejemplo	Ejemplo de indicadores/proxies
Barrera física y absorción en energía cinética	Moderación de los extremos ambientales	Tormentas, inundaciones y amortiguación de olas por barreras vegetales; Absorción de calor durante olas de calor severas	Cubrir la densidad de las barreras de vegetación que separan las áreas construidas del mar
Eliminación o descomposición de nutrientes xénicos	Tratamiento de desechos	Filtración de efluentes y fijación de nutrientes por humedales urbanos	P, K, Mg y Ca en mgkg-1 en comparación con las normas de calidad del suelo y del agua
Secuestro y fijación de carbono en la fotosíntesis	Regulación climática	Secuestro y almacenamiento de carbono por la biomasa de arbustos urbanos y de tres	Secuestro de CO2 por árboles (carbono multiplicado por 3.67 para convertir a CO2)

Funciones y componentes	Servicio Ecosistémico	Ejemplo	Ejemplo de indicadores/proxies
Movimiento de los gametos florales por la biota	Polinización y dispersión de semillas	El ecosistema urbano provee hábitat para aves, insectos y polinizadores	Diversidad de especies y abundancia de aves y abejorros
Ecosistemas con valores recreativos y educativos	Recreación y desarrollo cognitivo	Los parques urbanos ofrecen múltiples oportunidades para la recreación, la meditación y la pedagogía	Superficie de los espacios públicos verdes [ha / habitante (o cada 1000 habitantes)]
Disposición del hábitat para las especies animales	Avistamiento de animales	El espacio verde urbano proporciona un hábitat para las aves y otros animales a los que les gusta ver	Abundancia de aves, mariposas y otros animales valorados por sus atributos estéticos

Existen diferentes perspectivas para la evaluación de los servicios ecosistémicos, en razón de ellos se crean diferentes indicadores, métodos estadísticos y metodologías para capturar de forma directa o indirecta los beneficios ambientales. (Gómez-Baggethun and Barton, 2013) propone 2 tipos de valoración: la económica y la sociocultural. La valoración socio-cultural se enfoca en la percepción y preferencias de los ciudadanos que están ligadas a

sus costumbres y sistemas de valores. Resalta lo difícil de medir y que suele ser mejor abordado por instrumentos cualitativos, construcción de escalas y el uso de narrativas. Un ejemplo de este tipo de trabajos es (Garzón et al., 2004) se hace referencia a la vegetación en las ciudades y su incidencia en la vida de las personas, sobre todo en aquellas comunidades de menores recursos. “Los árboles están estrechamente vinculados a la historia de la humanidad; están cargados de historia, son testigos de varias generaciones, que enlazan su existencia con la anterior, constituyéndose en documentos vivos de acontecimientos culturales.” (Ferro Medina, 2010). (Konijnendijk et al., 2005) propone además que se valore lo ambiental, la biodiversidad y sostenibilidad por su carácter fundamental en la existencia misma del los ecosistemas.

Las metodologías de valoración económica se basan principalmente en análisis costo beneficios que buscan un aprovechamiento eficiente del uso del suelo (Bolund and Hunhammar, 1999). La inversión en el arbolado urbano y las zonas verdes arroja resultados positivos consistentemente en la literatura p.e (McPherson et al., 1997) reporta que los beneficios exceden los costos en ciudades como Chicago y en Adelaide (Australia) según (Killicoat et al., 2002) cada árbol da beneficios por AUD\$172.

Algunos análisis no se limitan a evaluar el costo de ahorro o el dispuesto a pagar por la población, para valorar económicoamente los servicios ecosistémicos se basan en modelos biofísicos de los individuos arbóreos e incluyen variables ambientales, características de los suelos, la infraestructura de las zonas y sus habitantes (Nelson et al., 2009) p.e el modelo UFORE (Nowak and Crane, 2000) o CBAT (McPherson et al., 1997) , y que permiten tomar acciones específicas sobre el tipo de vegetación y su distribución. Sin embargo es importante tener en cuenta que este tipo de análisis pueden generar incentivos para la conversión indeseable de ecosistemas urbanos en infraestructura construida, con la consiguiente pérdida de servicios de los ecosistemas (Gómez-Baggethun and Barton, 2013).

La lógica económica de los servicios ecosistémicos puede conducir también a incentivar procesos paradójicos, como el incrementos de los precios de las casas y arriendos que

derivan en procesos de gentrificación y desplazamiento de la población que fue beneficiada por las estrategias de mejoramiento de EV y AU con el propósito de resolver problemas relacionados con la justicia ambiental. (Wolch et al., 2014)

## 2.2. La perspectiva de la justicia ambiental

La justicia ambiental es un concepto que ha evolucionado desde su aparición en la década de los 1980 a través de organizaciones dedicadas al activismo ambiental y las redes que ellas forman en conjunto con la academia, acuñaron conceptos de ecología política como justicia ambiental, deuda ecológica, epidemiología ambiental, racismo ambiental, justicia climática, soberanía alimentaria, y responsabilidad ambiental empresarial, que han sido adoptados también por académicos y por tomadores de decisiones (Martinez Alier et al., 2014; Cerdà, 2011).

La planeación urbana, las relaciones entre los ciudadanos con los espacios públicos y con las instituciones que los rigen son la base de la idea de justicia ambiental propuesta en (Low, 2013), y que tiene 3 componentes que la definen: *i)* la justicia distributiva, que en términos de espacio público se basa asegurar disponibilidad y acceso equitativo de los espacios y servicios a los ciudadanos; *ii)* la justicia procedural o procesal, que se refieren a los procesos de negociación y toma de decisiones, en concreto a la percepción de los individuos sobre qué tan justos y equilibrados son, y por tal motivo más dispuesto a aceptar los resultados aunque no les favorezca; *iii)* la justicia interaccional, que refiere al comportamiento y trato de las personas en el espacio público que configuren comportamiento violento o discriminatorios sobre grupos de la población. El autor argumenta que las condiciones ambientales provocadas por los procesos de urbanización y/o contaminación deben analizarse pensando en satisfacer las tres dimensiones, de lo contrario no es posible hablar de justicia. (Schlosberg, 2013) lleva la reflexión un poco más lejos, argumentando que las nuevas extensiones de la justicia ambiental se han movido del discurso a un nuevo dominio, donde lo natural y ambiental crean las condiciones para la justicia social.

En (Braverman, 2008) se exploran las implicaciones entre las intervenciones en el arbolado urbano y el control de fenómenos como el crimen y la gobernabilidad dada las relaciones afectivas y morales de la población con los árboles. El uso de los árboles puede verse también una forma de discriminación y de discurso político o tecnología de gobierno. El uso de zonas verdes y arboles ha sido usado también como una forma de simbolica de estatus y de poder, y esta afirmación es consistente que la tendencia en tener distribuciones inequitativas (Braverman, 2008).

La perspectiva distributiva de la justicia ambiental busca relacionar entonces métricas usadas para cuantificar los servicios ecosistémicos con métricas sobre la población y sus condiciones de vida usando unidades espaciales para caracterizar su comportamiento en el área de estudio.Típicamente se usan variables ambientales que representan aspectos biofísicos de los ecosistemas p.e superficie de la hoja, índice de área foliar, el área de cobertura de la copa o los efectos directos e indirectos de los ecosistemas sobre variables climáticas p.e temperatura o humedad, o físico-químicas para representar la composición del suelo y del aire, o mediciones de la capacidad de secuestrar carbono de los árboles y la de filtrar agua del suelo.

Los primeros trabajos enmarcados en la justicia ambiental se enfocaron en la ubicación de plantas de residuos y manejo de desechos relacionándolos con variables sociales como la etnicidad e indicadores de segregación racial o en comunidades con bajos ingresos (Heynen et al., 2006) y que desarrollan trabajos como (Chakraborty and Armstrong, 1997; Cutter et al., 1996). Posteriormente con el avance de modelos y tecnologías de la información para la caracterización de la infraestructura ecológica urbana y la cuantificación de servicios ambientales se desarrollan metodologías para establecer relaciones entre distribuciones desiguales adoptando el uso de la cobertura de copa de los árboles como variable que se consolida para este tipo de estudios, con variables socioeconómicas principalmente de raza/etnicidad de la población, que tradicionalmente ha sido una preocupación en el estudio de las desigualdades sociales (Heynen et al., 2006; Landry and Chakraborty, 2009; Phelps, 2012; Zhou and Kim, 2013; Schwarz et al., 2015; Watkins et al., 2016).

De los espacios verdes suelen caracterizarse dimensiones sobre el acceso, su dimensiones físicas, los equipamientos y servicios que prestan, y si los beneficiarios de son del ámbito local al EV o más amplio. Igualmente es importante preguntarse por la calidad del parque y zonas verdes así como los usos, que pueden variar dependiendo de las comunidades que están disfrutando del parque (Kabisch and Haase, 2014). El acceso suele ser un concepto complejo de medir, pues el análisis espacial de los datos arroja variaciones significativas usando diferentes métricas que proviene de los diferentes conceptos de acceso usados para su definición (Talen and Anselin, 1998).

En Cali se han realizado trabajos que caracterizan procesos de desigualdades entre la población analizando factores como la segregación racial, brechas salariales, empleabilidad e índices socioeconómicos de segregación espacial(Arroyo Mina et al., 2016; Mora and Arcila, 2014; Cerón and Escobar, 2014). Aunque todos ellos incluyen una dimensión espacial, en tanto que usan las comunas como unidades de agregación de las variables socioeconómicas, su análisis no hace uso de los datos espaciales de los objetos geográficos, su base teórica son los modelos econométricos de regresión de mínimos cuadrados o la creación de índices (escalas) para clasificar las unidades gráficas con la escala p.e (Cerón and Escobar, 2014) que hace uso del escalograma de Guttman, un método para normalizar las diferentes escalas usando igual número de rangos discretos y construir un índice acumulativo de todas las dimensiones. Mora and Arcila (2014) estudian la discriminación, tanto racial como por sexo a través de los modelos econométricos de Oaxaca-Blinder usado para analizar las diferencias salariales entre dos grupos usando una función que tiene un componente de discriminación. Para incluir en la ecuación la procedencia se crean los conglomerados a priori y se agregan dummies por conglomerado de comunas en la ecuación. Arroyo Mina et al. (2016) evalúan la calidad del empleo bajo el supuesto de ser buen proxy de la calidad de vida y encuentra que en Cali existe evidencia de que las poblaciones afro son discriminadas laboralmente, y que esta discriminación está explicada por el lugar de residencia. Sin embargo, el método usado es similar al de (Mora and Arcila, 2014) e inclusive usa los mismo conglomerados para las variables dummies de la procedencia. Entre los trabajos sobre segregación racial el de (Pacheco, 2013) si hace uso de modelos de regresión espacial

y logra identificar zonas homogéneas con alta correlación espacial. Además describe los cambios en la segregación residencial en Cali entre 1993 y 2005 mostrando que la segregación racial es un fenómeno que se ha mantenido vigente en una ciudad cuya población afrocolombiana es de 26,2 % del total.

Los estudios que se han llevado a cabo sobre segregación espacial muestran que la exclusión de grupos étnicos en Cali tiende a coincidir espacialmente con la segregación de los grupos socioeconómicos de estratos bajos (Cerón and Escobar, 2014). Tendencia que también se ven estudios de ciudades norteamericanas (Heynen et al., 2006; Landry and Chakraborty, 2009; Zhou and Kim, 2013; Nesbitt and Meitner, 2016).

## **2.3. Modelamiento y análisis espacial de variables ambientales y sociales**

Las fuentes de datos con información espacial usada para capturar variables ambientales relacionadas con la vegetación provienen principalmente de imágenes satelitales (Landry and Chakraborty, 2009; Troy et al., 2007; Vásquez Fuentes and Romero Aravena, 2008; Nesbitt and Meitner, 2016), imágenes aéreas (Azócar et al., 2007; Heynen et al., 2006; Tratalos et al., 2007), datos de Lidar (Shanahan et al., 2014; Schwarz et al., 2015) e inventarios producidos por muestras o censos arbóreos y de espacios verdes (Comber et al., 2008; Killicoat et al., 2002; Nowak and Crane, 2000, Nowak and Crane (2002); Talen and Anselin, 1998). Las imágenes satelitales son usadas en gran cantidad de los estudios dadas su creciente disponibilidad y la frecuencia de actualización —Landsat 8 revisita un mismo punto sobre la superficie de la tierra cada 16 días con un desfase de 8 días con respecto al satélite Landsat 7—, lo que permite hacer monitoreo y seguimiento a escalas entre los 15 m a 100 m por ancho de píxel. Los indicadores de cobertura calculados con base en datos de imágenes satelitales son estimados usando la escala de resolución de la imagen en combinación con el índice de Vegetación de Diferencia Normalizada (NDVI) que permite

diferenciar entre densidad y tipo de vegetación leñosa o vegetación herbosa (Nesbitt and Meitner, 2016).

El uso de inventarios arbóreos permite estudios más detallados sobre las características del arbolado y son usados para evaluar la salud y estructura de los individuos arbóreos y las capacidades específicas de las especies para proveer servicios ambientales. (Killicoat et al., 2002; Nowak and Crane, 2002; Cowett, 2014).

Entre los distintos indicadores desarrollados para capturar la extensión y distribución de los servicios ambientales la cobertura de copas ha probado ser sensible y eficaz para cuantificar hasta qué punto los árboles y bosques están proporcionando servicios críticos a los residentes (Nowak et al., 2010). Se usan otros indicadores además de la cobertura en la literatura, y en muchas ocasiones se normalizan los valores de las variables ambientales por unidad de área, usando unidades geográficas definidas e introduciendo métricas sobre la densidad y cantidad de población beneficiada en el cálculo. De hecho (Cowett, 2014) propone que para analizar con precisión la distribución espacial de los árboles de las calles y los beneficios que proporcionan es importante migrar hacia métricas en las que las especies arbóreas y el tamaño del árbol sean un factor en el cálculo, pues la mayoría de los servicios de los ecosistemas arbóreos son proporcionales a la cantidad de área superficial de la hoja; en esta medida las especies de árboles de mayor estatura típicamente proporcionan muchos más beneficios que las especies de menor estatura. Trabajos como (Alanís et al., 2014) usan indicadores ecológicos de las especies como abundancia, dominancia y frecuencia para construir índices de importancia para valorar las especies nativas y estructura de los bosques urbanos, cuya conservación también hace parte de las metas de manejo del AU (Nowak et al., 2010).

La tabla 2.2 resume los indicadores para cuantificar servicios/beneficios y estado de los ecosistemas arbóreos usados en la literatura revisada.

Cuadro 2.2: Métricas para caracterizar servicios del AU

Métrica	Referencia
Producción de alimentos [toneladas/año]	(Gómez-Baggethun and Barton, 2013)
Índice de área foliar	(Gómez-Baggethun and Barton, 2013)
Disminución de la temperatura [° C] por cobertura arbórea[m <sup>2</sup> ] en parcelas/sitios cubiertas de árboles	(Gómez-Baggethun and Barton, 2013)
Superficie de la hoja [m <sup>2</sup> ]	(Gómez-Baggethun and Barton, 2013)
O <sub>3</sub> , SO <sub>2</sub> , NO <sub>2</sub> , CO y PM10 m removido en [toneladas/año] multiplicado por la cobertura arbórea (m <sup>2</sup> )	(Gómez-Baggethun and Barton, 2013)
Secuestro de CO <sub>2</sub> por árbol (carbono multiplicado por 3.67 para convertir a CO <sub>2</sub> )	(Gómez-Baggethun and Barton, 2013)
Abundancia (por especie) [individuos]	(Alanís et al., 2014)
Densidad [individuos/m <sup>2</sup> ]	(Nowak et al., 2010)
Cobertura de copa de árbol por persona[m <sup>2</sup> /persona]	(Nowak et al., 2010)
Cobertura de copa de árbol[m <sup>2</sup> ]	(Nowak et al., 2010) y muchos antes también
porcentaje de cobertura de copa de árbol en un área	(Nowak et al., 2010)
Dominancia de una especie en función a la cobertura de copa [m <sup>2</sup> /ha]	(Alanís et al., 2014)
Dominancia relativa de la especie i respecto a la dominancia total	(Alanís et al., 2014)

Métrica	Referencia
Frecuencia absoluta de una especie (porcentaje de presencia en los sitios)	(Alanís et al., 2014)
Abundancia relativa de la especie i respecto a la abundancia total en el área de estudio	(Alanís et al., 2014)
Árboles por habitante	POT2014
índice de cobertura de copa de árbol alrededor de un punto de muestreo	(Zhou and Kim, 2013)
Porcentaje de cobertura de copa de árbol en servidumbres y espacios públicos	(Landry and Chakraborty, 2009)

En cuanto a los espacios verdes se valoran típicamente los aspectos geométricos y el acceso. Los aspectos físicos se caracterizan con medidas de superficie de los espacios, el porcentaje de área o número de espacios respecto de las unidades espaciales del análisis; este tipo de medidas son llamadas por (Talen and Anselin, 1998) la aproximación contenedor (como el representado en la ecuación (2.1)), y aunque muy extendidas en la literatura para caracterizar acceso a los EV asumen que los beneficios del espacio verde tiene un impacto local: “*Sin embargo, si el investigador está seguro de que la esfera de influencia de un servicio dado se limita a un límite geográfico específico, puede seguir siendo apropiado el enfoque unidimensional tradicional de la accesibilidad por medio de conteos por unidad*” (Talen and Anselin, 1998). En esta misma línea, (Kabisch and Haase, 2014) asegura que aunque el secuestro de carbono tiene beneficios a nivel de toda la ciudad, algunos procesos biofísicos que rigen los beneficios de los EV ocurren a nivel local. Aunque no es fácil definir la escala a la que operan los procesos biofísicos, los beneficiarios de los servicios ambientales son a menudo aquellos que viven cerca de los EV.

Respecto de las unidades espaciales de agregación de las variables (Cutter et al., 1996) encontró que el coeficiente de correlación tiende a aumentar con el aumento de la escala de las unidades espaciales de agregación. Para verificar la robustez de las estimaciones de

correlación entre la cantidad de instalaciones tóxicas/desperdicios e indicadores socioeconómicos Cutter et al. (1996) usó pruebas de hipótesis para probar que tanto explican los indicadores de etnia y de ingreso la presencia de las instalaciones, agregando los datos a diferentes escalas de la cartografía censal en EE.UU (condado > sectores > bloque). Los resultados sugieren que los sectores y bloques son la escala espacial más apropiada para evaluar las desigualdades debido a la alta variabilidad de los indicadores socioeconómicos dentro del condado en comparación con los sectores y los bloques.

Estos hallazgos ponen en cuestión la selección del tamaño de las unidades espaciales del análisis en relación con el área de estudio. Para el caso de Cali se suele usar las comunas como unidad espacial, pues es la unidad sobre la que se define la inversión, pero puede no representar muy bien la escala de los procesos que dominan las transformaciones en el AU y el acceso a los espacios verdes cuando se presenta alta variabilidad de los indicadores socioeconómicos y ambientales.

En esta línea es interesante analizar variables que capturan aspectos físicos de la estructura del espacio disponible. (Landry and Chakraborty, 2009) relaciona las coberturas con las áreas de los sectores censales completos y los compara con las áreas de espacio público sin los lotes de las manzanas de cada sector censal, introduciendo un indicador espacial que apunta a espacios donde la responsabilidad es claramente de las instituciones estatales por el énfasis en el área de jurisdicción para la intervención de la estructura arbórea y de los espacios verdes (ver tabla 2.2).

Además de las medidas de superficie y sus variantes para caracterizar los beneficios en una unidad espacial definida, se usan medidas que relacionan origen destino, como la distancia mínima del centroide de la unidad espacial al borde del EV más cercano, que proporciona una estimación fiable de la distancia media desde cualquier punto dentro de una unidad de análisis. Este método trata a todos los espacios verdes por igual, independientemente del tamaño y la propiedad (Nesbitt and Meitner, 2016).

Para medir el acceso se pueden usar medidas de distancia, que operan bajo el concepto de

costo de viaje. Una forma es calcular la suma de las distancias desde el centroide de una unidad espacial a **todos** los EV, como lo hace la ecuación (2.2) y su variante (2.3) que divide el costo de viaje entre el número de EV (Talen and Anselin, 1998). Una alternativa a la suma de distancias es usar la distancia del centroide al espacio verde más cercano (ecuación (2.4)). Otra métrica relevante de acceso es la distancia de red a través de la estructura de las vías de la ciudad o distancia a pie, que produce una versión más realista de la experiencia de acceso, y puede usarse para calcular la distancia mínima promedio de puntos aleatorios o para cada manzana, por ejemplo, dentro de una unidad de espacial. Estas métricas exigen marcar los puntos de acceso a las zonas verdes, tarea que se realiza de forma no automática (Zhou and Kim, 2013).

*índice contenedor*

$$A_i^C = \sum_j s_j \quad \forall j \in I \quad (2.1)$$

*costo de viaje*

$$A_i^T = \sum_j d_{ij} \quad (2.2)$$

*costo de viaje normalizado*

$$\bar{A}_i^T = \sum_j d_{ij}/N \quad (2.3)$$

*distancia mínima*

$$A_i^M = \min |d_{ij}| \quad (2.4)$$

*índice de accesibilidad a pie*

$$A^W = \sum_{\int R_b} (r_{min}/d_j) \quad \forall d_j > 0, \quad r_{min} < R_b \quad (2.5)$$

El cálculo de estos indicadores e índices de acceso se ha complejizado con base en recomendación de las agencias ambientales que definen valores de referencia como la cantidad de EV mínimo disponible ( 2 ha ) en un radio determinado (300m). Esto puede lograrse, por ejemplo, con la suma de los cocientes entre el radio recomendado y la distancia de un punto de muestreo dentro de la unidad espacial a cada uno de las EV en el radio de búsqueda  $R_b$ , que suele ser mayor que la distancia de acceso recomendada  $r_{min}$ . Se pueden seleccionar un número de muestras para cada unidad espacial y sumar o promediar las distancias obtenidas (Zhou and Kim, 2013) o usar el centroide como único punto (ecuación (2.5)). Modificaciones a este índice pueden incluir el área de la zona verde para cuantificar el acceso y las características del EV al que se accede (Comber et al., 2008).

El tipo de espacios usualmente aceptados como espacio verde urbano en los estudios incluyen parques y jardines públicos, corredores verdes (por ejemplo, adyacentes a ríos y canales), reservas naturales locales, áreas comunes, pequeñas áreas de bosques con sotobosque, sitios de importancia para la conservación de la naturaleza, áreas de drenaje (es decir, áreas regularmente inundadas cercanas a los ríos), cementerios, instalaciones desocupadas cubiertas de vegetación, lagos y humedales, campos de golf, áreas privadas de asociaciones de propietarios, terrenos escolares, parcelas y hasta senderos que no son lo suficientemente grandes para calificar como parques(Nesbitt and Meitner, 2016; Comber et al., 2008; Kabisch and Haase, 2014; Zhou and Kim, 2013).

La tabla 2.3 resume los indicadores usados en la literatura revisada sobre el acceso a EV.

Cuadro 2.3: Métricas para caracterizar servicios del EV

Métrica	Referencia
Superficie de los espacios públicos verdes [ha]	varios

Métrica	Referencia
Número de instalaciones o servicios contenidos en una unidad dada (por ejemplo, distrito censal o división político-administrativa)	(Talen and Anselin, 1998)
Potencial de la gravedad	(Talen and Anselin, 1998)
Distancia al EV más cercano[m]	(Talen and Anselin, 1998)
Coste medio de viaje EV más cercano	(Talen and Anselin, 1998)
Superficie de los espacios públicos verdes por habitante (o cada 1000 habitantes)[ha/habitantante]	
Cobertura de espacios verdes[ %]	varios
Distancias entre puntos o nodos en la red de movilidad urbana	(Comber et al., 2008)
Índice de accesibilidad caminando al parque dentro de un radio	(Zhou and Kim, 2013)

Para la caracterización de las condiciones de vida de la población en relación con el disfrute a servicios ambientales los indicadores sociales usados suelen provenir de los censos de población, encuestas de trabajo o calidad de vida, registros catastrales para el avalúo de predios y uso de los suelo. La tabla 2.4 resume algunos de los indicadores socioeconómicos usados para la evaluación de inequidades ambientales en la revisión de la literatura realizada.

Cuadro 2.4: Métricas para caracterizar aspectos socioeconómicos de la población y estructura de las unidades espaciales del EV

Métrica	Referencia
Ingreso medio del hogar	varios

Métrica	Referencia
Valor medio vivienda ocupada por el propietario	(Cowett, 2014)
Porcentaje de viviendas ocupadas	(Cowett, 2014)
Porcentaje de población blanca	(Cowett, 2014)
Porcentaje de personas con pregrado	(Cowett, 2014)
Porcentaje de población afro	[Cowett (2014)]
Porcentaje de viviendas arrendadas	(Heynen et al., 2006)
Densidad de población [residentes/ha]	(Troy et al., 2007)
Densidad de hogares[hogares/ha]	(Troy et al., 2007)
Valor medio vivienda ocupada	(Troy et al., 2007)
Porcentaje de viviendas desocupadas	(Troy et al., 2007)
Índice de delincuencia	(Troy et al., 2007)
Porcentaje de viviendas que son viviendas unifamiliares	(Troy et al., 2007)
Tamaño medio del hogar	(Landry and Chakraborty, 2009)
Tamaño medio del lote de vivienda	[Shanahan et al. (2014)]
Densidad de viviendas	
Grados escolares cursados[años]	
Porcentaje de habitantes con grado profesional	(Nesbitt and Meitner, 2016)
Porcentaje de habitantes con maestría	(Nesbitt and Meitner, 2016)
Edad del barrio	
Edad media de las construcciones	(Zhou and Kim, 2013)

### 2.3.1. Enfoque estadístico

El instrumento matemático más popular para establecer relaciones de dependencia y asociación entre dos variables aleatorias son los índices de correlación, sin que la relación cuantificada sea necesariamente de causalidad. La causalidad entre las variables suele ser una apuesta del investigador y de su conocimiento o intuición sobre los procesos que dominan las características de las variables aleatorias (Gibbons and Overman, 2012). El coeficiente de correlación de Pearson es medida usada para caracterizar la fuerza de relaciones lineales, mientras que coeficientes de correlación como el de Spearman son usados para encontrar relaciones no necesariamente lineales, pues se calcula observando si los cambios en una de las variables está relacionado con incrementos o decrementos de la otra, lo que permite cuantificar dependencias no lineales. Estos coeficientes son análisis eficaces para seleccionar variables candidatas a ser incluidos en modelos de regresión, tanto para seleccionar las variables que tienen relación fuerte con la variable a predecir, como la independencia lineal de las variables explicativas, condiciones necesarias para que los métodos de ajuste de parámetros de los modelos lineales (Gibbons and Overman, 2012). Los coeficientes de correlación, como estadística de resumen, no pueden reemplazar el examen visual de los datos y la construcción de relaciones que tengan un fundamento teórico.

Los modelos de regresión lineal son ampliamente usados en la econometría para construir modelos explicativos con términos que describen componentes teóricos de los procesos de estudio e inferir el comportamiento o tendencias en una población con base en una muestra. De esta forma se pretende cuantificar el cambio de la variable dependiente ( $y$ ) ante aumentos o disminuciones del valor de una de las  $p$  variables dependientes  $x_j$  donde  $j = 1...p$ . Cada una de las observaciones  $i$ ,  $i = 1...n$  del vector  $X_p$  denominadas  $x_{ij}$  forman un sistema de ecuaciones lineales donde el parámetro  $\beta_j$  pesa el aporte explicativo de la variable correspondiente al estimarlo (ecuación (2.6) y (2.7)). Así, los coeficientes estimados representan la importancia de una variable independiente en los cambios de la dependiente. Los métodos para resolver son estimadores de mínimos cuadrados (OLS ordinary least-square) o de máxima verosimilitud (ML maximum likelihood). Si se cumple que las variables in-

dependientes del modelo y los términos del error tienen una distribución normal y una varianza constante (homocedasticidad), entonces el estimador es óptimo entre todos los estimadores no sesgados lineales y no lineales. Es clave indagar el cumplimiento de las condiciones de normalidad, en especial en los residuos. (Gibbons and Overman, 2012).

modelo lineal de la i-esima observación

$$y_i = \beta_0 + \beta_1 \cdot x_{i1} + \dots + \beta_j \cdot x_{ij} + \dots + \beta_p \cdot x_{ip} + e_i \quad (2.6)$$

forma matricial del sistema de ecuaciones a resolver

$$Y_{n \times 1} = X_{n \times p} \beta_{p \times 1} + e_{(n \times 1)} \quad (2.7)$$

Estudios como (Heynen et al., 2006; Vásquez Fuentes and Romero Aravena, 2008) usan coeficientes de correlación entre los pares de variables ambientales y socioeconómicas agregadas en unidades censales o administrativas para luego usar modelos de regresión de lineal con el fin de estimar la importancia explicativa de las variables a través de la estimación de los coeficientes.

Tratalos et al. (2007) indaga sobre relaciones entre indicadores ambientales y el estatus social de los residente en cinco ciudades del Reino Unido; se usa coeficientes de correlación de Spearman para seleccionar las variables usadas en un modelo de regresión lineal usando OLS para probar que los desarrollos urbanos de alta densidad generalmente se asociaban con un pobre desempeño ambiental, medido por el tamaño del parche en el espacio verde y los niveles de provisión de servicios ambientales.

### 2.3.2. Enfoque geoestadístico

Cuando se analiza los residuos  $e_i$  en un regresión lineal estos debería ser ruido. Sin embargo si se encuentra que existe algún tipo de relación de los residuos con la posición de las observaciones  $i$  en plano geográfico, entonces se puede usar variaciones en el modelo lineal, que consiste en incluir un término lineal usando la variable dependiente (modelo autoregresivo SAR, ecuación (2.8)) , las independientes ( spatial lag o retardo espacial en  $X$  SLX, ecuación (2.10)), en el error (modelo espacial del error SEM,ecuación (2.9)) o usando una combinación del modelo de error y autoregresivo (modelo espacial de Durbin SD,ecuación (2.11)). Todas estas aproximaciones introducen una matriz de  $W_{n \times n}$ , donde  $n$  es el número de sitios, que captura la influencia de las variables en relación con su proximidad.

Esto significa que se propone a priori una relación espacial al dar estructura a  $W$ , por ejemplo usando solo valores de 1 (vecino) y 0 (no vecino) a las unidades espaciales contiguas o expresa una relación global usando una métrica de distancia en los valores de la matriz  $W$ , por ejemplo  $w_{ij} = 1/d_{ij}$  donde  $d_{ij}$  es la distancia euclíadiana entre los centroides de los sitios  $i$  y  $j$ . La idea es que los vecinos más cercanos influencian más la unidad  $i$ , basados en la primera ley de la geografía o Ley de Tobler<sup>1</sup> (Tobler, 1970). Este es un supuesto fuerte, y debe ser coherente con el fenómeno y los datos que lo representan. Justamente con la matriz  $W$  se calcula el coeficiente de Moran'I (Moran, 1950), que se usa para probar una asociación global entre una variable y su posición en el plano. Si los residuos  $e_i$  muestran una asociación fuerte con la estructura de  $W$  en la prueba de Moran'I, se sugiere usar alguno de los modelos espaciales (SAR,SLX,SEM o SD), pues esto significa que no se puede confiar en los coeficientes estimados por la regresión lineal. Se aduce que  $W$  puede ayudar a explicar el proceso que se está modelando. De lo contrario es mejor usar un modelo lineal u otra técnica. (Páez and Scott, 2005; Kissling and Carl, 2008; Anselin, 2002) hacen una revisión de los procesos espaciales, sus supuesto y los criterios de ajuste como errores cuadráticos y criterio de información de Akaike para la selección

---

<sup>1</sup>“Todas las cosas están relacionadas entre sí, pero las cosas más próximas en el espacio tienen una relación mayor que las distantes.”

de un modelo y diseño bien formulado y ajustado.

Modelo espacial autoregresivo (SAR)

$$Y = X\beta + \rho WY + e \quad (2.8)$$

Modelo de error espacial (SEM)

$$Y = X\beta + uu = \rho Wu + e \quad (2.9)$$

Modelo de retardo espacial en los terminos independiente (SLX)

$$Y = X\beta + WX\lambda + e \quad (2.10)$$

Modelo espacial de Durbin SD

$$Y = X\beta + \rho WY + WX\lambda + e \quad (2.11)$$

Estudios como (Landry and Chakraborty, 2009; Schwarz et al., 2015; Zhou and Kim, 2013; Shanahan et al., 2014) hacen uso de modelos autorregresivos y de retardo espacial para explicar los errores en el modelo lineal y mejorar el ajuste comparando modelos que relacionan de las variables ambientales(cobertura de copa y acceso a EV) con indicadores socioeconómicos.

Sin embargo trabajos como [LeSage and Pace (2014); Kissling and Carl (2008); Gibbons & Overman, 2012] hacen una crítica muy relevante a esta aproximación, y muestran que la inclusión de los términos de auto regresión (SAR) no aportan mucho, pues existe un acople entre los estimadores de los coeficientes que le quita peso a la posibilidad de interpretar los coeficientes como factores del efecto aislado de uno de las variables explicativas. Además,

muestra que el ajuste no suele ser muy sensible a variaciones en  $W$ , lo que derrumba un poco la idea de importancia de la vecindad definida a priori. (LeSage and Pace, 2014) propone que si la meta de usar modelos de regresión espacial es tener aproximadamente medidas escalares correctas de resumen de los efectos directos, indirectos y totales sobre la variable dependiente que surgen de cambios en las variables explicativas, entonces sí es adecuado el uso de estos modelos (LeSage and Pace, 2014).

Además del uso del análisis de autocorrelación espacial global, se puede explorar métodos que hacen análisis local de la variabilidad y sensibilidad de los parámetros usados para las estimaciones OLS. Si no hay autocorrelación global o ninguna agrupación, todavía podemos encontrar grupos a nivel local utilizando autocorrelación espacial local. El hecho de que de la Morán I es una suma de productos cruzados individuales es explotado por los “indicadores locales de asociación espacial” (LISA) para evaluar la agrupación de las unidades individuales mediante el cálculo del índice de Moran local para cada unidad espacial y la evaluación de la significación estadística p-value para cada sitio  $i$  (Talen and Anselin, 1998). En Fotheringham et al. (1998) se muestra que la regresión geográficamente ponderada (Geographically weighted regression GWR) produce resultados más informativos con respecto a la variación de parámetros en el espacio que la simple utilización de los agregados estadísticos. Estos métodos puede ser utilizado para examinar visualmente la variabilidad espacial de los resultados de la regresión o de las variables del estudio a través de una región y así informar sobre la presencia de no-estacionariedad espacial. La GWR es usada con éxito en (Comber et al., 2011) para analizar los posibles factores que pueden ayudar a identificar y caracterizar las dimensiones a tener en cuenta para mejorar el acceso a la salud.

### **2.3.3. Otras técnicas**

(Nesbitt and Meitner, 2016) investigan la relación de usos de suelo para establecer diferencias entre las coberturas usando análisis de varianza (ANOVA). En (Kabisch and Haase,

2014) se usa el análisis cluster para identificar grupos de población con características socioeconómicas similares y similar acceso a espacios verdes, para luego analizar si existen diferencias entre la distribución de los espacios verdes usando el coeficiente de Gini para diferentes grupos de población definidos en un tipo de análisis de disimilitud.

# Capítulo 3

## Metodología

Este trabajo se concentra en indagar en particular sobre la justicia ambiental distributiva por medio de modelos estadísticos, geoestadísticos apoyados por el análisis visual de los patrones espaciales. El motivo principal para hacer uso de este tipo de aproximación pasa por la identificación de relaciones estadísticamente significativas entre variables sociales, económicas, estructurales y ambientales incluyendo la capacidad de detectar agrupaciones de unidades geográficas con características homogéneas donde intervenir y disminuir las inequidades en el acceso a servicios ambientales.

Para ello se hará uso de datos del censo arbóreo de 2015<sup>1</sup>, el censo de población del 2005<sup>2</sup> y aspectos estructurales del espacio público y privado de las unidades espaciales de análisis<sup>3</sup>.

---

<sup>1</sup>Desafortunadamente los datos del censo arbóreo no están disponibles al público general por motivos desconocidos, pues según la legislación Colombiana no existe ningún tipo de impedimento ni restricciones de privacidad sobre estos datos, que fueron recolectados con dineros públicos. Trabajé con una copia a la que tuve acceso por vínculos laborales con la Alcaldía de Cali

<sup>2</sup>Disponibles en (cen; DANE, 2005; geo, a)

<sup>3</sup>Disponibles en (geo, b; noa, b)

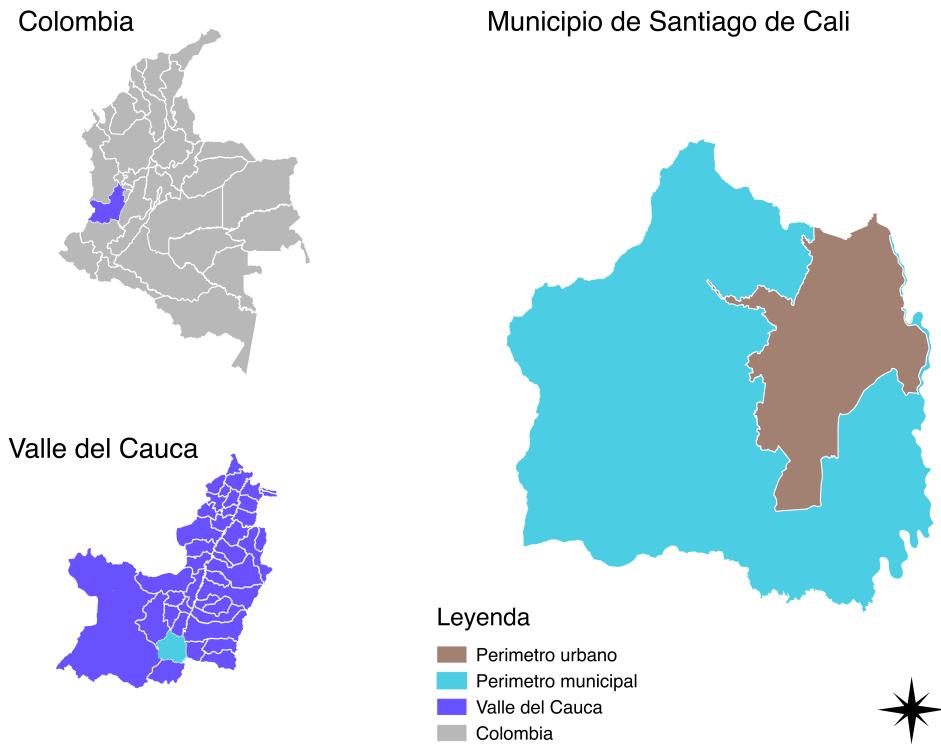


Figura 3.1: Área de estudio

### 3.1. Área de estudio

El municipio de Santiago de Cali se encuentra ubicado al suroccidente colombiano. Es la capital del departamento del Valle del Cauca y es la tercera ciudad más poblada del país, después de Bogotá y Medellín, con 2.420.114 habitantes según Cali en Cifras 2015(Escobar, 2015) . El municipio tiene un área 561.7 Km<sup>2</sup>, un área del perímetro urbano 120.4 Km<sup>2</sup> (21.4 %).(Ver figura 3.1).La división administrativa de la zona urbana son comunas y las comunas se componen de barrios.

Santiago de Cali presenta dos zonas topográficas: el valle del río Cauca hacia el oriente, el terreno más plano donde se ubica el casco urbano, y la zona de piedemonte hacia el

occidente sobre la margen derecha de la cordillera Occidental. El área urbana limita al oeste y sur con el área rural del municipio, al este con el río Cauca y los municipios de Palmira y Candelaria, y al norte con el municipio de Yumbo.

El clima del municipio varía en relación al rango altitudinal que abarca entre 916 y 1,438 msnm. En la zona plana, se presenta un clima cálido con características semihúmedas hacia el sur y semiáridas hacia el norte mientras el piedemonte presenta condiciones de clima templado. La precipitación anual promedio es de 1.500 mm y la temperatura promedio anual es de 24 °C aproximadamente (CIAT, 2015a). La ciudad de Cali es de clima caliente, donde la sombra y la brisa son bien valoradas por sus habitantes.

## **3.2. Datos**

### **3.2.1. Datos de registros oficiales del municipio**

La cartografía disponible en la Infraestructura de Datos Espaciales de Santiago de Cali, IDESC (geo, b), incluye información sobre los objetos geográficos naturales, de infraestructura urbana, límites y divisiones político administrativas y la clasificación de predios en cuanto a espacio público disponibles en coordenadas planas del sistema (noa, a). Además está la base de datos geográfica del Plan de Ordenamiento territorial de Cali 2014, POT2014<sup>4</sup>. Del POT2014 se seleccionaron conjuntos de datos de equipamientos y espacio público contenido en la estructura ecológica complementaria (ECC) que incluye cementerios, universidades, EV de acceso no restringido aunque algunos sea predios privados contenidos en EEC. De la IDESC se seleccionó la capa de barrios, espacio público, humedales, ríos y corredores ambientales disponibles vía WFS. La capa de manzanas es necesaria para refinar las capas de espacio verde y poder calcular el área de calle , área privada y otras métricas sobre la estructura de cada sector sector censal y que servirán

---

<sup>4</sup>Toda la información del POT2014 se encuentra en la web de la Alcaldía y puede descargarse como archivo GDB compatible con ArcGIS 10.4 o consumirse de Geoserver de IDESC como WFS o mapas en formato pdf del acuerdo

como criterios para la selección de sectores urbanos a incluir en los análisis de regresión.

En la figura 3.2 se muestra un mapa con las capas seleccionadas para el realizar el procesamiento y los análisis.

### 3.2.2. El censo arbóreo

En el año 2015 la ciudad de Santiago de Cali (Colombia) concretó la realización de un censo arbóreo que dejó como resultado una base de datos de aproximadamente 290.000 individuos censados. Los datos dan cuenta de la identificación de especies, sus características dasométricas, de emplazamiento y estado fitosanitario. Estos datos constituyen un insumo fundamental para la caracterización de los beneficios y cargas que supone el mantenimiento y desarrollo del arbolado urbano. De hecho, su realización ocurre en el marco del proceso de formulación del plan silvicultura urbana o estatuto arbóreo<sup>5</sup> (Acuerdo 0335 de 2013). Las variables biofísicas recolectadas y la georeferenciación de los individuos permite agregar las características del arbolado a diferentes escalas de las unidades administrativas p.e divisiones censales, para identificar y caracterizar su distribución espacial y correlación con variables sociales o/y económicas. Algunas de las variables incluidas en el censo se resumen en la tabla 3.1 y en la tabla 3.2.

Cuadro 3.1: Variables para caracterizar el AU

variable	{valores}[unidades]
id_arbol	número entero único
diametro	[m <sup>2</sup> ]
copa	

---

<sup>5</sup>El proyecto de censo arbóreo se formuló en dos fases; la primera se ejecutó mediante convenio No 095 de 2013 entre la CVC y la Universidad Autónoma de Occidente, y la segunda fase mediante convenio No 049 de 2014 entre las mismas entidades. Los datos no se encuentran publicados y fueron solicitados mediante un derecho de petición.

Cuadro 3.2: Muestra de los datos del censo arbóreo

id	vegetacion	edad	emplazamiento	diametro_copa	altura_arbol
0199G41070768	Arbol	Maduro	Ronda de rios	12	11
0199G41070769	Arbol	Maduro	Ronda de rios	7	8
0199G41070770	Arbol	Maduro	Ronda de rios	5	3
0199G41070771	Arbol	Maduro	Ronda de rios	6	7
0199G41070772	Arbol	Maduro	Ronda de rios	9	8
0199G41070773	Arbol	Maduro	Ronda de rios	10	7

variable	{valores}[unidades]
altura	[m]
arbol	
vitalidad	{Regular, Sano, Seco, Muerto}
edad	{Juvenil, Maduro, Longevo}
emplazamiento	{Anden, Bahias de estacionamiento, Bulevares, Corredor Ferreo, Escenario deportivo y/o Cultural, Glorieta, Parque Urbano, Paseos, Plaza, Plazoleta, Ronda de rios, Rondas de canales, Separador Vial}
vegetación	{Arbol, Arbusto, Bambu, Muerto, Palma, Planta arbustiva, Seco}
Este	[m] MAGNA - SIRGAS-CALI
Norte	[m] MAGNA - SIRGAS-CALI

Existe una diferencia de 10 años entre censo de población de 2005 y el censo arbóreo de la ciudad de Cali. Aunque esto pueda parecer una situación que reduce la legitimidad de los resultados que se hallen en este estudio, autores como Boone et al. (2010) y Schwarz et al. (2015) reconocen que los paisajes que vemos hoy son legados de patrones de consumo

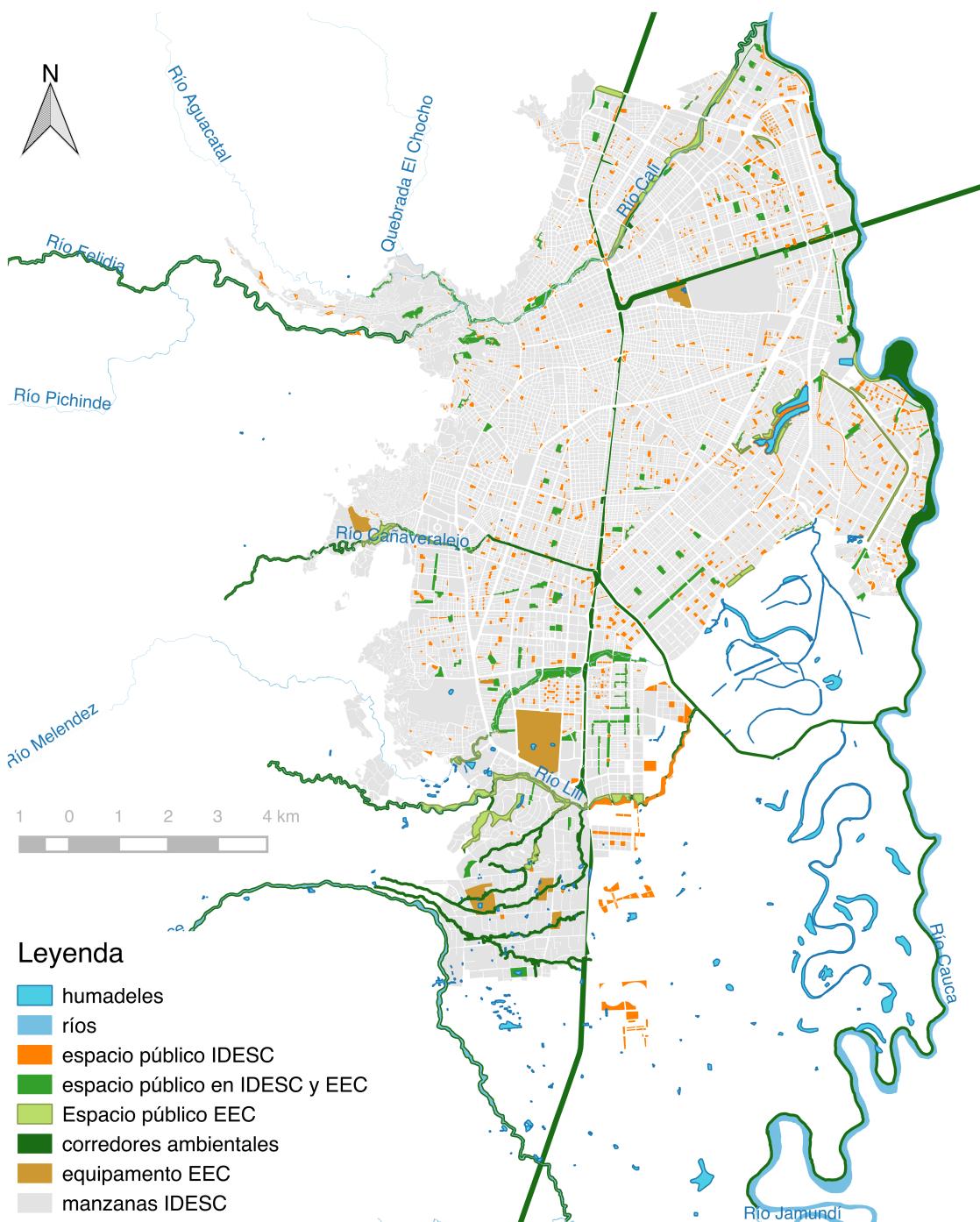


Figura 3.2: Capas usadas para el procesamiento de los espacios verdes y las características de las manzanas

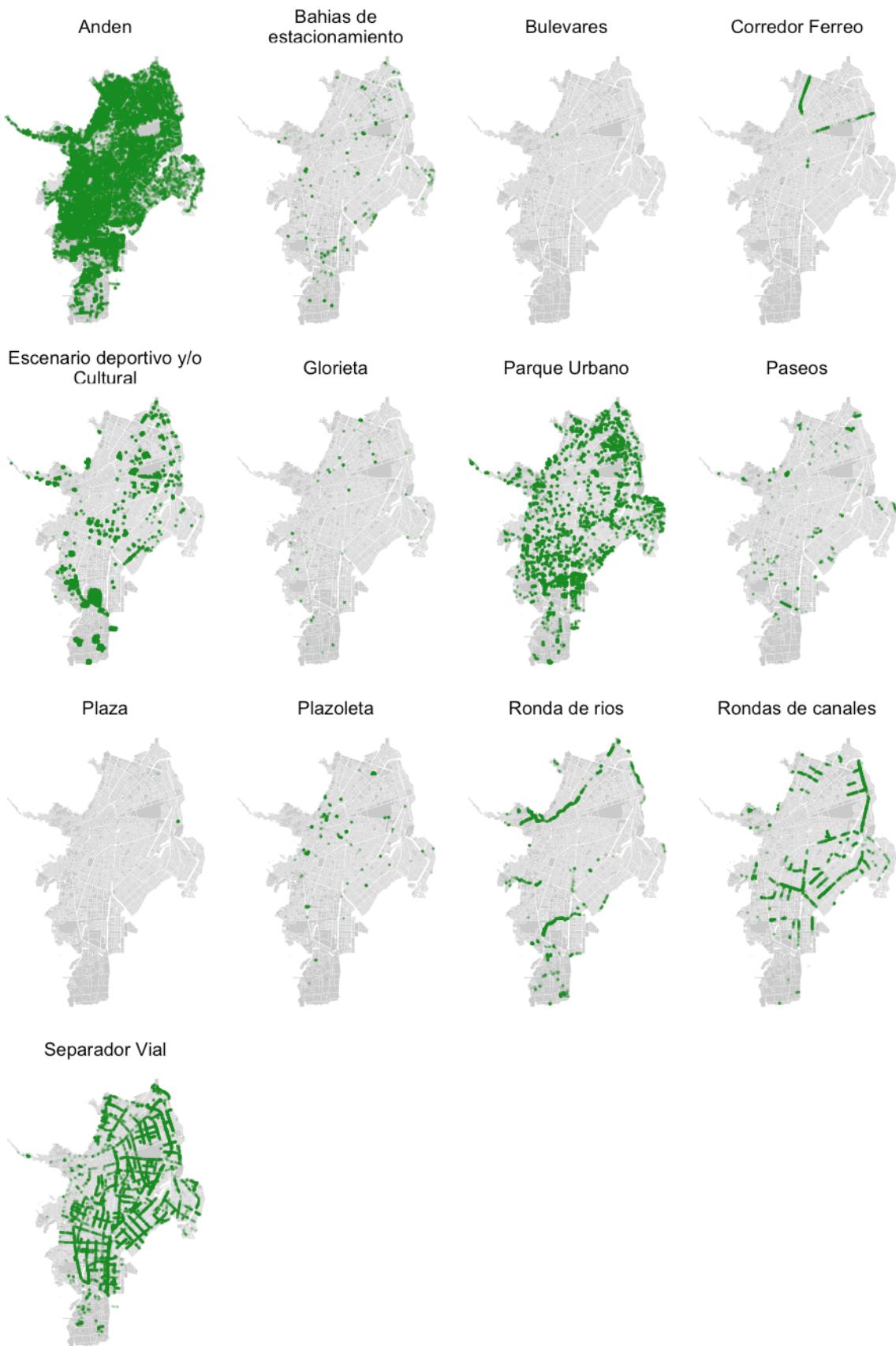


Figura 3.3: Small multiples de los individuos arbóreos por emplazamiento

pasados, y que en el caso de la vegetación urbana tratamos con organismos de larga vida que pueden tardar mucho tiempo en establecerse y crecer. En contraste, la estructura social de las ciudades puede cambiar más rápidamente.

Una apuesta para reducir la brecha es la exclusión de los árboles jóvenes del inventario, que posiblemente no estaban ahí en 2005. Aunque no conocemos las tasa anual de tala de árboles en la ciudad, y dado es posible que una parte importante de los árboles jóvenes haya reemplazado a los que fueron talados, no parece realista mantener el inventario entero.

En general toda la vegetación aporta beneficios ambientales a los habitantes, en este estudio descartamos la vegetación arbustiva y los árboles, palmas y bambú de menos de 1.9 m de altura para circunscribirnos a los individuos más desarrollados.

Una vez aplicado este filtro contamos con 203112 individuos.

### **3.2.3. El censo de población**

El último censo de población en Colombia se realizó en el año 2005, y los datos se pueden consultar y agregar en las diferentes unidades censales (sector, sección, manzana) usando una sistema de consulta web de censos Redatam<sup>6</sup>. Estos datos sirven para caracterizar la población con base en indicadores y rasgos de las personas, aspectos sobre el uso del suelo y los tipos de vivienda. Las variables disponibles para el análisis están resumidas en las tablas 3.3 y 3.4.

El otro componente de los datos es la cartografía censal del DANE (geo, a) disponible para las diferentes unidades espaciales de agregación en el sistema de coordenadas WGS84. Para el análisis se tiene en cuenta todos las unidades censales que se interceptan con el perímetro urbano disponible en la IDESC, pues el censo arbóreo se limitó al este perímetro.(ver figura 3.4)

---

<sup>6</sup>El sistema de consulta es el (CEPAL), que se puede acceder directamente desde (DANE, 2005) y en la página web del DANE dónde está organizada la documentación metodológica y otros servicios del censo.

Cuadro 3.3: Variables sobre la población

variable	{valores}[unidades]
Pertenencia Étnica	[personas]{indígenas, ROM, gitanos, raizales del Archipiélago de San Andrés, Providencia y Santa Catalina, palenqueros de San Basilio, afrocolombianos}
Con alguna limitación	[personas]{sí,no}
Con estudios superiores o postgrado	[personas]
Ningún estudio	[personas]

Cuadro 3.4: Variables sobre las viviendas

variable	{valores}[unidades]
tipo vivienda	[viviendas] {Casa,Casa indígena,Apartamento,Tipo cuarto,Otro tipo de vivienda}
uso vivienda	[predio]{Uso Vivienda.Uso Unidad Económica,Uso LEA}
cantidad predios	[predios]

variable	{valores}[unidades]
cantidad	[viviendas]
viviendas	

Una de las apuestas del proyecto es incluir aspectos de los procesos de desarrollo urbano a través de la idea de barrio: como unidad de identidad cultural urbana y estructural, de características físicas y habitacionales en las que confluyen las transformaciones que hacen los habitantes y los diseños urbanos e intervenciones arquitectónicas de los planeadores y constructores en la ciudad. La unidad espacial de análisis sobre la cual se harán todas las agregaciones es el sector urbano (SU) de la cartografía censal 2005.

### 3.3. Métodos y técnicas

El análisis propuesto se compone de las siguientes actividades:

1. Preparación de los datos: una tarea común pero crucial para el análisis de datos. La estandarización de las variables categóricas y la identificación de valores atípicos o inconsistentes es una base firme para la estimación de parámetros y obtener soluciones confiables y sensibles de interpretación. Los datos suelen estar usualmente en formatos para la lectura humana o con distinta estructura de las variables de los modelos. La preparación de los datos consume la mayor parte de los esfuerzos de las tareas de procesamiento de los datos. Los datos del censo arbóreo se encuentran en tablas bien conformadas lo que facilita su manipulación. Los datos de consulta del censo de población vienen en tablas independientes para cada unidad espacial seleccionada, con diferentes longitud de variables. A esto se suma el componente espacial, donde hay que prestar particular atención a los sistemas de coordenadas y usar coordenadas planas consistentes con las unidades de espacio.
2. Procesamiento y análisis estadístico: cálculo de indicadores de cobertura, acceso y

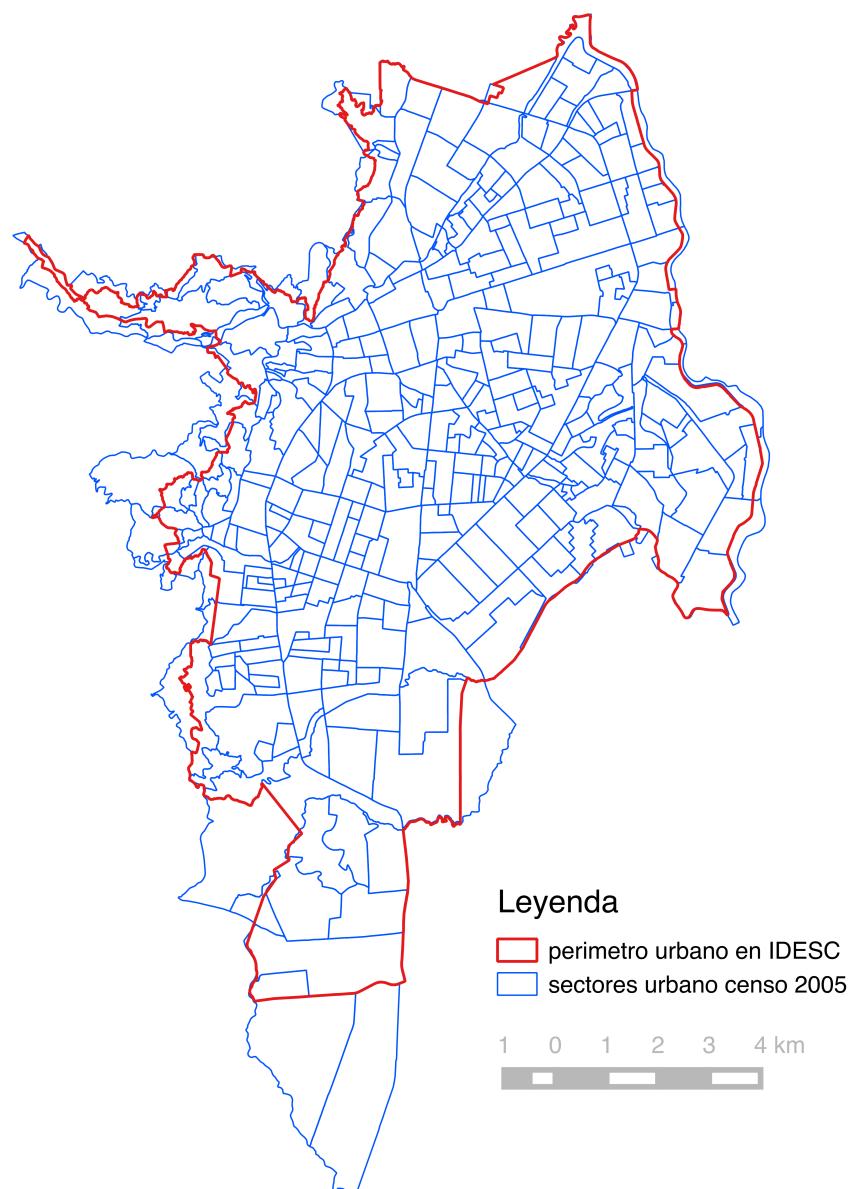


Figura 3.4: División en barrios y sectores urbanos de Santiago de Cali

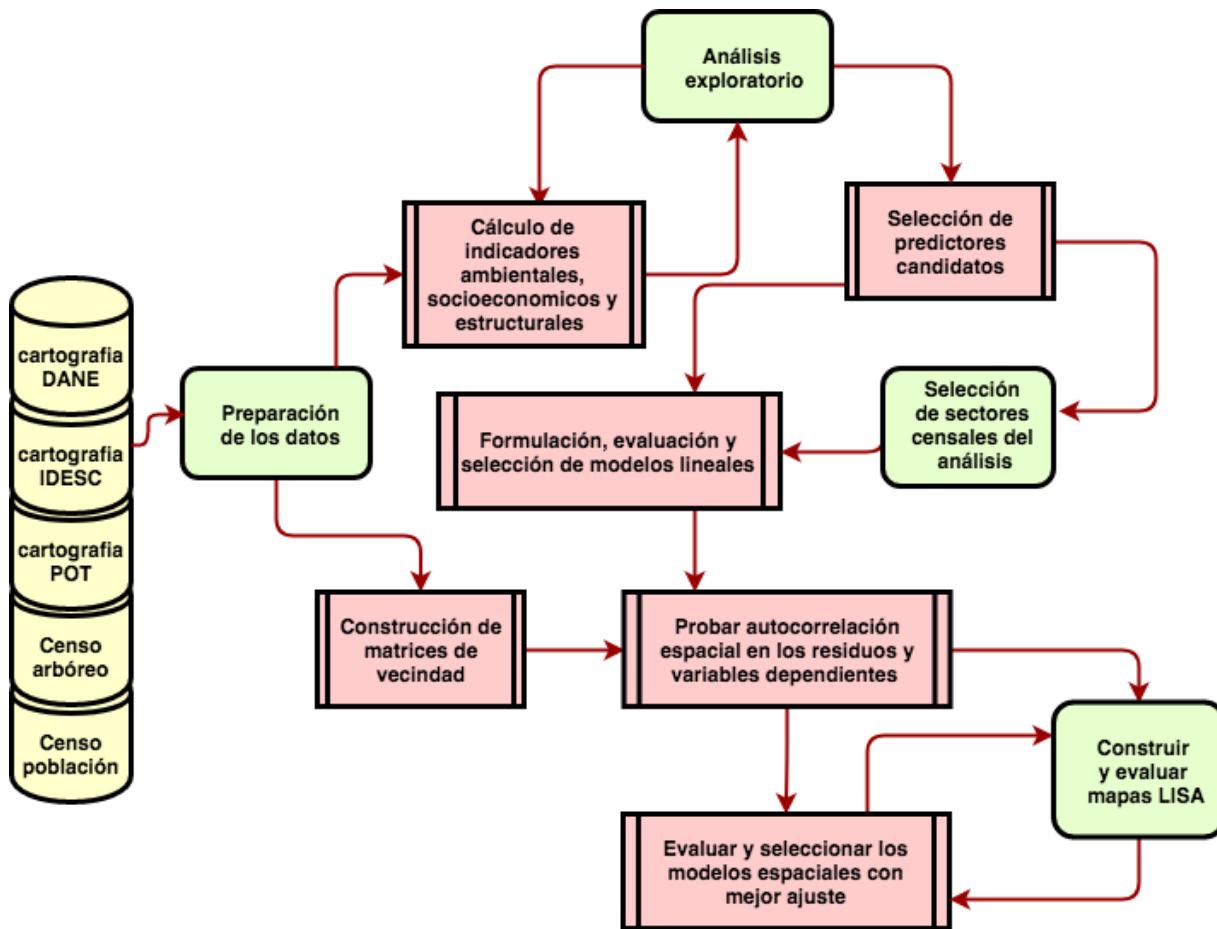


Figura 3.5: Diagrama de flujo de la metodología

variables socioeconómicas. Cálculo de estadísticos para probar normalidad, normalización de las variables e indicadores, cálculos de coeficientes de correlación Pearson y de Spearman entre todos los pares de variables.

3. Inspección visual de los datos: hacer uso de gráficas estadísticas, mapeos y mapas para evaluar y seleccionar los indicadores a usar en un modelo de regresión lineal.
4. Evaluar los residuos usando la prueba de correlación espacial de Moran's I usando al menos dos diseños de matriz W. Si la prueba muestra una correlación y un valor de significancia alta, se prueban modelos tipo SAR, SEM o SLX para comparar su desempeño.
5. Selección del modelo que mejor se ajusta usando métricas de error y de ajuste como  $R^2$  y el criterio de Akaike.

### 3.3.1. Procesamiento de datos

El procesamiento de los datos se realizó principalmente en R Core Team (2017). Se usó QGIS para conectarse a los servicios WFS del IDESC y previsualizar las capas de información geográfica recolectada y la realización de algunos de los mapas detallados.

El código que implementa los análisis está dividido en archivos para facilitar su lectura, cada uno de los cuales se encargan de transformar los datos de las fuentes y construir estructuras de datos necesarias para realizar las regresiones, las gráficas y los análisis de tipo estadístico y geoestadístico. Cada script implementa una fase de la metodología y produce resultados intermedios que facilitan seguir y reproducir dichas transformaciones sobre los datos de un dominio del problema. El archivo de `funciones.R` agrupa funciones que encapsulan funcionalidades recurrentes dentro del desarrollo del análisis. El script de `geodata.R` opera sobre los fuentes de datos geográficas necesarias para consolidar los índices de acceso a espacios verdes (EV), los indicadores y variables de la estructura de física de los sectores censales y unidades geográficas del análisis. El script `arboles.R` consolida la información de cada uno de los individuos del censo arbóreo agregandolos por sector censal. El script `censopoblacion.R` consolida los datos del Censo de Población 2005. Los scripts `consolidarDatos.R` y `analisis_exploratorio.R` consolidan una única estructura con todos los datos y produce una serie de gráficas y medidas de correlación, que son base para la identificación de supuestos y selección de las variables independientes para los análisis estadísticos y las regresiones espaciales. Finalmente los script de `analisis_estadistico.R` y `analisis_geoestadistico.R` implementan las regresiones lineales y las regresiones espaciales respectivamente, así como los test y tablas para la verificación de los supuesto matemáticos y la verificación de la calidad de los resultados. Todos estos están reunidos en un script que carga las librerías necesarias y ejecuta secuencialmente cada de los scripts descritos.

Para usar la información geográfica de la cartografía censal y la información del IDESC es necesario establecer un sistema de coordenadas común, en unidades métricas, que facilite

integrar la información y produzca resultados consistentes. El sistema de coordenadas proyectadas que vamos a usar es noa (a). Para cargar y manipular los datos espaciales hacemos uso de las librerías `rgdal` (Bivand et al., 2017), `rgeos` (Bivand and Rundel, 2017) y `sp` (Pebesma and Bivand, 2018).

El código y los datos están disponibles en este repositorio de GitHub.

### 3.3.2. Cálculo de métricas de acceso a servicios ambientales

#### 3.3.2.1. Indicadores de beneficios del arbolado urbano

Entre los distintos indicadores desarrollados para capturar la extensión y distribución de los servicios ambientales la cobertura de copas ha probado ser sensible y eficaz para cuantificar hasta qué punto los árboles y bosques están proporcionando servicios críticos a los residentes (Nowak et al., 2010).

En este trabajo usaremos dos variantes de la cobertura de copa: el área de copa en metros<sup>2</sup> (`area_copa`) y la cobertura de copa como porcentaje del área pública total (`cobertura_copa.ap`), conformada por la vías y calles más el área de espacio públicos) (ver figura 3.6).

#### 3.3.2.2. Índices de acceso a espacios verdes

Para mejorar la lectura de esta sección se incluyen a continuación las ecuaciones que definen los índices de acceso seleccionados con las variantes definidas en este trabajo.

**índice contenedor porcentual** (`area_ep.porcentaje`)

$$A_i^{C_p} = 1/a_i \sum_j s_j \quad \forall j \in I \tag{3.1}$$

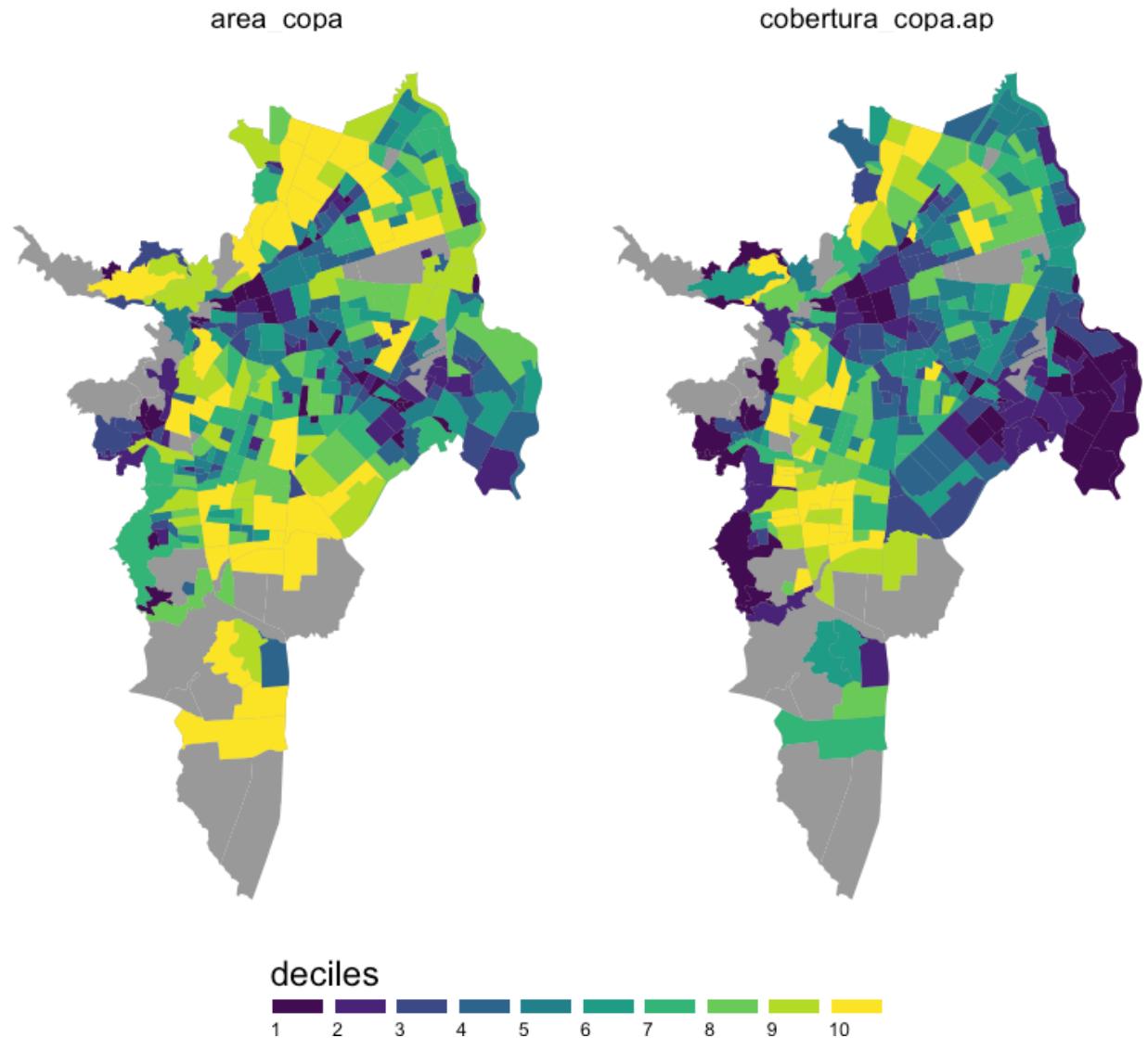


Figura 3.6: Sectores urbanos de las variables dependientes sobre cobertura de copa

donde  $s_j$  es el área de cada espacio verde  $j$  que pertenece al conjunto  $I$  de EV dentro del sector  $i$  y  $a_i$  es el área del sector  $i$ .

#### **razón área disponible distancia (ia.areas.dist)**

$$\bar{A}_i^{AD} = \frac{\sum_{j \in I_{R_b}} s_j}{\sum_{j \in I_{R_b}} d_{ij}} \quad \forall j \in I_{R_b} \quad (3.2)$$

donde  $R_b$  es el radio de búsqueda,  $s_j$  es el área de cada espacio verde  $j$ ,  $d_{ij}$  es la distancia del centroide del sector  $i$  al espacio  $j$  que pertenecen al conjunto  $I_{R_b}$  de EVs en el radio de búsqueda.

Este índice muestran un dimensión relacionada no con solo con el acceso sino con la cantidad de espacio disponible en el radio de búsqueda definido desde el centroide del sector censal. Para hacernos una idea del radio de búsqueda seleccionado, el siguiente mapa muestra los radios búsqueda y los espacios verdes.

### **3.3.3. Cálculo de métricas de sobre la población**

#### **3.3.3.1. Características de la población**

La tabla 3.3 resumen las variables consideradas inicialmente en este trabajo, sin embargo, algunas de ellas no contienen suficiente variabilidad o el número de individuos es muy bajo en comparación con el total de la población. En la tabla 3.5 se observa el bajo número de personas que pertenecen al pueblo Rom (gitanos), Palenqueros de San Basilio (departamento de Bolívar) y de Raizales del Archipiélago de San Andrés, Providencia y Santa Catalina (SAI), por lo que son descartados del análisis al igual que la población indígena.

Las variables del censo de población seleccionadas para el análisis se muestran en la figura 3.8

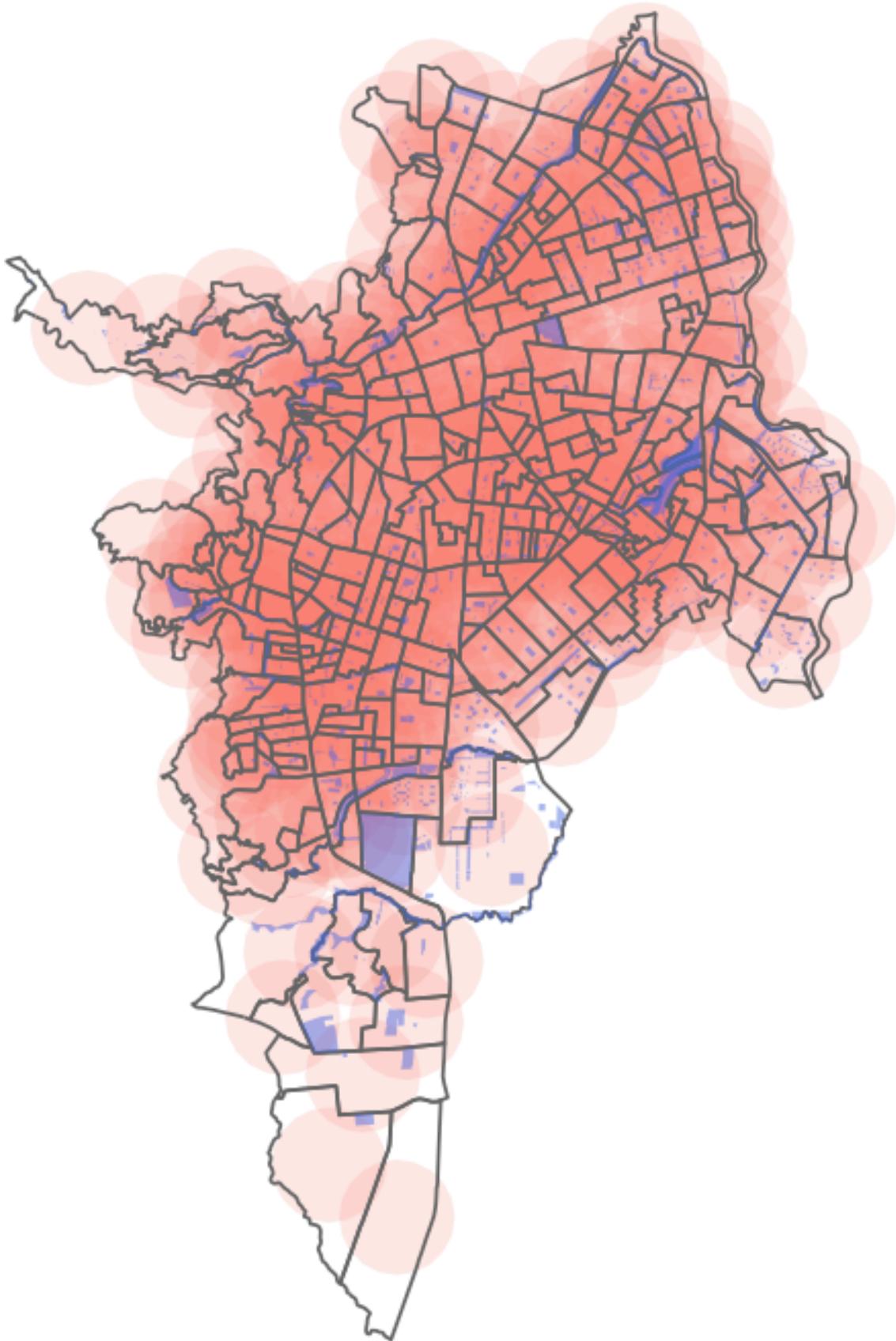
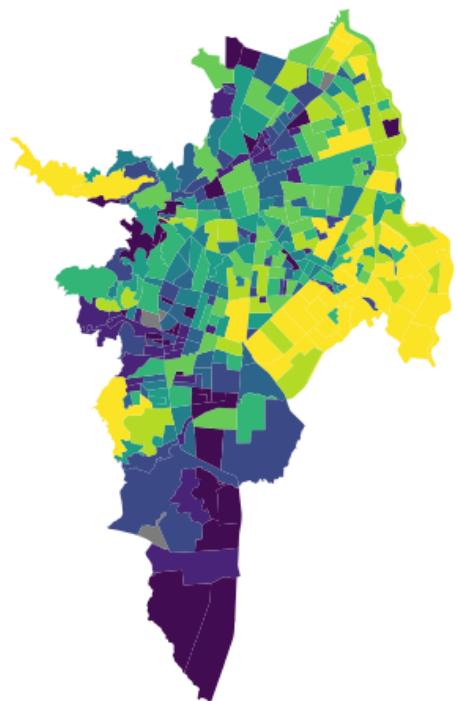
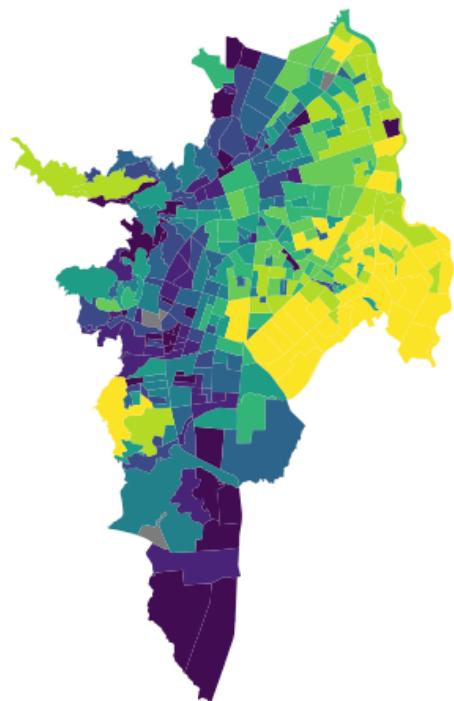


Figura 3.7: Espacio verdes y rango de 1 km desde centriodes de SU

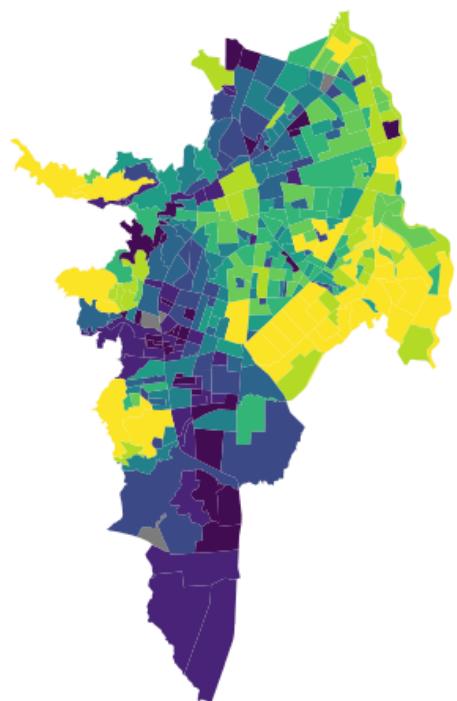
con\_alguna\_limitacion



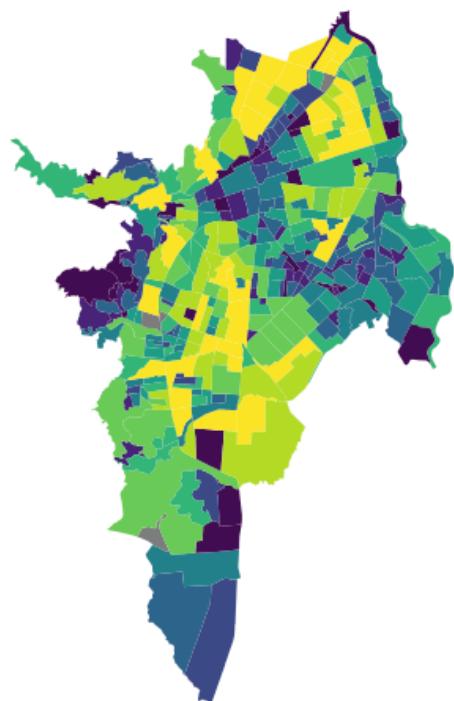
negro\_mulato\_afrocolombiano



ningun estudio



superior postgrado



deciles



Figura 3.8: Mapas de las variables de población seleccionadas (en deciles)

Cuadro 3.5: Totales de población en la ciudad de Cali

Tipo	Cantidad
Población Total	2027024
Población afrodescendiente, negros o mulatos	530990
Población indígena	9195
Población Rom	690
Población Palenqueros	1
Población raizales de SAI	851

Además de las variables seleccionadas podemos calcular indicadores como la densidad de población: dado que los árboles compiten por el espacio con los seres humanos es de esperarse que a mayor cantidad de personas haya menos lugar para los árboles. Podemos de nuevo calcular indicadores porcentuales respecto de la población total de cada unidad geográfica para facilitar la comparaciones y acentuar las diferencias entre los diferentes sectores (figura 3.9).

### 3.3.3.2. Características de las viviendas

Además de las rasgos étnicos, condiciones de estudio y limitaciones de la población el censo de 2005 tiene disponibles datos sobre el tipo de viviendas (casa, apartamento, tipo cuarto, casa indígena, otros), y el uso habitacional, comercial y la cantidad de unidades especiales de alojamiento L.E.A dado a los predios. La vocación comercial o residencial de un barrio puede ser un factor en el desarrollo del arbolado urbano o de las disposiciones urbanísticas de la ciudad en relación a los EV, ya sea por las condiciones físicas como por la intervención de sus habitantes. Estas variables pueden también expresarse como porcentaje de la cantidad de predios de vivienda en el caso de los tipos o como porcentaje de la cantidad de predios en el caso del uso como unidad de vivienda, económica o L.E.A.

A continuación presentamos el resumen, los mapas por sector urbano (figuras 3.10).

El uso de L.E.A tiene una distribución concentrada en uno o pocos SU, por lo que podemos descartarla para los análisis de regresión. Existe también cierta complementariedad entre

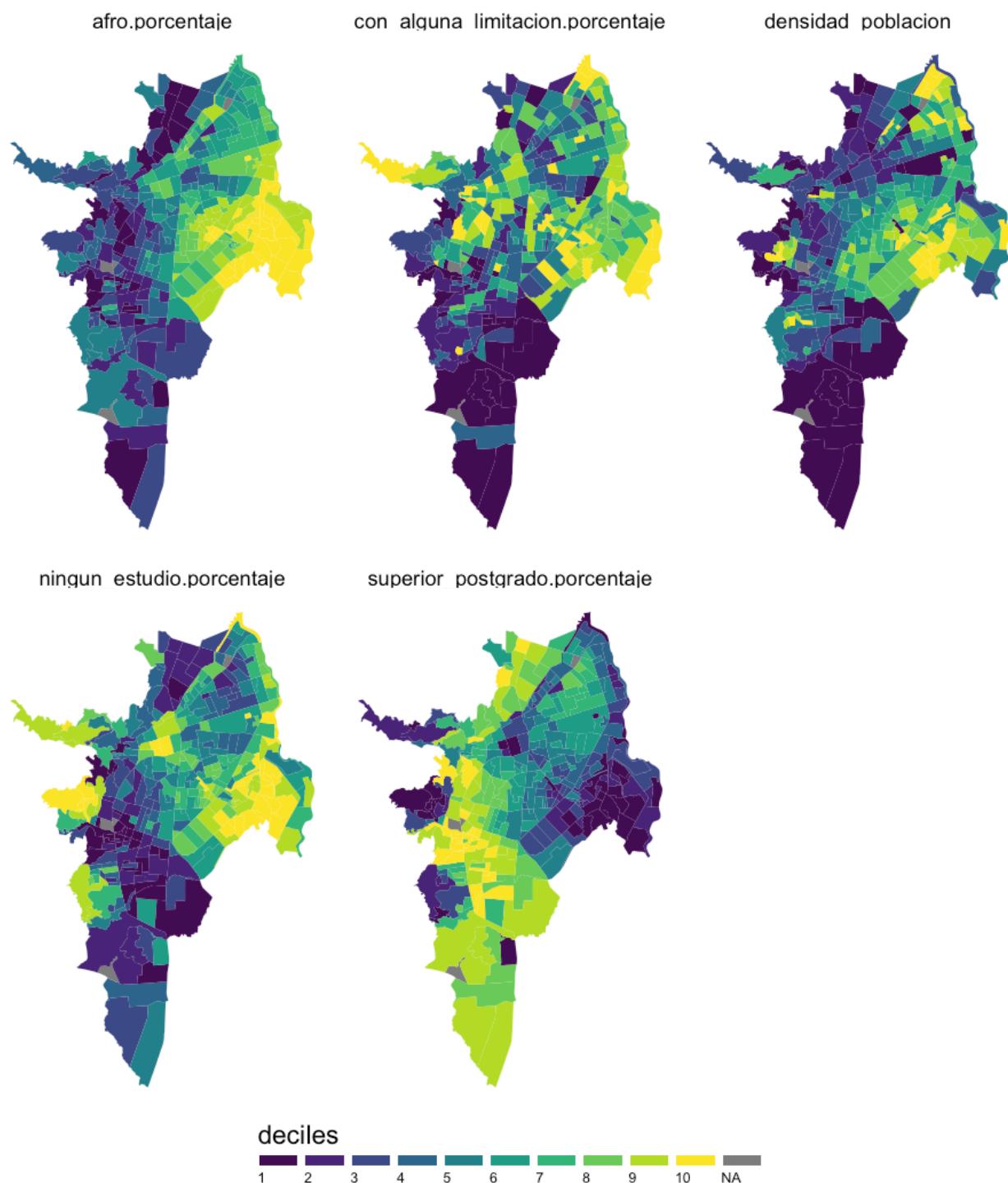


Figura 3.9: Mapas de las variables de población seleccionadas como porcentajes (en deciles)

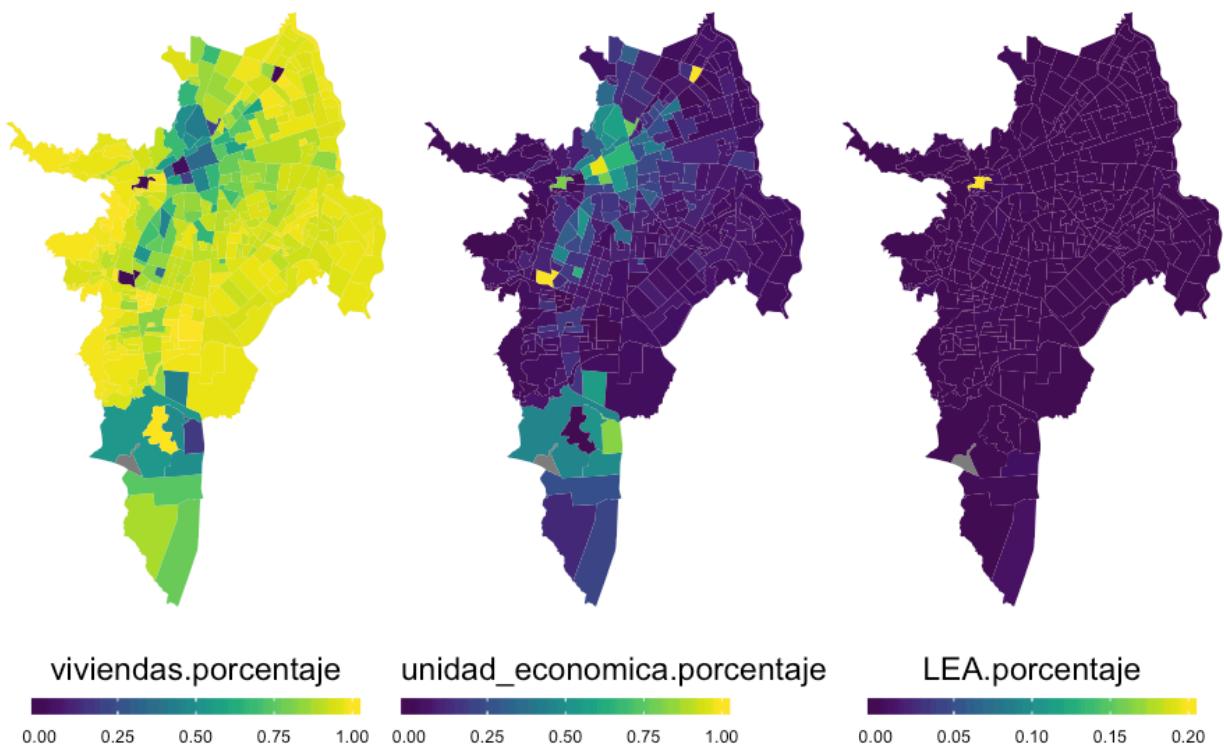


Figura 3.10: Mapas de las variables sobre el tipo de uso de los predios como porcentaje de la cantidad de predios (escala continua)

el uso de vivienda y los usos económicos de los predios, porque seguramente, si existe una correlación entre estas variables y la cobertura de copa o el acceso a espacios verdes una de las dos puede bastar para incluir esta dimensión en los modelos de regresión.

### 3.3.4. Criterios y selección de sectores censales

Antes de iniciar un análisis de regresión establecemos los criterios para inclusión o no de ciertos datos dentro del conjunto de valores para la regresión y cálculo de la correlación. Estos criterios se listan a continuación:

- sectores sin personas
- sectores sin viviendas
- sectores área de espacio público mayor que el 60 % del área del sector
- sectores área de calle mayor que el 80 % del área del sector

- sectores área privada mayor que el 90 % del área del sector

Además de estos criterios se excluyeron los sectores donde está la Laguna el Pondaje, que cubre una porción muy importante del sector que no se ve reflejado en las otras métricas, los sectores con una porción mayor al 60 % por fuera del perímetro urbano o sin urbanización visible en las imágenes satelitales. Así los sectores excluidos del análisis se muestran en los mapas 3.11.

### **3.3.5. Selección de variables dependientes y regresiones lineales**

Las variables a incluir en los modelos lineales deben cumplir una serie de condiciones para ser elegidas como candidatas:

- *Mostrar una correlación fuerte* (típicamente mayor a 0.6 se considera una asociación fuerte).
- *Las variables independientes o predictoras no deben estar fuertemente correlacionadas entre ellas.*
- *Las observaciones deben ser independientes.* En nuestro caso significa que no debe existir relación espacial o temporal entre los diferentes sectores. Justamente esto se pondrá a prueba con los test estadísticos y los graficos de diagnostico sobre la distribución de los residuos de la regresión: se espera que dicha dependencia esté motivada por la vecindad de los sectores.
- *Las variables dependientes e independientes deben tener una distribución normal.* Esta condición no suele ser estricta, pues lo importante es que al calcular los coeficientes de la regresión obtengamos una distribución normal de los residuos (sin ningún patrón, ruido). De no ser así, es posible que las variables no sean independientes o que exista información significativa en los residuos, por ejemplo, porque existe autocorrelación espacial en la variable dependiente y entonces la regresión lineal no obtiene resultados confiables para los coeficientes.

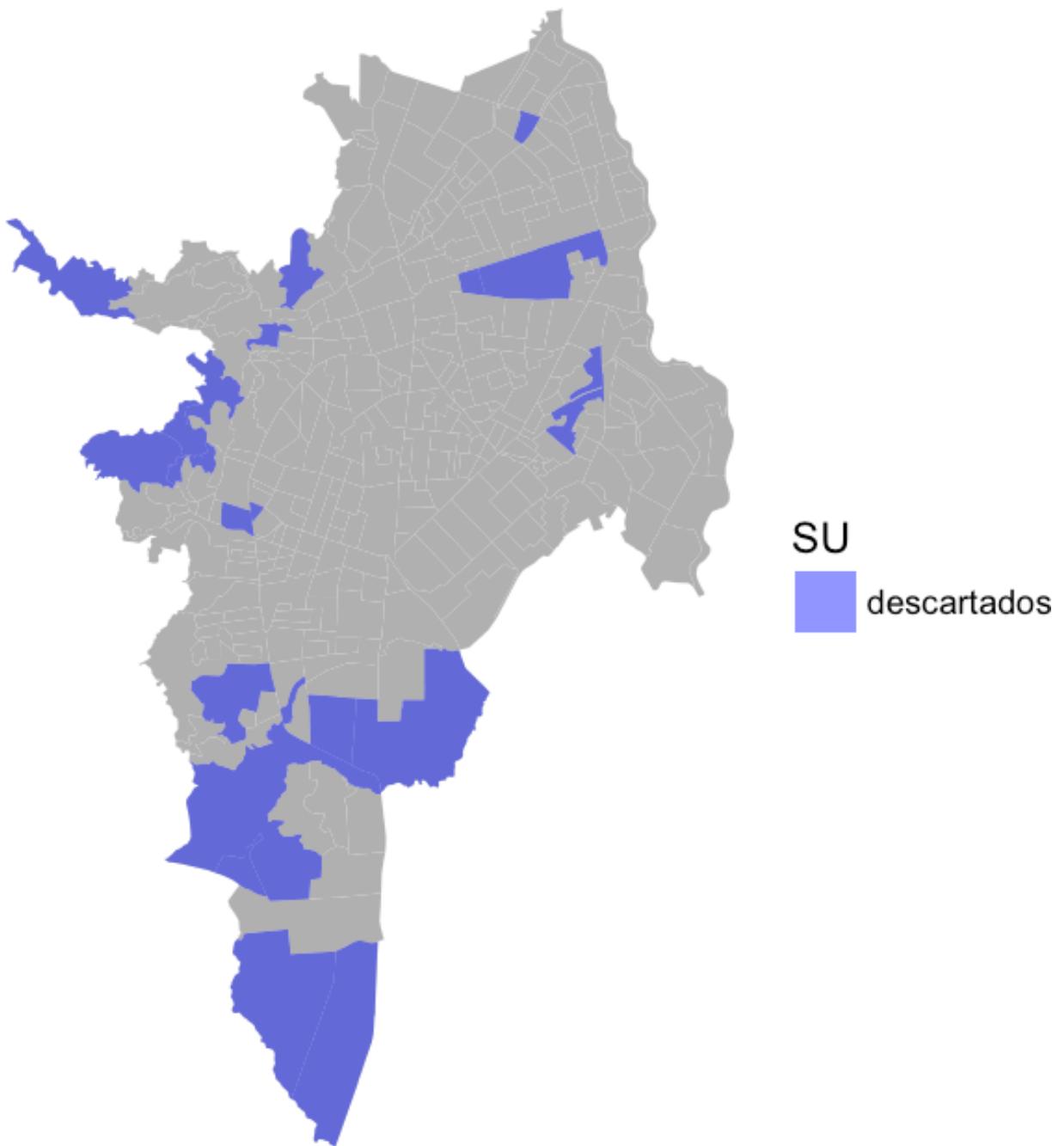


Figura 3.11: Mapa de los sectores excluidos

La modelación inicia con el análisis con las variables sobre la población, que son las de mayor interés en un estudio dado su enfoque en la justicia ambiental, para luego incorporar las variables de los dominios relacionados con el uso de los predios, los tipos de viviendas y la existencia de espacios verdes como parques, bulevares, escenarios deportivos o plazas, alojan una cantidad considerable de los individuos arbóreos de la ciudad.

Para garantizar que las variables no están correlacionadas entre sí, usaremos los coeficientes de correlación de Pearson, usado para detectar relaciones lineales, usualmente en variables con distribución normal, y el coeficiente de Spearman para detectar relaciones en variables con otras distribuciones o que exhiben relaciones no lineales. Para tener una idea más amplia sobre esa relación que expresan los coeficiente de correlación se incluyen gráficas de dispersión entre las variables independientes, y entre dependientes.

Para seleccionar las variables que mejor predicen la cobertura de copa aplicamos un procedimiento análogo al realizado con las variables dependientes entre sí. Con base en los coeficientes de correlación de Pearson y Spearman entre las variables dependientes e independientes, y teniendo en cuenta las restricciones de colinealidad entre las variables dependientes, seleccionamos las variables a usar en el modelo lineal.

Antes ajustar los modelos suele ser común en los modelos de regresión ajustar la distribución de las variables dependientes (y a veces las independientes) por motivos teóricos usando transformaciones logarítmicas o de raíz cuadra para eliminar no linealidades entre las variables dependientes y las independientes, y reducir posibles fenómenos de heterocedasticidad debido a estas no-linealidades.

Dividir o multiplicar por alguna constante no tiene ningún efecto en la calidad de las estimaciones , pero sí sobre los coeficientes de la regresión. Esto suele ser sensible a la hora de interpretar los cambios marginales de cada una de las variables independientes y su efecto sobre la variable dependiente. Sin embargo, lo que interesa para este estudio no es la interpretación de esos cambio sino la importancia relativa de cada variable y comparar los cambios de los coeficientes de regresión para el ajuste de cada modelo y/o las mejoras

que pueda operar un modelo autorregresivo en caso de encontrarse autocorrelación en los residuos de la regresión lineal. Por esta razón, normalizar los valores puede ser una ventaja pues mantiene los coeficientes mejor acotados. La normalización se aplica posterior a las transformaciones propuestas y se realiza dividiendo por el máximo valor de los datos de cada variable para mantener valores en el intervalo [0,1], dado que los valores son todos iguales o mayores que 0.

Al aplicar test para verificar que las condiciones de un buen ajuste (no hay sesgos en el estimador o una mala especificación del modelo) de un modelo lineal se cumplen:

- La media de los residuos es 0 o muy cercana.
- La distribución de los residuos es normal.
- Los residuos muestran homocedasticidad (la varianza es constante)

Para verificar la normalidad de los residuos se hace uso del test de Shapiro–Wilk (Shapiro and Wilk, 1965, ) y para la verificar si existe homocedasticidad el test de Breusch–Pagan (Breusch and Pagan, 1979).

### 3.3.6. Análisis geoestadísticos

Para los análisis geoestadísticos introducimos los modelos autoregresivos para obtener mejoras en la estimación de los coeficientes y en el ajuste de los modelos si existe algún tipo de autocorrelación espacial en los residuos. Existe una variedad de estos modelos que capturan diferentes tipos de efectos: modelo autoregresivo SAR que capturan efectos de la variable dependiente, ecuación (2.8), sobre las variables independientes (spatial lag o retardo espacial en  $X$  SLX, ecuación (2.10), en el error (modelo espacial del error SEM, ecuación (2.9) o usando una combinación del modelo de error y autoregresivo (modelo espacial de Durbin SD, ecuación (2.11)). Todas estas aproximaciones introducen una matriz de  $W_{n \times n}$ , donde  $n$  es el número de sitios, que captura la influencia de las variables en relación con su proximidad. Esta matriz  $W$  es una estructura que restringe la influencia a

priori en el modelo. Para observar el efecto que tiene esta matriz sobre los resultados del modelo usaremos 2 matrices distintas, y veremos su impacto en la estimación.

Para los análisis espaciales usaremos la librería `spdep` (Bivand, 2017)

### 3.3.6.1. Matrices de vecindad

La matriz  $W$  representa la topología de vecindad entre los sectores censales. Existen la literatura diferentes tipos de vecindad: *rook*, *bishop* y *queen* son las más referenciadas. Esta vecindad está representada en la matriz con 1 cuando existe vecindad y 0 cuando no. Otra forma de cuantificar la interacción de esa vecindad es usando una matriz de inversos de la distancia entre los centroides de los sectores censales, con el fin de atenuar la interacción entre sectores muy alejados y tener una variable continua que representa esa influencia. En la figura 3.12 se muestra la matriz  $W$  definida para vecinos que comparten un lado del polígono (vecindad *rook*) para todos los sectores de la ciudad de Cali.

Sin embargo, la regresiones se realizan sobre un subconjunto de los datos, y por tanto la estructura de esta matriz debe tener esto en cuenta, o mejor, no tener en cuenta la influencia de estos sectores excluidos. Se calculó cada par de matrices de vecindad propuestas con base en el subconjunto de datos usados en los espacios verdes y para sectores del análisis del arbolado urbanos. Así la matriz de vecindad para los SU usados para la estimación de los coeficientes de las regresiones lineales se ve en la figura 3.13.

La matriz  $W_d$  impone un estructura de interacciones que puede relacionar sectores en una zona que no necesariamente comparten ningún lado o esquina pero que están cercanos, mientras que  $W_q$  se restringe a condiciones de vecindad sólo entre sectores contiguos. En esa medida puede existir juego para dar interpretación teórica al fenómeno de derrame o influencia que ejercen sobre el ajuste de los modelos. Un ejemplo puede ser que la dependencia espacial de las coberturas de árboles que expresa  $W_d$  es una característica en una zona no limitada por las divisiones del territorio con base en los desarrollos urbanísticos (barrios) sino que se ajustan más a fenómenos de dispersión continuos con base en el alcance



Figura 3.12: Grafo de vecindad entre todos los SU de la ciudad de Cali

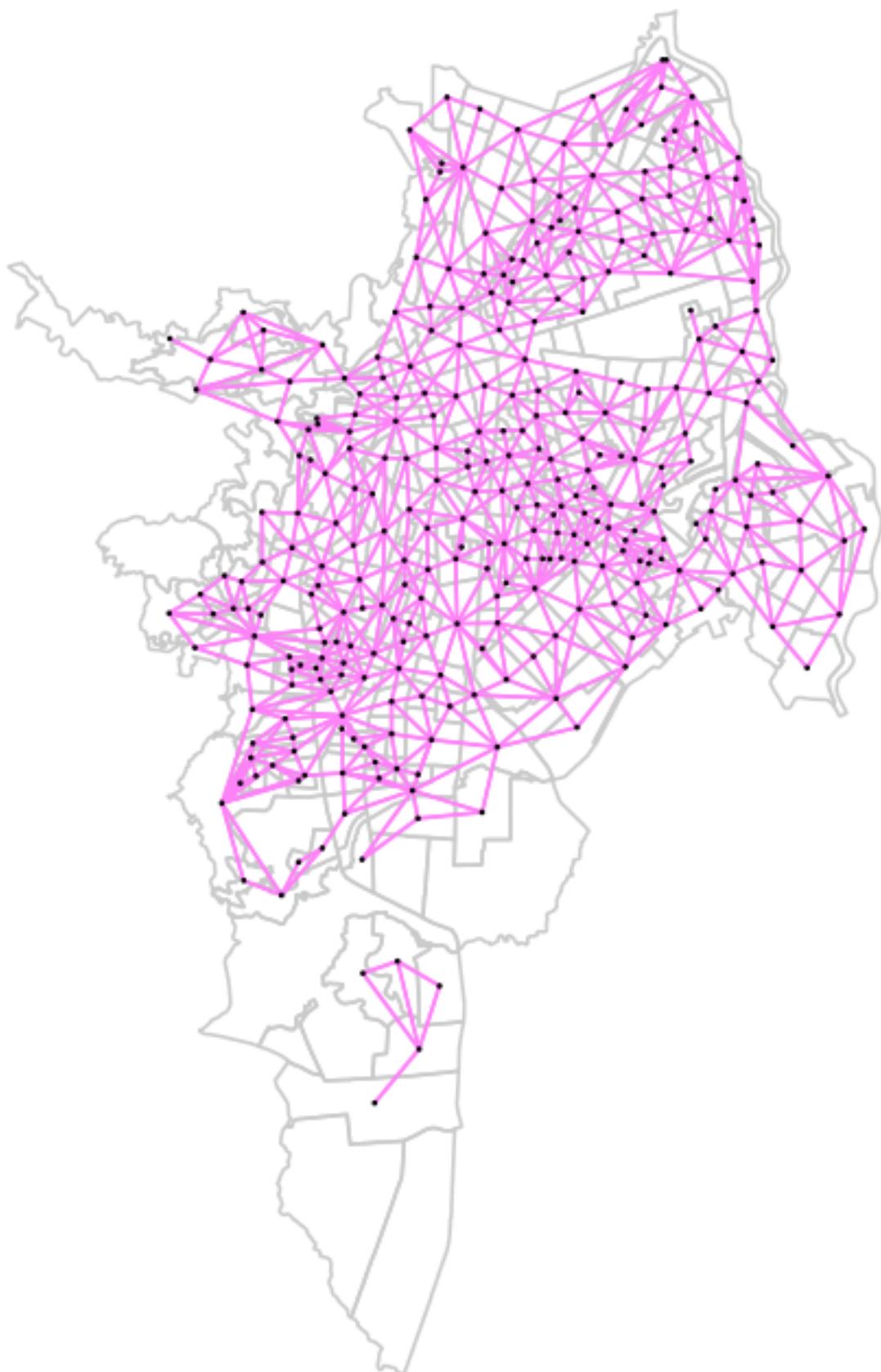


Figura 3.13: Grafo de vecindad entre los SU seleccionados para el análisis del AU

escogido. Así, la  $W_q$  puede interpretarse como una forma de dar relevancia a la continuidad entre barrios y su importancia como unidad de desarrollo urbano en las variaciones de la variable a predecir.

### 3.3.6.2. Autocorrelación espacial

Para indagar sobre la información o patrones espaciales de los residuos de los modelos de regresión usaremos el índice de Moran I. El índice de Moran I es el coeficiente de correlación para la relación entre una variable y sus valores circundantes. Si encontramos una correlación espacial significativa en los residuos, esto sugiere que agregando esa estructura en el modelo podremos obtener una estimación más eficiente de los coeficientes, y en consecuencia un mejor entendimiento de la relación entre esas variables. Hay que recordar que en este ejercicio no estamos queriendo entender una población por una muestra, estamos calculando estos coeficientes sobre el total de la población, y por tanto los coeficientes pueden interpretarse como la fuerza de esa relación. La confianza en esa estimación depende de que los residuos obtenidos sean tenga un valor medio de 0, y que no puedan distinguirse del ruido. La ecuación (3.3) define matemáticamente el índice:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2} \quad (3.3)$$

donde  $N$  es el número de unidades espaciales indexados por  $i$  y  $j$ ;  $X$  es la variable de interés;  $\bar{X}$  es la media de  $X$ ; y  $w_{ij}$  es un elemento de una matriz de pesos espaciales  $W$ . Un valor de 0 de Moran'I indica un patrón espacial aleatorio. Si existe autocorrelación los valores son positivos y el máximo es 1. Si los valores son negativos decimos que existe dispersión, siendo -1 el mínimo valor posible representando la dispersión perfecta.

El gráfico de Moran es una forma de observar el valor de la pendiente (el índice de autocorrelación) graficando los valores retardados (spatial lag: es como dijimos previamente el valor medio de los valores vecinos) de la variable en cuestión en el eje  $y$  y la variable en el

$x$ . El valor  $p$  del test estadístico nos dice qué tan seguros estamos que esa pendiente no es plana, por lo que se espera que sean menores que el valor límite de significancia  $\alpha = 0,05$

El hecho de que de la Morán I es una suma de productos cruzados individuales es explotado por los “indicadores locales de asociación espacial” (LISA) para evaluar la agrupación de las unidades individuales mediante el cálculo de la I de Moran local para cada unidad espacial y la evaluación de la significación estadística para cada I. (Wikipedia, 2017)

Estos mapas acompañan el resultado numérico y el gráfico de Moran representan el valor z-normalizado del LISA, el valor  $p$  y el mapa de clusters. En este último mapa las regiones resaltadas en rojo tienen valores altos de la variable y tienen vecinos con valores altos también (*high-high*). El área azul es *low-low* los grupos presentan valores bajos al igual que sus vecinos. Mientras que las regiones azul pálido son *low-high* y las áreas rosadas son *high-low* muestran correlación negativa, es decir valores muy diferentes a los de sus vecinos. Las regiones fuertemente coloreadas son aquellas que contribuyen significativamente a un resultado positivo de autocorrelación espacial global, mientras que los colores más claros contribuyen significativamente a un resultado de autocorrelación negativo.

### 3.3.6.3. Ajuste de modelos espaciales

Mejorar la especificación de los modelos lineales incluyendo términos de retardo espacial en la variable dependiente (SAR (2.8)) se hace para obtener una adecuada estimación de los coeficientes de las otras covariables en el modelo. Si optamos por un modelo de error espacial (SEM (2.9)) implica que no es necesario plantear efectos distintivos de la variable dependiente rezagada, y que es posible que ese efecto sea por otras variables no tenidas en cuenta: el agrupamiento espacial observado en la variable dependiente se explica simplemente por el patrón geográfico de variables independientes medidas y no medidas. El modelo SAR, en cambio, incorpora la influencia de variables independientes no medidas, pero también estipula un efecto adicional de valores de atributos vecinos,

es decir, la variable dependiente rezagada. Si incluimos el retardo sólo de las variables independientes (SLX (2.10)) esperamos que los cambios en las dimensiones expresadas con las predictores produce un efecto de derrame o influencia en los sectores vecinos.

¿Qué significa decir que la cantidad de cobertura de copa está relacionada con la de los sectores vecinos? ¿Son los procesos de reproducción del arbolado urbano un fenómeno independiente de las intervenciones de sus habitantes y de los urbanizadores? ¿Los habitantes que ven árboles en las cuadras o barrios aledaños deciden sembrar árboles en su vecindario? ¿Existen similitudes en las condiciones estructurales de los barrios en ciertas zonas de la ciudad que prefieren las personas con mejores condiciones sociales (tener estudios superiores p.e.)? ¿Qué tipo de pérdidas en la cobertura de copa están motivadas por la densificación de un sector? ¿Cómo afectan los cambios en los tipos de oferta habitacional en un sector la cobertura de copa de los sectores vecinos? (las viviendas tipo cuarto suelen ofrecerse en pensiones y ser más económicas que las casas o apartamentos).

La pregunta a hacerse es cómo saber cuál de los diferentes modelos espaciales es el que mejor representa el fenómeno que estamos modelando y si los datos respaldan nuestras convicciones teóricas. Si el modelo de retraso espacial que especifique es realmente el correcto, entonces ninguna dependencia espacial debe permanecer en los residuos, y podremos elaborar sobre el tipo de procesos que pueden verse representados.

Una alternativa metodológica es probar los 4 tipos de modelos con la matriz  $W$  que resultó capturar mejor la asociación espacial en los datos y comparar sus resultados.



# **Capítulo 4**

## **Resultados**

### **4.1. Modelando la cobertura de copa**

#### **4.1.1. Correlaciones y gráficos de dispersión bivariados**

En la figura 4.1 se explora las relaciones entre las variables de población en las unidades originales de los datos (número de personas); la matriz triangular superior muestra los coeficientes de correlación de Pearson, la diagonal contiene el histograma de frecuencias de la variable y la matriz triangular inferior muestra un gráfico de dispersión y la línea de tendencia usando un modelo lineal entre cada par de variables. Es notoria la alta correlación entre tener ningún estudio y tener alguna limitación física (0.88); pertenecer a una comunidad afrodescendiente y carecer de estudios (0.92) o ser afrodescendiente y tener alguna limitación (0.88). Esto representa una suma de condiciones desfavorables relacionadas entre sí, que desde el punto de vista del modelo sólo podrán ser representadas por una de las variables, la que mejor se relacione con la cobertura de copa y evitar así colinealidad entre los predictores.

Cuando realizamos la misma comparación entre las variables porcentuales (más la densi-

Comparación por pares entre predictores de población

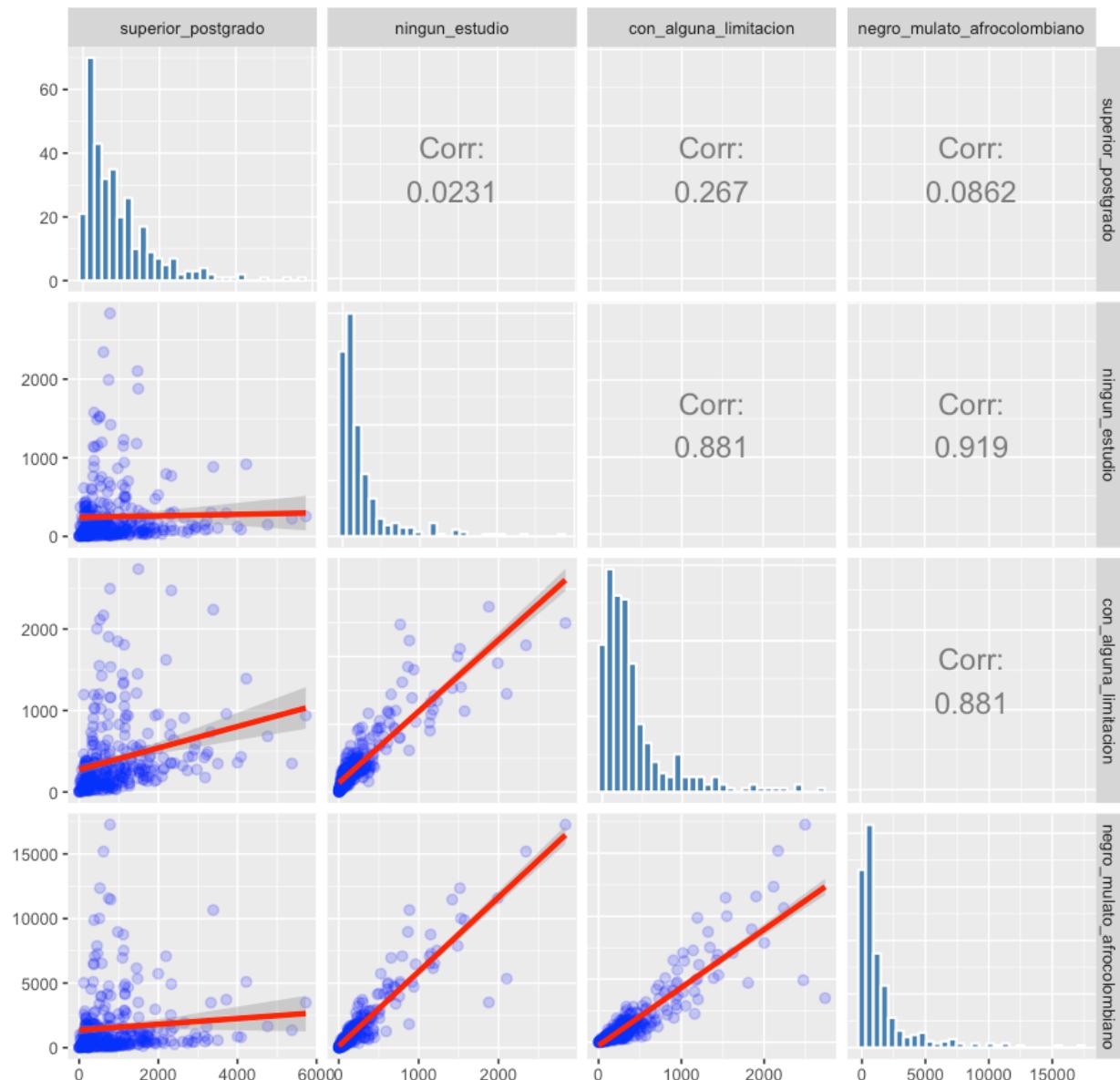


Figura 4.1: Comparación por pares entre predictores de población

dad poblacional) se observan patrones similares (ver figura 4.2): existe una alta correlación negativa entre el porcentaje de población afro de un sector y la tenencia de estudios superiores (-0.71), una fuerte asociación positiva entre el porcentaje de personas afro de un sector y el porcentaje de personas que carecen de estudios (0.68). También hay una fuerte relación inversa entre el porcentaje de personas de un sector sin estudios y el porcentaje de ellos que tiene estudios superiores (-0.8). Estas variables evitaremos usarlas como predictores en una misma formulación para no sesgar la estimación con problemas de colinealidad. Existe también una asociación, no tan fuerte pero importante, entre la densidad de población y sectores con mayor porcentaje de personas afro (0.47) y una asociación negativa entre la densidad de población y el porcentaje de personas con estudios superiores (-0.51). Estos resultados hablan de una concentración de condiciones desfavorables para la población, posiblemente acompañado de una segregación racial alta.

Ya que hemos explorado visualmente la dispersión entre los datos crudos, podemos usar una forma resumida, como los gráficos de azulejos o de matriz para consultar la intensidad de estas relaciones (figura 4.3 sintetiza las relaciones lineales entre las variables dependientes, mientras que la figura 4.3 lo hace para las no lineales)

Para seleccionar las variables que mejor predicen la cobertura de copa aplicamos un procedimiento análogo al realizado con las variables dependientes entre sí. Con base en los coeficientes de correlación de Pearson (figura 4.5) y Spearman (figura 4.6) entre las variables dependientes e independientes, y teniendo en cuenta las restricciones de colinealidad entre las variables dependientes, seleccionamos las variables a usar en el modelo lineal.

Así pues, para el área de copa (`area_copa`) los predictores seleccionados son `superior_postgrado`, `densidad_poblacion`, `con_alguna_limitacion.porcentaje`, `afro.porcentaje` y para la cobertura de copa en área pública (`cobertura_copa.ap`) los predictores seleccionados son `superior_postgrado.porcentaje`, `densidad_poblacion`, `con_alguna_limitacion.porcentaje`, `afro.porcentaje`

Para la selección de variables sobre uso de los predios, los tipos de vivienda, áreas de

Comparación por pares entre predictores de población porcentuales y la densidad de población

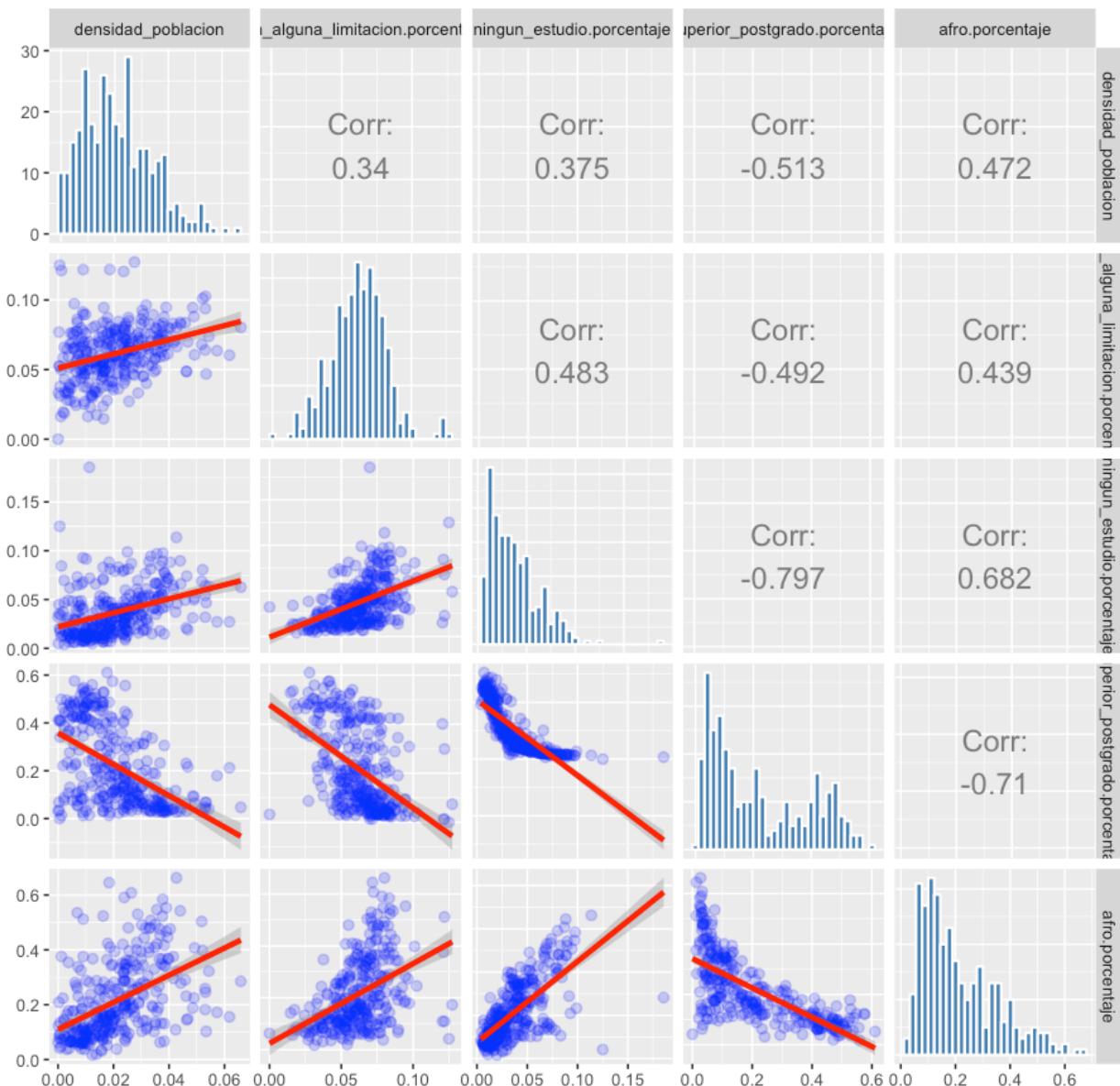


Figura 4.2: Comparación por pares entre predictores de población porcentuales

## Coeficiente Pearson entre variables de población

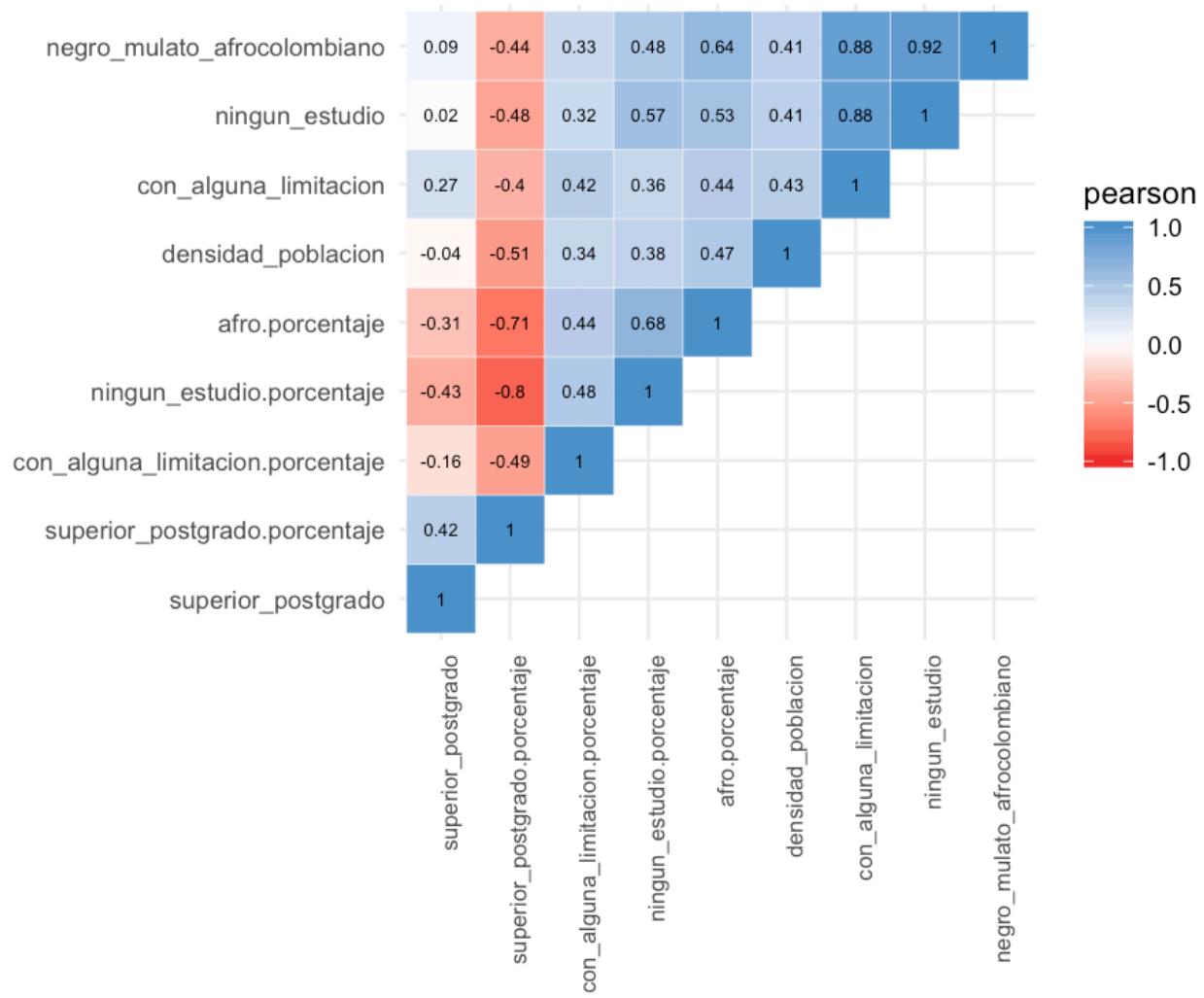


Figura 4.3: Coeficiente Pearson entre variables de población

## Coeficiente Spearman entre variables de población

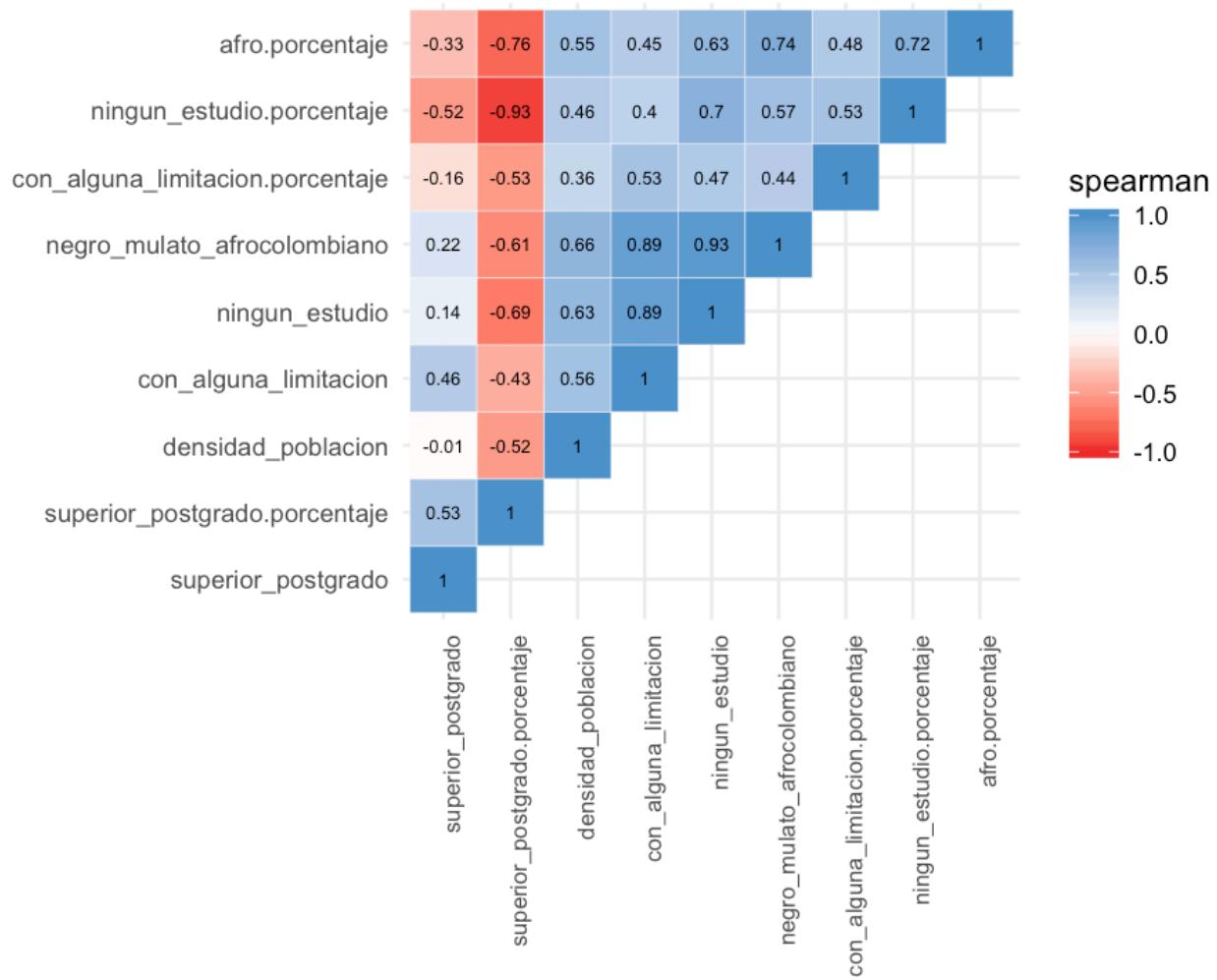


Figura 4.4: Coeficiente Spearman entre variables de población

## Pearson entre cobertura de copa y variables de población

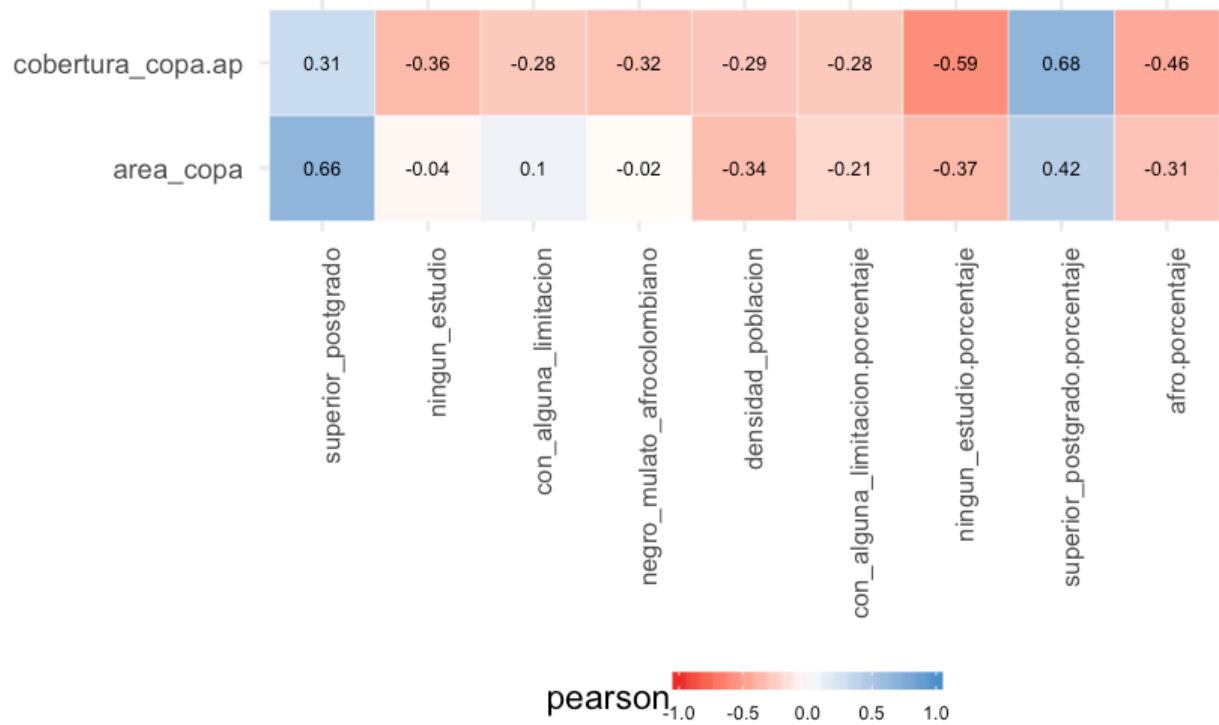


Figura 4.5: Coeficiente Pearson entre cobertura de copa y variables de población

## Spearman entre cobertura de copa y variables de población

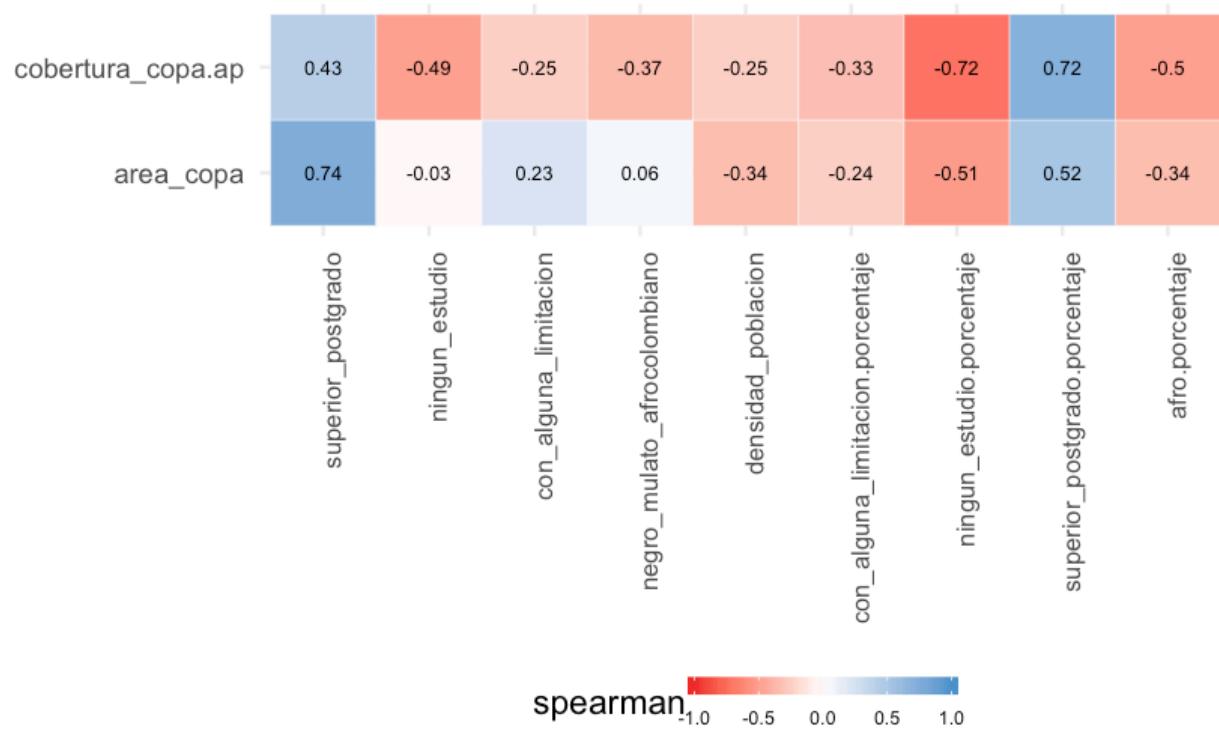


Figura 4.6: Coeficiente Spearman entre cobertura de copa y variables de población

espacio público y manzanas agregadas por sector urbano aplicamos el mismo proceso realizado para la selección de las variables de población usando como criterio la inclusión de variables con coeficientes de correlación que muestran una asociación fuerte con las variables de cobertura y área de copa.

Las figuras 4.7 y 4.8 se ven los coeficientes de Pearson y Spearman, respectivamente, entre las variables sobre el uso los predios, tipo de viviendas y área (y porcentaje) de espacios verdes en cada sector urbano. Como se observa, existe una fuerte (perfecta) asociación negativa entre el porcentaje de casas y apartamentos, lo que obliga a solo escoger una de las dos en caso de haber una fuerte relación entre alguna de ellas con las variables de cobertura de copa. También hay una fuerte asociación positiva entre el área de espacios verdes y el porcentaje de área de espacio verdes en un sector urbano. Como mencionamos antes solo incluiremos una de las dos en caso de que ambas resulten fuertemente asociadas con las variables dependientes.

Las figuras 4.9 y 4.10 muestran la correlación entre las potenciales nuevas variables a incluir en el modelo y las variables dependientes de cobertura y área de copa. Para el área de copa se seleccionan el área de espacios verdes (`area_ep`) y el porcentaje de viviendas tipo cuarto (`cuarto.porcentaje`). Para el modelo de porcentaje de cobertura de se seleccionan `apartamento.porcentaje`, `cuarto.porcentaje` y `area_ep.porcentaje`.

#### 4.1.2. Modelos de regresión lineal

Antes de evaluar los modelos se aplicaron varias transformaciones en busca de normalizar las distribuciones de las variables dependientes. Las que mejor resultado arrojaron en la formulación de los modelos fueron la transformación logarítmica para el caso de el área de copa y la variable sin transformar en el caso de la cobertura de copa. A continuación se presentan los resultados de los modelos de regresión

## Pearson entre características de predios y EV

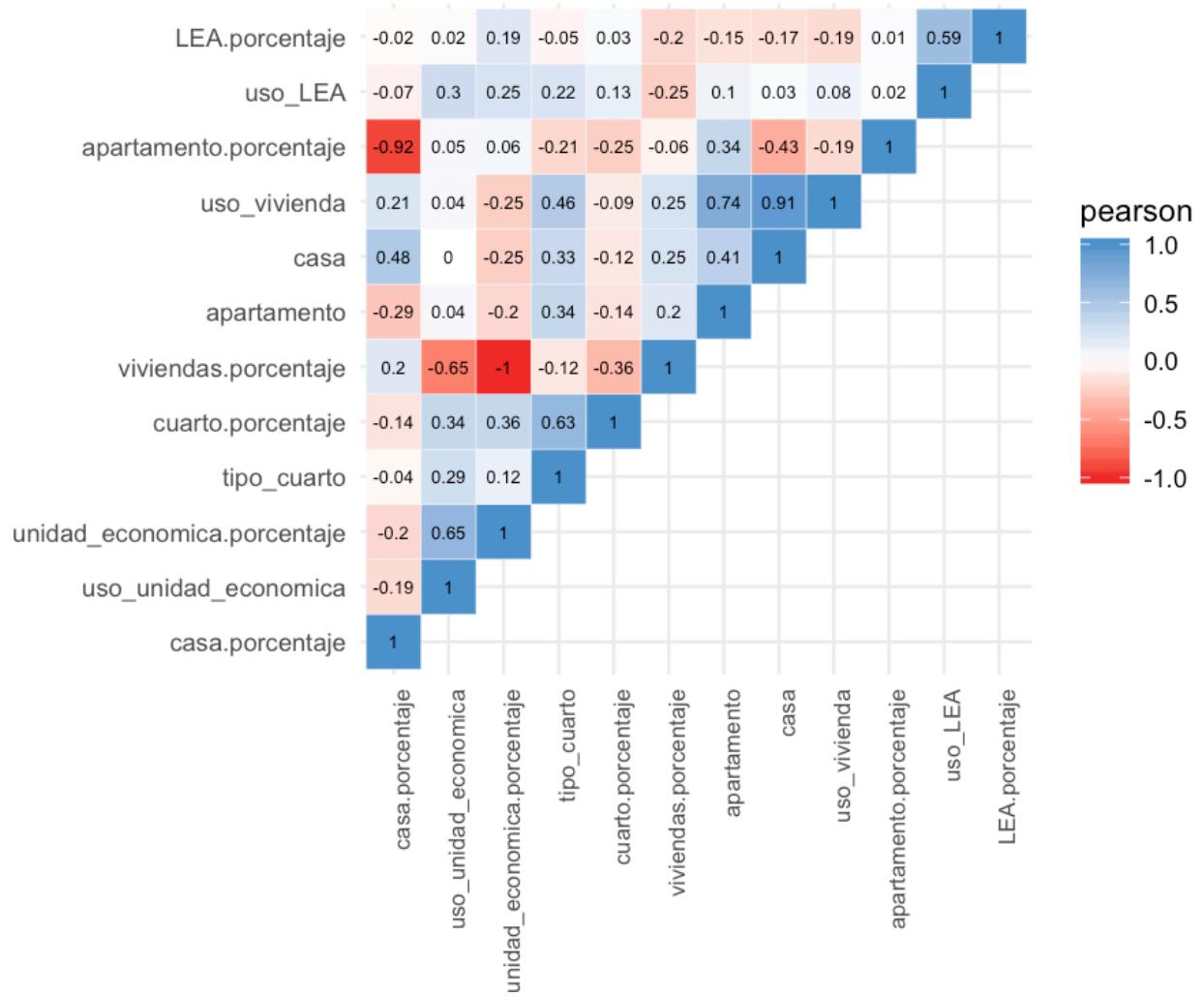


Figura 4.7: Coeficiente Pearson entre variables de uso de los predios y espacios verdes en los sectores urbanos

## Spearman entre entre características de predios y EV

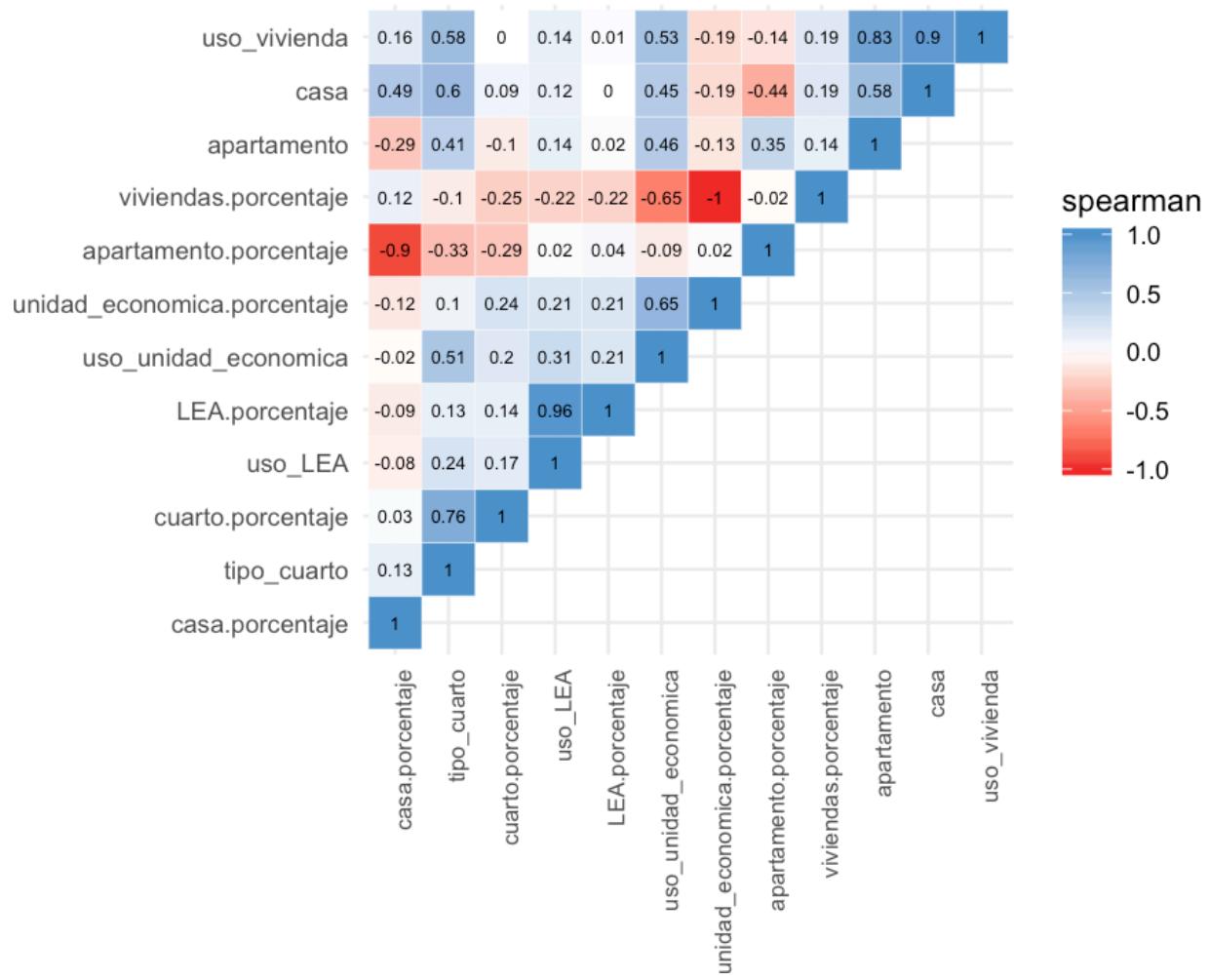


Figura 4.8: Coeficiente Pearson entre variables de uso de los predios y espacios verdes en los sectores urbanos

### Coeficiente Pearson entre cobertura de copa y uso de los predios

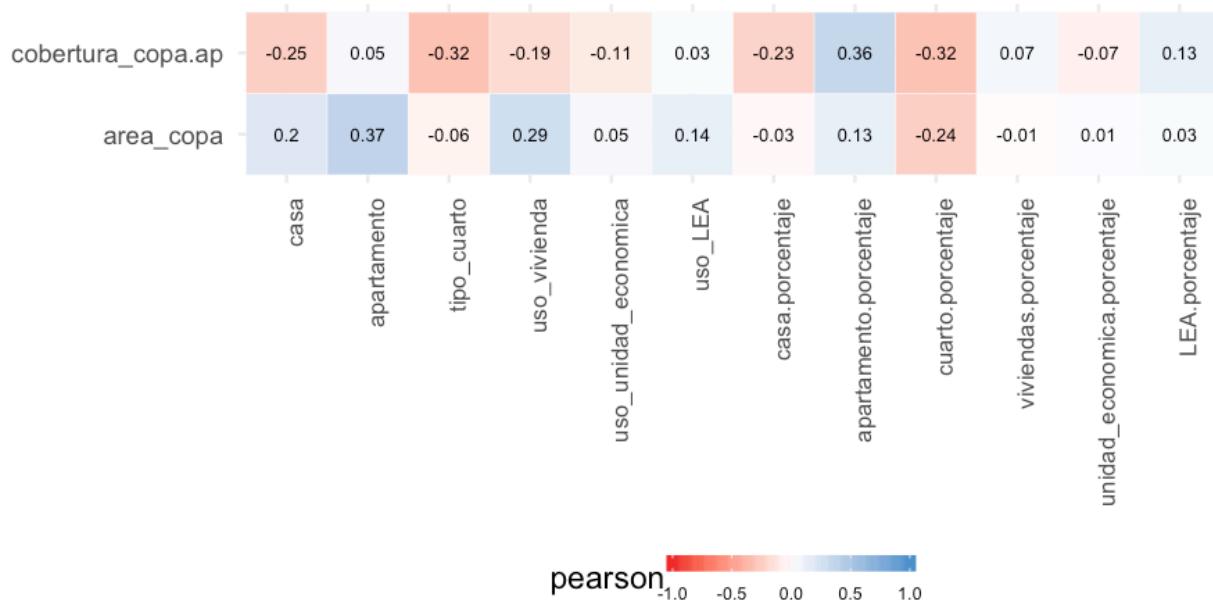


Figura 4.9: Coeficiente Pearson entre coberturas de copa y variables de predios y EV

### Coeficiente Spearman entre cobertura de copa y uso de los predios

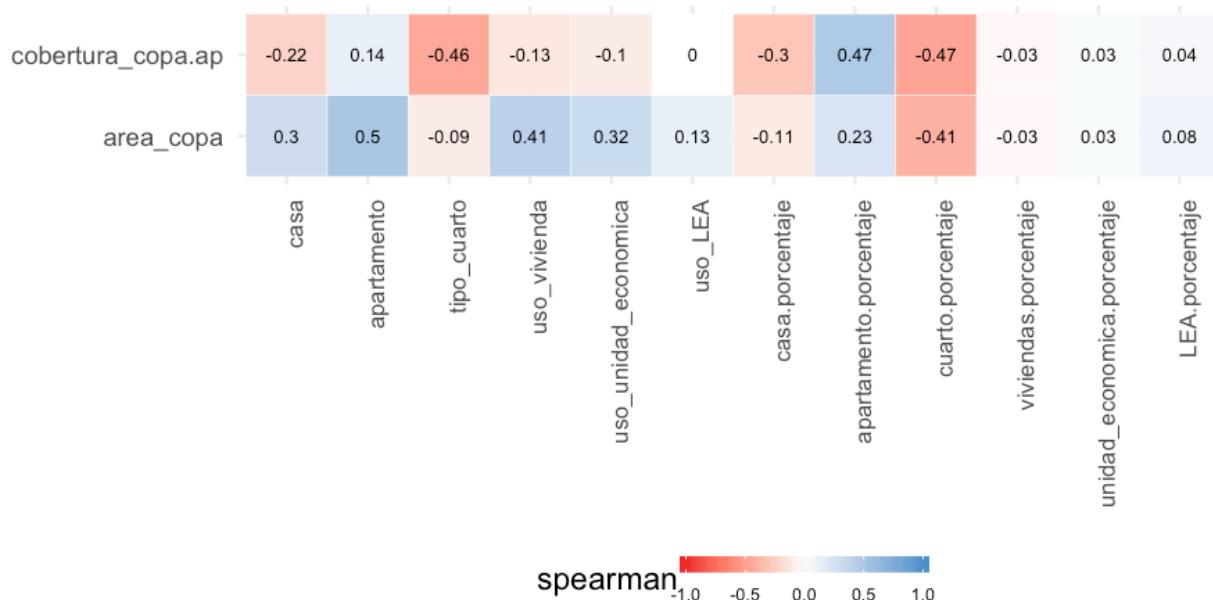


Figura 4.10: Coeficiente Pearson entre coberturas de copa y variables de predios y EV

#### 4.1.2.1. Área de copa

```
dependiente <- "log.area_copa"
independientes <- c(indep.poblacion.copa.sel,indep.predios.copa.sel)
# max normalizado
var_names<-c(dependiente,names(regresion.arboles[,independientes]))
regresion.arboles.mn<-max_nomalization(regresion.arboles,var_names)
f <- paste(str_c(dependiente,".mxn"),
           "~", paste(str_c(independientes,".mxn"),
                      collapse=" + "))
lm.mod.area_copa<-lm(f,
                       data = regresion.arboles.mn)
sm<-summary(lm.mod.area_copa)
sm

##
## Call:
## lm(formula = f, data = regresion.arboles.mn)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.214196 -0.031198  0.002999  0.040367  0.130671
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.78779   0.01265 62.256 < 2e-16
## superior_postgrado.mxn      0.31226   0.02400 13.013 < 2e-16
## densidad_poblacion.mxn     -0.14646   0.02011 -7.283 2.60e-12
## con_alguna_limitacion.porcentaje.mxn -0.02011   0.02474 -0.813   0.417
## afro.porcentaje.mxn          0.01004   0.02087  0.481   0.631
```

```

## cuarto.porcentaje.mxn          -0.14492   0.03048  -4.754 3.03e-06
## area_ep.mxn                   0.11908   0.02763   4.310 2.18e-05
##
## (Intercept)                  ***
## superior_postgrado.mxn      ***
## densidad_poblacion.mxn      ***
## con_alguna_limitacion.porcentaje.mxn
## afro.porcentaje.mxn
## cuarto.porcentaje.mxn          ***
## area_ep.mxn                   ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05934 on 317 degrees of freedom
## Multiple R-squared:  0.6121, Adjusted R-squared:  0.6048
## F-statistic: 83.38 on 6 and 317 DF,  p-value: < 2.2e-16

```

#### 4.1.2.2. Cobertura de copa

```

# cobertura AP
dependiente <- "cobertura_copa.ap"
independientes <- c(indep.poblacion.copa.ap.sel,
                     indep.predios.copa.ap.sel)
# max normalizado
var_names<-c(dependiente,independientes)
regresion.arboles.mn<-max_nomalization(regresion.arboles,var_names)
f <- paste(str_c(dependiente,".mxn"),
           "~", paste(str_c(independientes,".mxn"),
                      collapse=" + ")))

```

```

lm.mod.cobertura.ap<-lm(f,
                           data = regresion.arboles.mn)
sm<-summary(lm.mod.cobertura.ap)

sm

##

## Call:
## lm(formula = f, data = regresion.arboles.mn)
## 

## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.27874 -0.05872 -0.01222  0.03964  0.64364
## 

## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)               0.03438   0.03989   0.862   0.389    
## superior_postgrado.porcentaje.mxn 0.38210   0.04227   9.040 <2e-16  
## densidad_poblacion.mxn        0.05573   0.03750   1.486   0.138    
## con_alguna_limitacion.porcentaje.mxn 0.05848   0.04470   1.308   0.192    
## afro.porcentaje.mxn          -0.02170   0.04410  -0.492   0.623    
## apartamento.porcentaje.mxn     -0.04561   0.03547  -1.286   0.199    
## cuarto.porcentaje.mxn         -0.06197   0.06163  -1.005   0.315    
## area_ep.porcentaje.mxn        0.03971   0.03796   1.046   0.296    
## 

## (Intercept)
## superior_postgrado.porcentaje.mxn *** 
## densidad_poblacion.mxn
## con_alguna_limitacion.porcentaje.mxn
## afro.porcentaje.mxn

```

Cuadro 4.1: Resumen métricas de ajuste OLS para el área de copa (AC) y cobertura de copa (CC)

medidasfit	Log(AC)	%CC
Shapiro-Wilk	0.98196	0.90416
SW p-value	0.00043	0.00000
Breusch-Pagan	21.62587	19.67634
BP p-value	0.00142	0.00631
Media Residuos	0.00000	0.00000
MSE	0.00345	0.01071
adj-Rsquare	0.60479	0.45931
AIC	-901.82032	-532.26094
Log likelihood	458.91016	275.13047

```
## apartamento.porcentaje.mxn
## cuarto.porcentaje.mxn
## area_ep.porcentaje.mxn
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1048 on 316 degrees of freedom
## Multiple R-squared:  0.471,  Adjusted R-squared:  0.4593
## F-statistic:  40.2 on 7 and 316 DF,  p-value: < 2.2e-16
```

La tabla 4.1 resume las métricas de ajuste de ambos modelos

Ignoraremos las variables no significativas de los modelos lineales, para los siguientes pasos de la metodología.

Existe no normalidad en los residuos y posibles no linealidades, como se observa en las gráficas diagnósticas de la regresión 4.11.

Finalmente se presentan los mapas de las variables dependientes de los 2 modelos seleccionados, comparándolas con el modelo ajustado y los residuos del ajuste, con el fin de observar donde se localizan los errores en la predicción. La figura 4.13 corresponde al ajuste

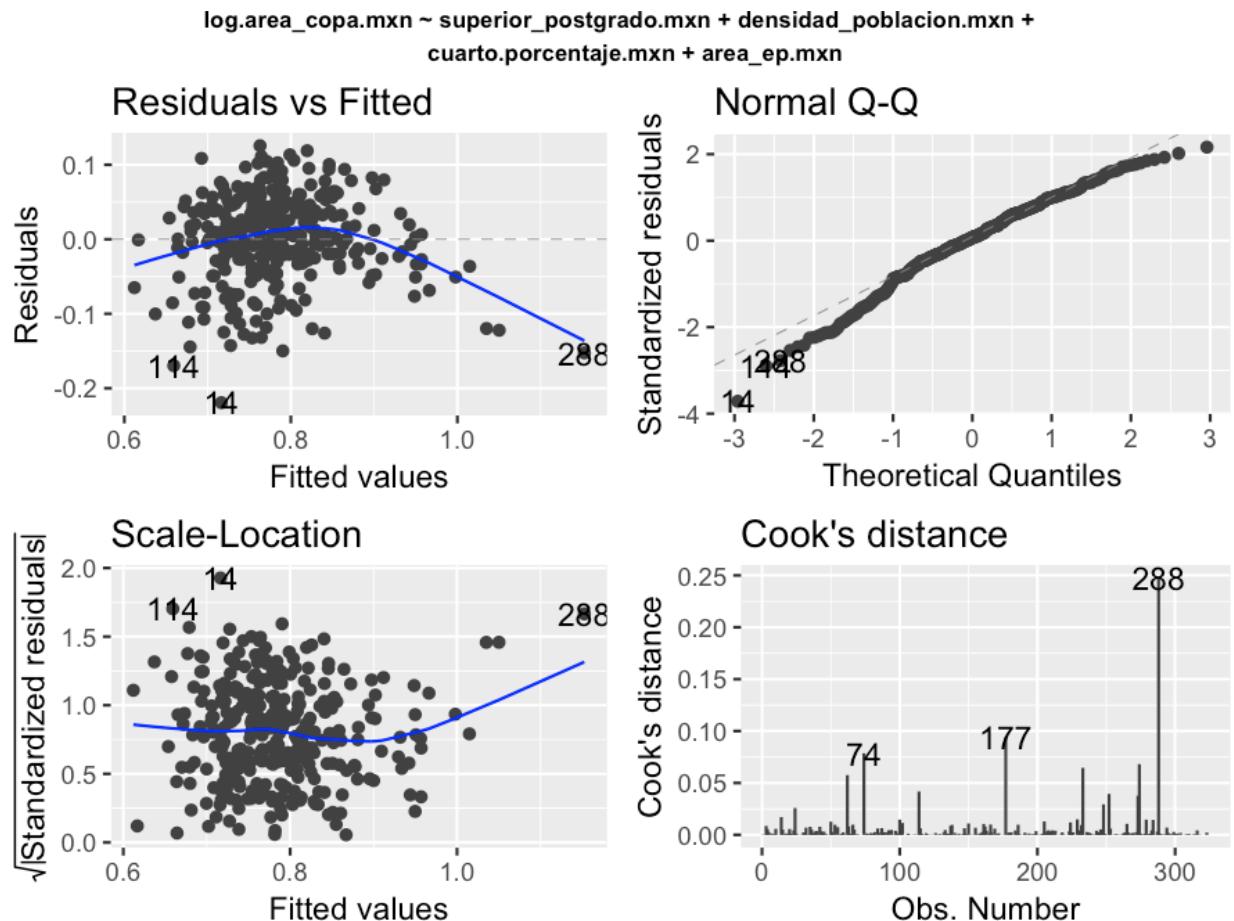


Figura 4.11: Gráficas diagnósticas para el análisis de regresión lineal de área de copa con los nuevos términos

```

cobertura_copa.ap.mx ~ superior_postgrado.porcentaje.mx + densidad_poblacion.mx +
con_alguna_limitacion.porcentaje.mx + afro.porcentaje.mx +
apartamento.porcentaje.mx + cuarto.porcentaje.mx + area_ep.porcentaje.mx

```

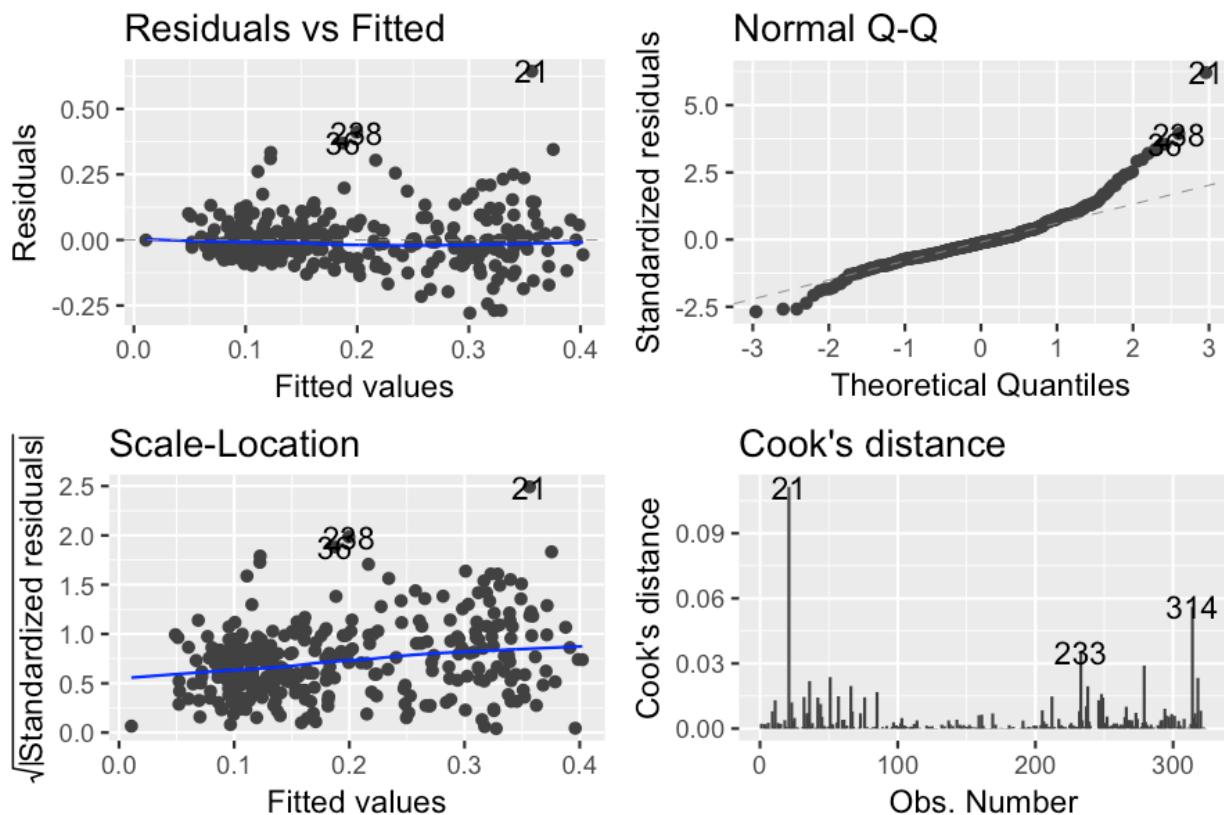


Figura 4.12: Gráficas diagnósticas para el análisis de regresión lineal de porcentaje de cobertura de copa.

$$\log.\text{area\_cpa.mx} \sim \text{superior\_postgrado.mx} + \text{densidad\_poblacion.mx} + \text{cuarto.porcentaje.mx} + \text{area\_ep.mx}$$

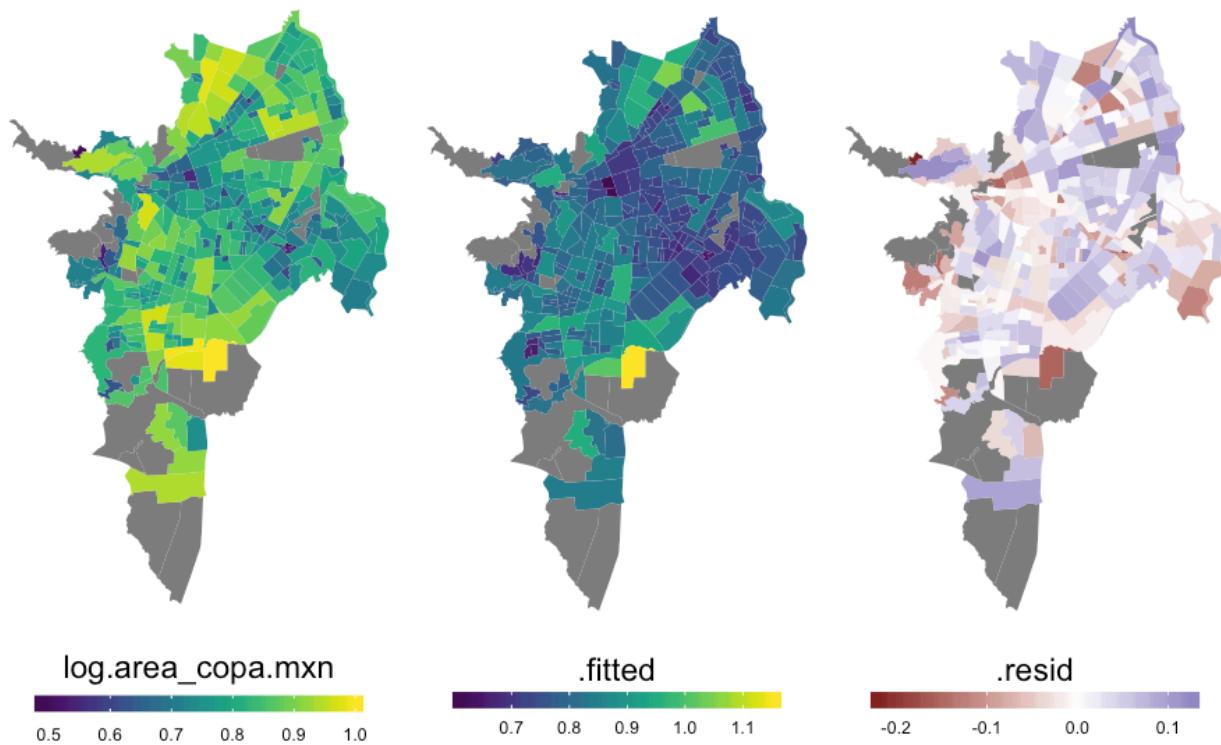


Figura 4.13: Variable dependiente, valor ajustado por el modelo y residuos del OLS para log.area\_copa normalizada

del área de copa y la figura 4.14 al porcentaje de la cobertura de copa.

#### 4.1.3. Modelado espacial AU

Las matrices de vecindad construidas para el análisis espacial son la *queen*  $W_q$ , que considera vecino a todos los sectores que comparten un lado o una esquina con un sector censal; y una matriz de distancia inversas entre los centroides de los SU, restringiendo la vecindad a aquellos centroides que están a menos de 1 km ( $W_d$ ). El valor de un kilómetro es arbitrario, aunque razonable en la escala humana. Los grafos que representan las 2 matrices  $W$  se muestran en la figura 4.15.

Examinemos primero la autocorrelación de las variables dependientes.

`cobertura_copa.ap.mx ~ superior_postgrado.porcentaje.mx`

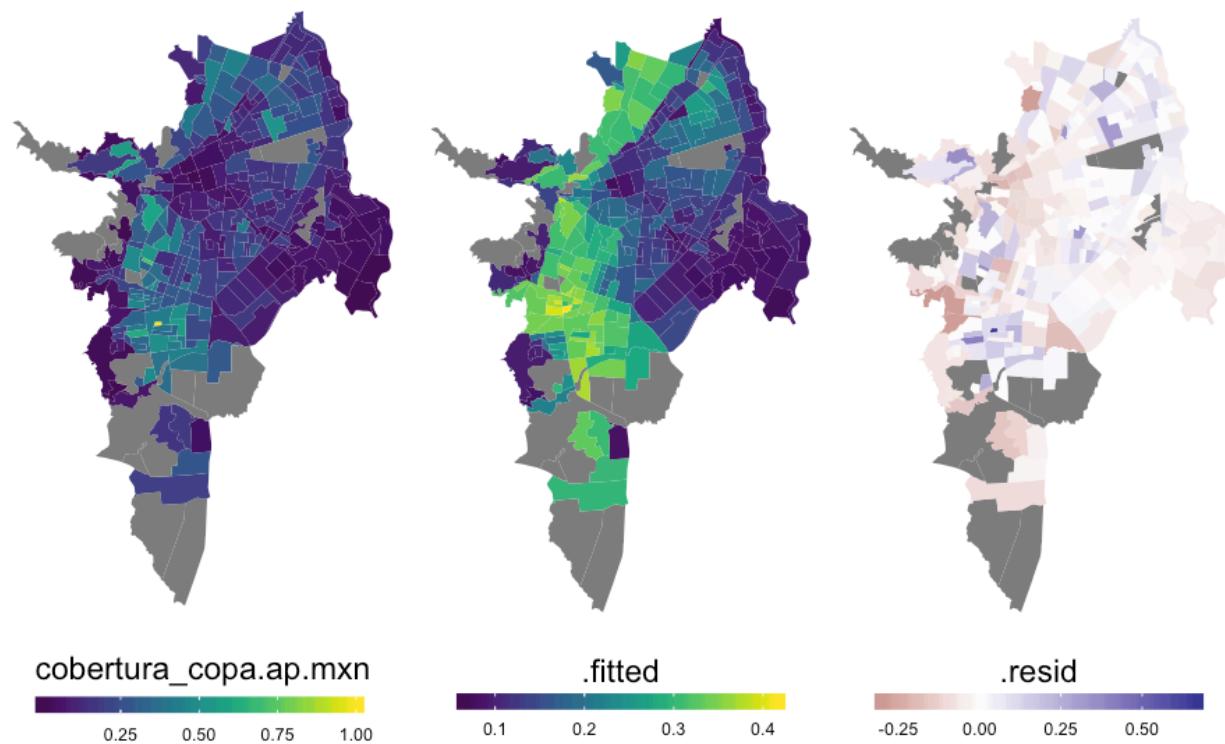


Figura 4.14: Variable dependiente, valor ajustado por el modelo y residuos del OLS para cobertura\_copa.ap normalizada

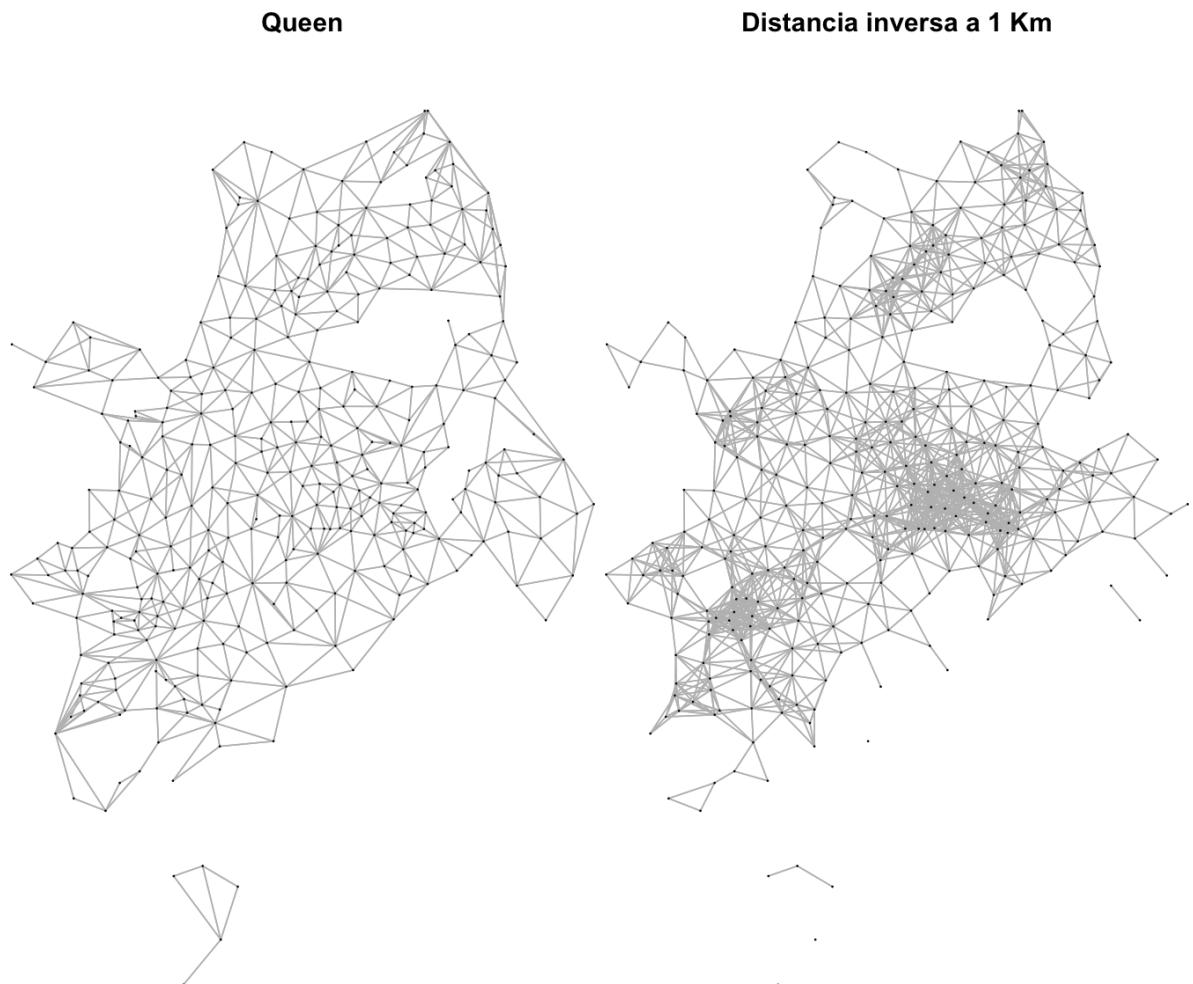


Figura 4.15: Matrices de vecindad del análisis espacial

#### 4.1.3.1. Autocorrelación variables dependientes

```
moran.test(lm.mod.area_copa$model$log.area_copa.mxn, listw=W_queen)

## 
## Moran I test under randomisation
## 
## data: lm.mod.area_copa$model$log.area_copa.mxn
## weights: W_queen
## 
## Moran I statistic standard deviate = 10.958, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.374881115     -0.003095975     0.001189781

moran.test(lm.mod.area_copa$model$log.area_copa.mxn, listw=W_dist1000.inv, zero.policy = 1)

## 
## Moran I test under randomisation
## 
## data: lm.mod.area_copa$model$log.area_copa.mxn
## weights: W_dist1000.inv
## 
## Moran I statistic standard deviate = 9.5098, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.2817402626    -0.0031250000    0.0008972897
```

La matriz  $W_q$  captura mucho mejor de forma global la autocorrelación del área de copa.

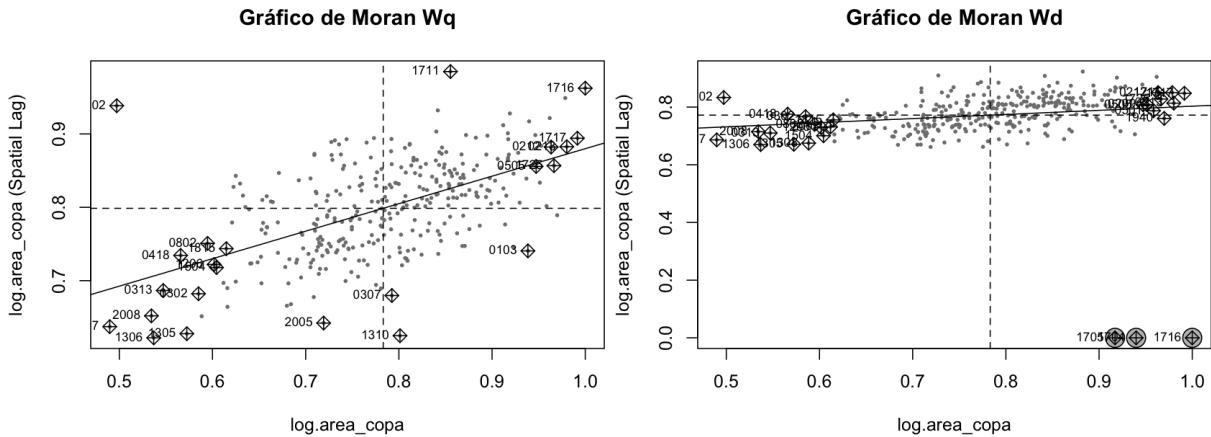


Figura 4.16: Gráfico de Moran para los residuos del modelo lineal para del área de cobertura de copa usando  $W_q$

Los mapas LISA muestran los focos de esta autocorrelación usando la matriz para  $W_q$  (figura 4.17) y para  $W_d$  (figura 4.18).

En los estos mapas de LISA de ambas matrices se observan que los sectores en rojo (H-H) y azul (L-L) identifican los lugares con autocorrelación positiva, formando grupos. No se presentan valores negativos de autocorrelación.

Para la variable dependiente porcentaje de cobertura de copa los resultados del análisis de autocorrelación se presentan a continuación.

```
moran.test(lm.best.cobertura.ap$model$cobertura_copa.ap.mxn, listw=W_queen)

##
##  Moran I test under randomisation
##
## data: lm.best.cobertura.ap$model$cobertura_copa.ap.mxn
## weights: W_queen
##
## Moran I statistic standard deviate = 14.107, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
```

### Mapas LISA - log.area\_copa - W\_q

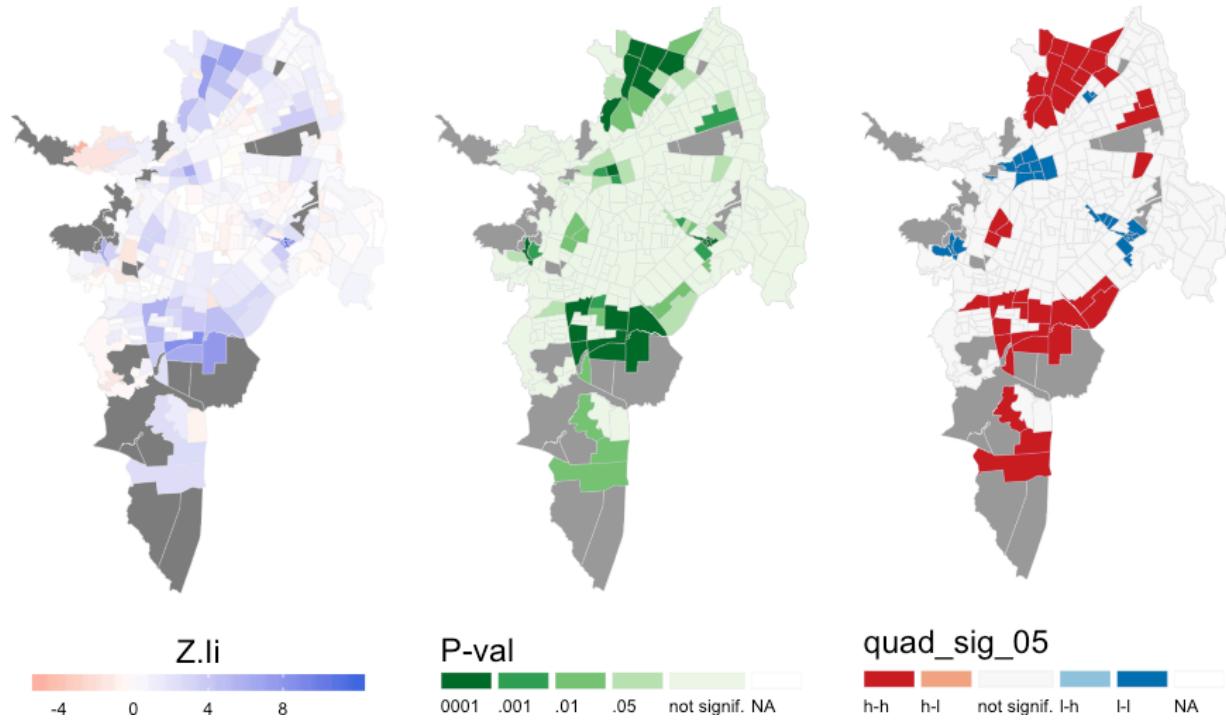


Figura 4.17: Mapas LISA para la matriz  $W_q$  de ‘log.area copa’

```

## Moran I statistic      Expectation      Variance
##                 0.481041234   -0.003095975   0.001177840

moran.test(lm.best.cobertura.ap$model$cobertura_copa.ap.mxnx, listw=W_dist1000.inv, zero.

##
## Moran I test under randomisation
##
## data: lm.best.cobertura.ap$model$cobertura_copa.ap.mxnx
## weights: W_dist1000.inv
##
## Moran I statistic standard deviate = 16.957, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance

```

Mapas LISA - log.area\_copa - W\_d

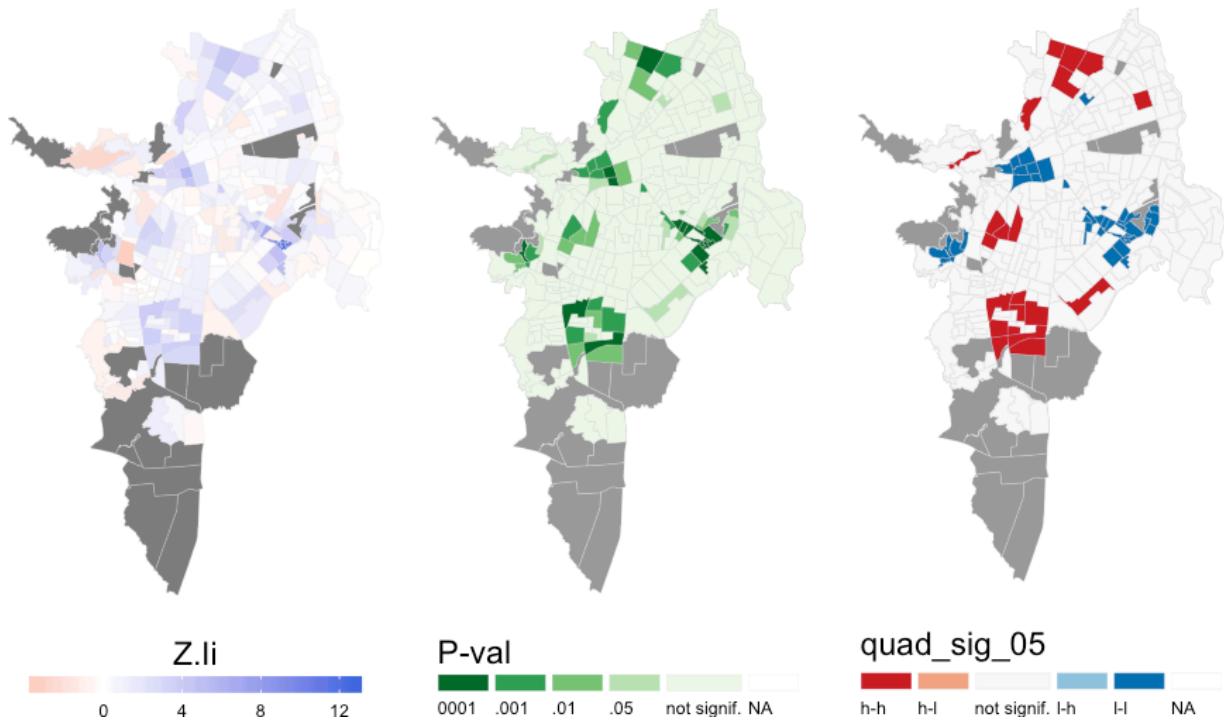


Figura 4.18: Mapas LISA para la matriz  $W_d$  de ‘log.area copa’

```
##          0.5022505537      -0.0031250000       0.0008881907
```

Ambos diseños de matriz revelan presencia clara de autocorrelación espacial, pero la matrix  $W_d$  la captura ligeramente mejor con un valor positivo y significativa, produciendo *clusters* de sectores urbanos más con más sectores urbanos. Los mapas LISA muestran los focos de esta autocorrelación usando la matriz para  $W_q$  (figura 4.20) y para  $W_d$  (figura 4.21) . No se presentan valores negativos de autocorrelación.

#### 4.1.3.2. Autocorrelación residuos de los OLS

El siguiente código calcula los índices de autocorrelación de los residuos del mejor modelo lineal de área de copa (`sqrt.copa_area.mxn ~ superior_postgrado.mxn + densidad_poblacion.mxn + con_alguna_limitacion.porcentaje.mxn + afro.porcentaje.mxn + cuarto.porcentaje.mxn + area_ep.mxn`) para ambas matrices  $W$  y construye los

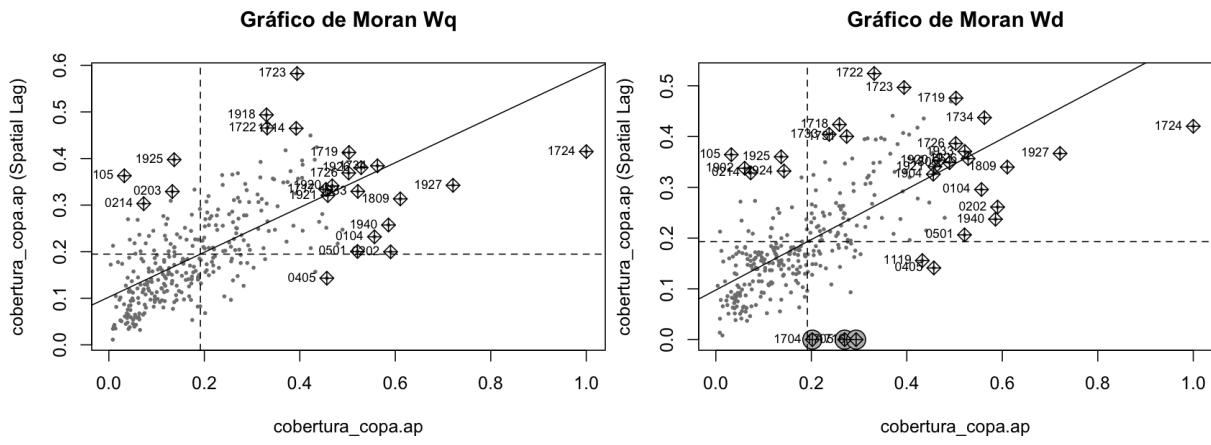


Figura 4.19: Gráfico de Moran para los residuos del modelo lineal para el porcentajes cobertura de copa usando  $W_q$

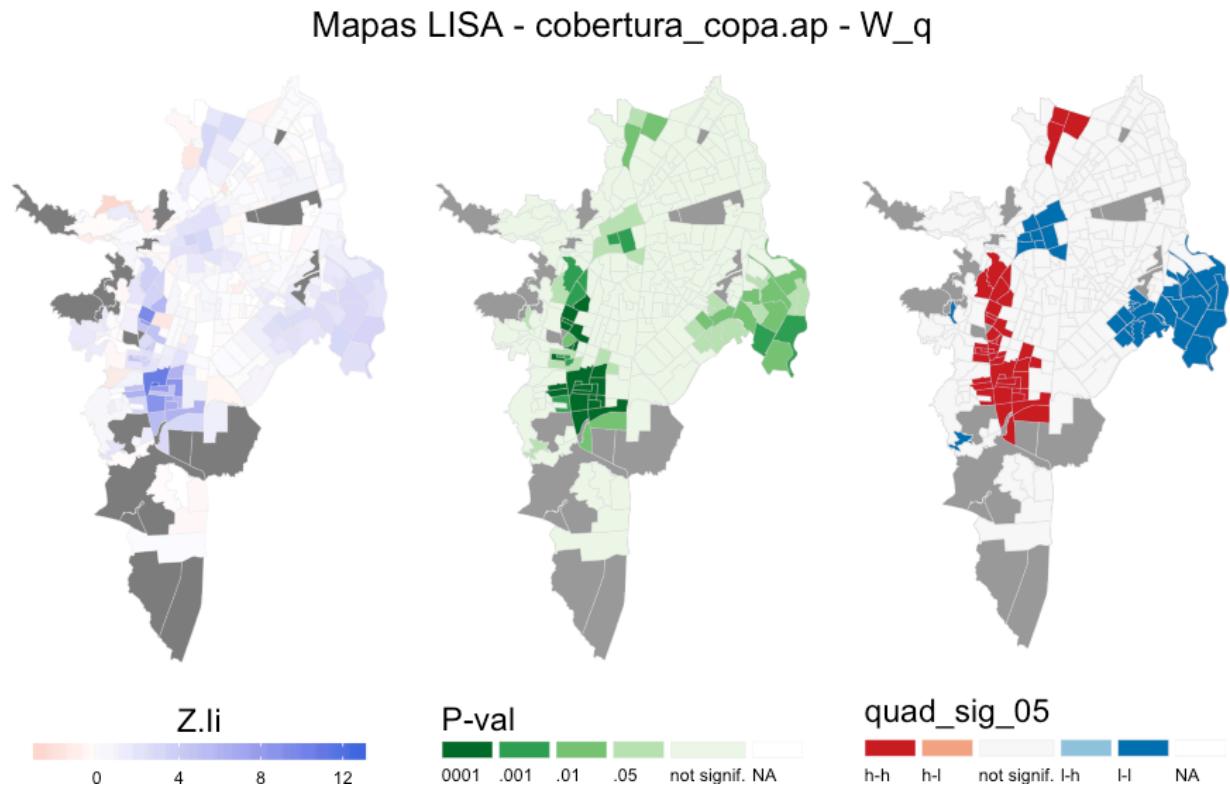


Figura 4.20: Mapas LISA para la matriz  $W_q$  de ‘cobertura\_copa.ap’

### Mapas LISA - cobertura\_copa.ap - W\_d

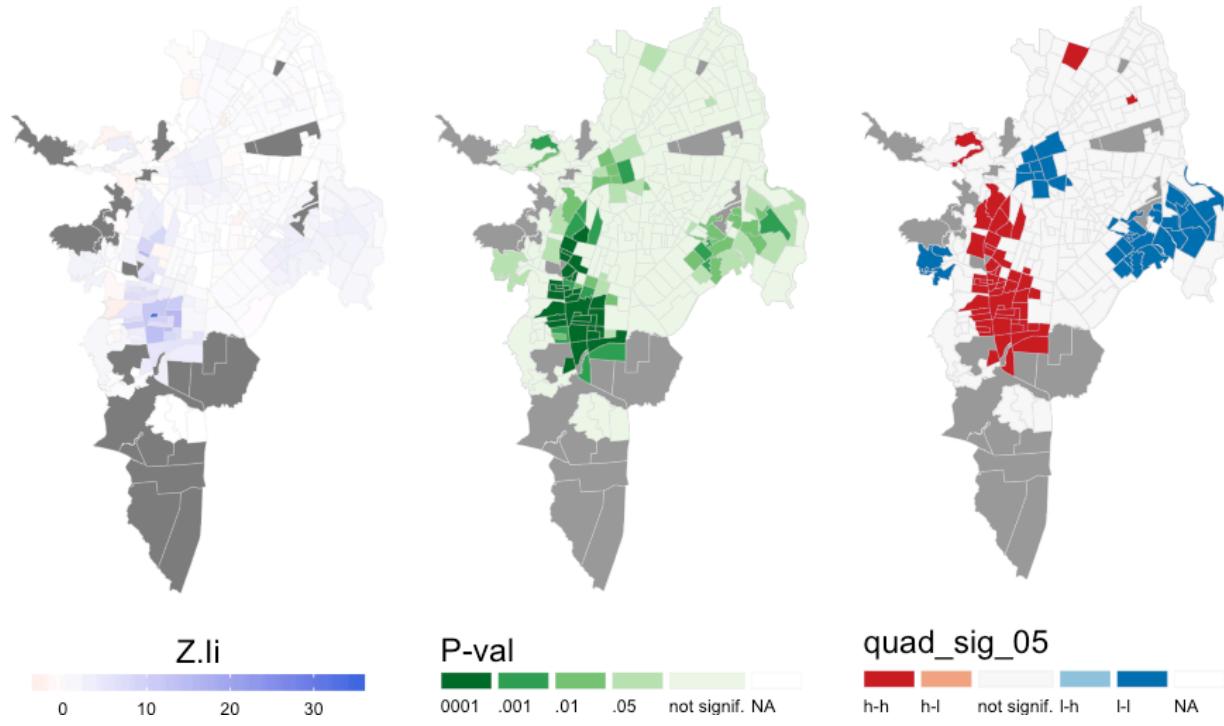


Figura 4.21: Mapas LISA para la matriz  $W_d$  de ‘cobertura copa.ap’

gráficos de Moran y mapas LISA 4.22 para  $W_q$  y para  $W_d$ .

```
lm.morantest(lm.mod.area_copa,
              W_queen, alternative="two.sided",
              zero.policy = T)

##
## Global Moran I for regression residuals
##
## data:
## model: lm(formula = f, data = regresion.arboles.mn)
## weights: W_queen
##
## Moran I statistic standard deviate = 3.0235, p-value = 0.002499
## alternative hypothesis: two.sided
```

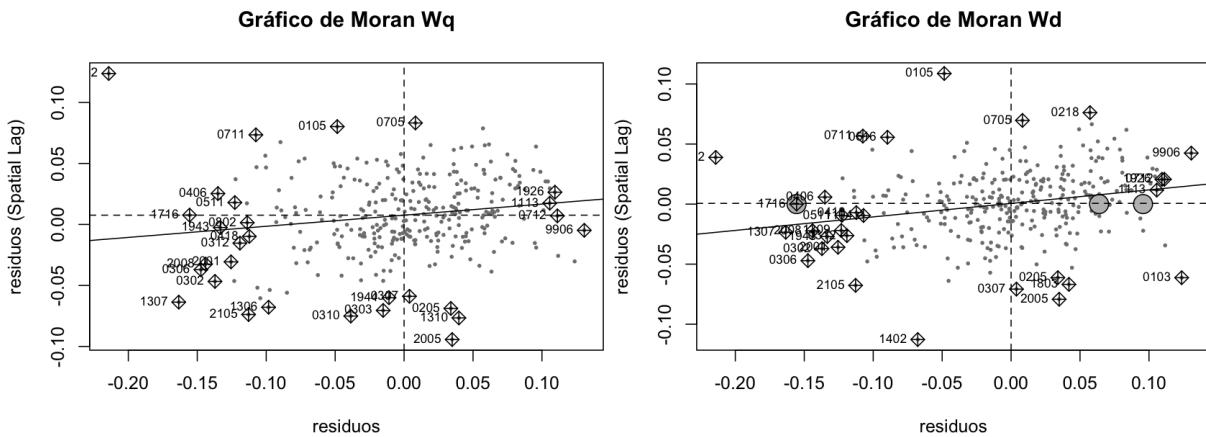


Figura 4.22: Gráfico de Moran para los residuos del modelo lineal para del área de cobertura de copa usando  $W_q$  y  $W_d$

```

## sample estimates:
## Observed Moran I      Expectation      Variance
##          0.09116750    -0.01123712     0.00114715

lm.morantest(lm.mod.area_copa,
             W_dist1000.inv,
             alternative="two.sided",zero.policy = T)

##
## Global Moran I for regression residuals
##
## data:
## model: lm(formula = f, data = regresion.arboles.mn)
## weights: W_dist1000.inv
##
## Moran I statistic standard deviate = 4.1797, p-value = 2.919e-05
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
##          0.1125770793   -0.0095502478     0.0008537482

```

## Mapas LISA - $W_q$

### Residuos:

`log.area_copa.mx ~ superior_postgrado.mx + densidad_poblacion.mx + cuarto.porcentaje.mx + area_ep.mx`

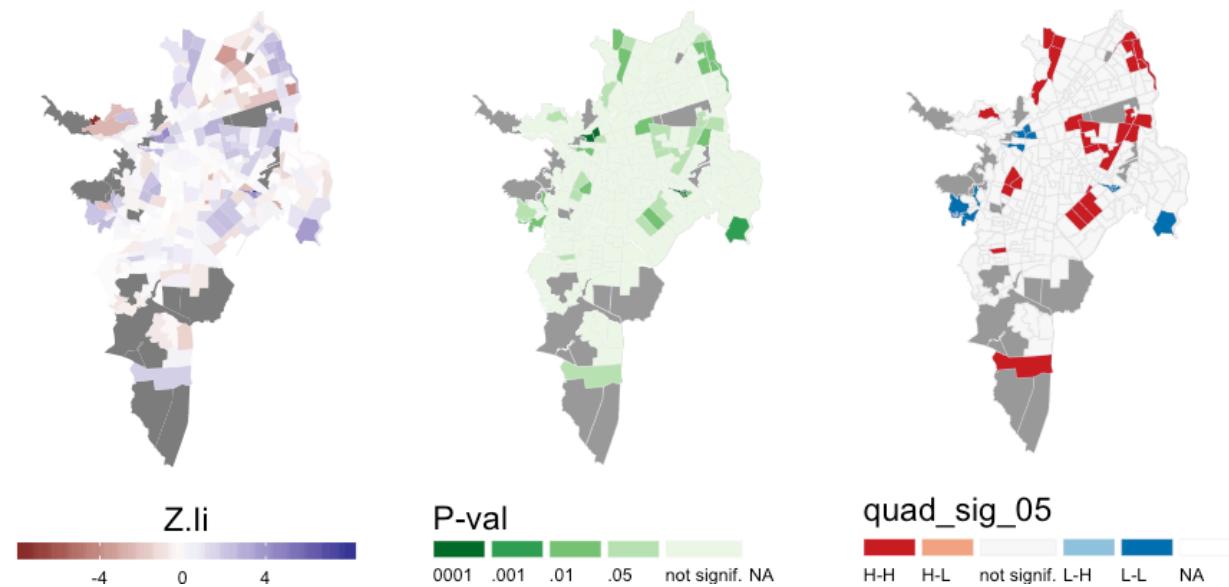


Figura 4.23: Mapas LISA para la matriz  $W_q$  de los residuos del modelo lineal para del área de copa

Para ambos casos el valor de Moran Global es mayor que 0 y significativo, aunque no es muy alto, si existe una tendencia en los datos y ambas matrices capturan el efecto.  $W_d$  lo hace mejor en los residuos, mientras  $W_q$  lo hace mejor con la variable dependiente. En las figuras 4.23 y 4.24 observamos que los grupos de sectores son más pequeños que los de las variables dependientes, lo que podría deberse a que las variables independientes presentan un patrón espacial similar, y que por consiguiente, al introducir estos retardos al modelo van a ayudar a absorber esa diferencias para mejorar la estimación.

El mismo análisis se aplica al modelo de porcentaje de cobertura de copa (`cobertura_copa.ap.mx ~ superior_postgrado.porcentaje.mx`).

```
lm.morantest(lm.best.cobertura.ap,
              W_queen, alternative="two.sided",
              zero.policy = T)
```

### Mapas LISA - Wd

Residuos:

$\log.\text{area\_copa.mxn} \sim \text{superior\_postgrado.mxn} + \text{densidad\_poblacion.mxn} + \text{cuarto.porcentaje.mxn} + \text{area\_ep.mxn}$

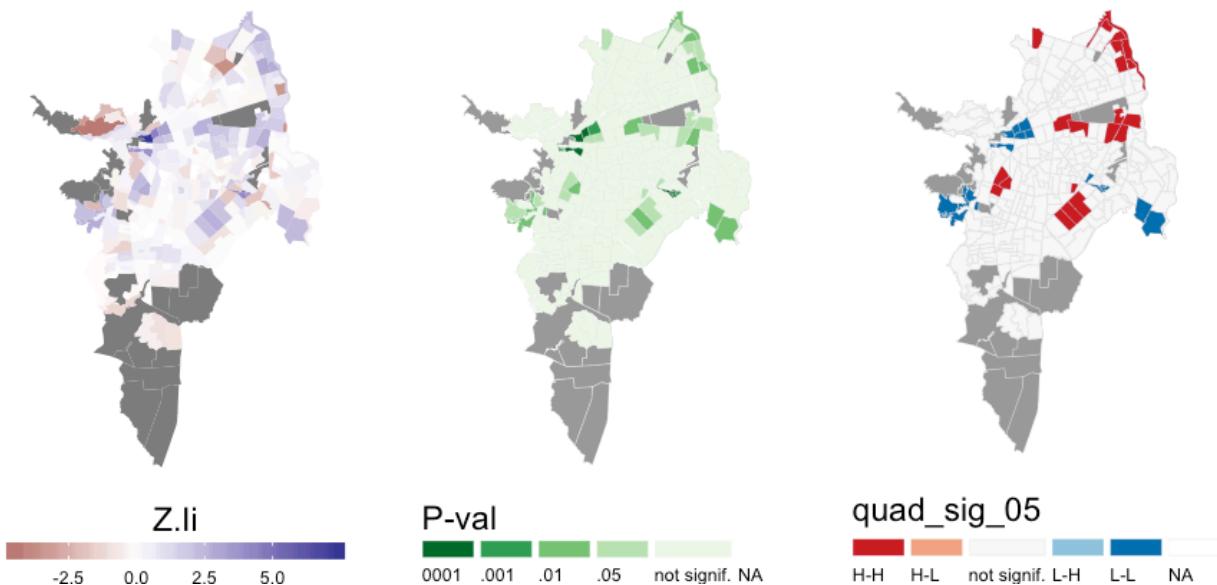


Figura 4.24: Mapas LISA para la matriz  $W_d$  dende los residuos del modelo lineal para del área de copa

```

##  

## Global Moran I for regression residuals  

##  

## data:  

## model: lm(formula = f, data = regresion.arboles.mn)  

## weights: W_queen  

##  

## Moran I statistic standard deviate = 5.8191, p-value = 5.915e-09  

## alternative hypothesis: two.sided  

## sample estimates:  

## Observed Moran I      Expectation      Variance  

##          0.194361169     -0.005463538     0.001179183

```

```

lm.morantest(lm.best.cobertura.ap,  

              W_dist1000.inv,  

              alternative="two.sided",  

              zero.policy = T)

```

```

##  

## Global Moran I for regression residuals  

##  

## data:  

## model: lm(formula = f, data = regresion.arboles.mn)  

## weights: W_dist1000.inv  

##  

## Moran I statistic standard deviate = 6.5028, p-value = 7.885e-11  

## alternative hypothesis: two.sided  

## sample estimates:  

## Observed Moran I      Expectation      Variance  

##          0.1882096889    -0.0053153442    0.0008856801

```

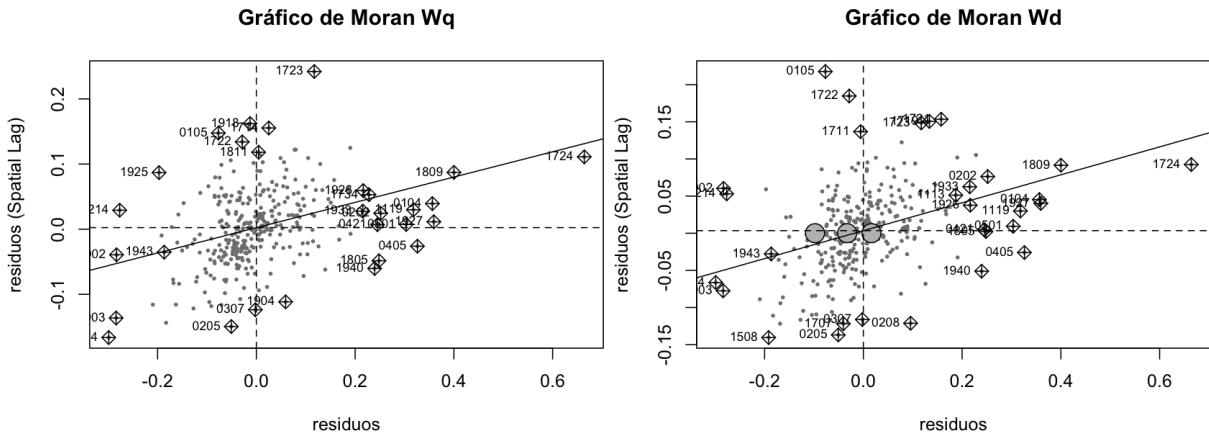


Figura 4.25: Gráfico de Moran para los residuos del modelo lineal para el porcentaje de cobertura de copa usando  $W_q$  y  $W_d$

Para ambas matrices el valor de Moran Global es mayor que 0 y significativo, casi el doble más alto que para el modelo de área de copa; el indicador porcentual de cobertura y el residuo del modelo para ajustarlo muestra el mismo comportamiento en la sensibilidad de las matrices: el efecto es más fuerte con  $W_d$  para la variable dependiente y levemente mayor con  $W_q$  en los residuos. En las figuras 4.26 y 4.27 una vez más observamos que los grupos de sectores son más pequeños que los de las variables dependientes, lo que, como dijimos, se debe a que las variables independientes presentan un patrón espacial similar, y se espera que al introducir estos efectos espaciales al modelo mejora la confiabilidad de la estimación.

Los resultados nos invitan a proceder a realizar un ajuste en ambos modelos introduciendo algún tipo de estructura espacial. Posiblemente la matriz  $W_q$  sea la candidata para el área de copa si se piensa en un modelo autorregresivo, pues parece capturar mejor en los test de autocorrelación global la existencia de grupos formados por fenómenos de dispersión o derrame en la distribución de la copa. En el caso de usar modelos de error espacial SEM posiblemente la candidata sea  $W_d$  por capturar mejor la autocorrelación en los residuos.

Mapas LISA -  $W_q$   
Residuos:  
`cobertura_copa.ap.mxn ~ superior_postgrado.porcentaje.mxn`

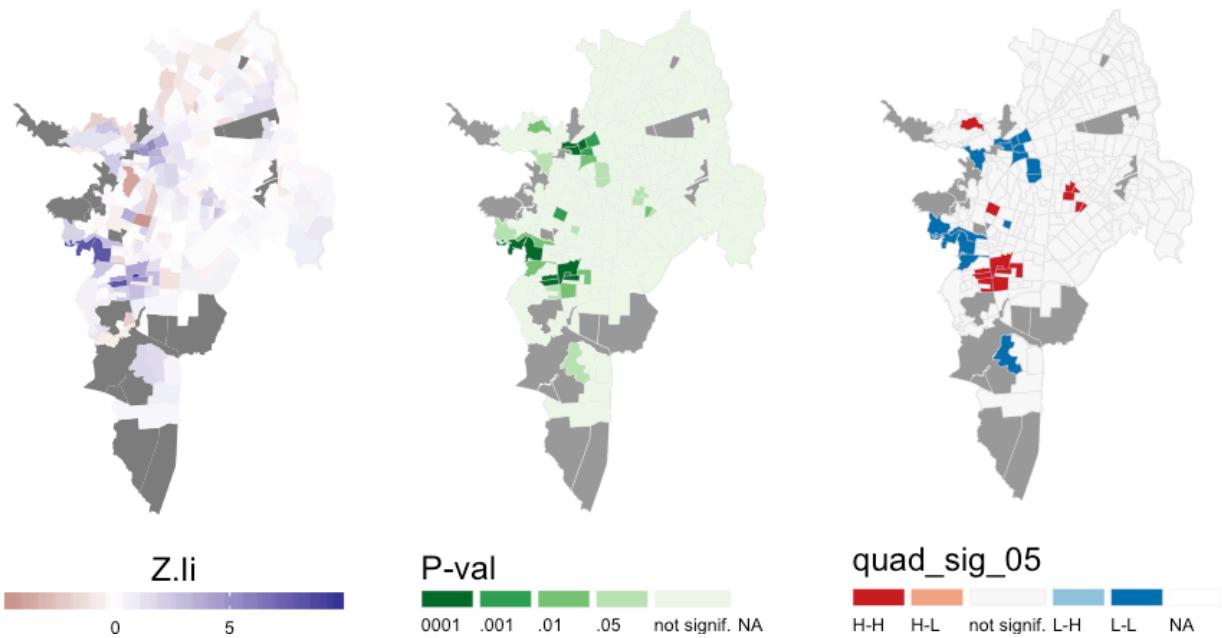


Figura 4.26: Mapas LISA para la matriz  $W_q$  de los residuos del modelo lineal para el porcentaje de cobertura copa

**Mapas LISA - Wd**  
**Residuos:**  
**cobertura\_copa.ap.mxn ~ superior\_postgrado.porcentaje.mxn**

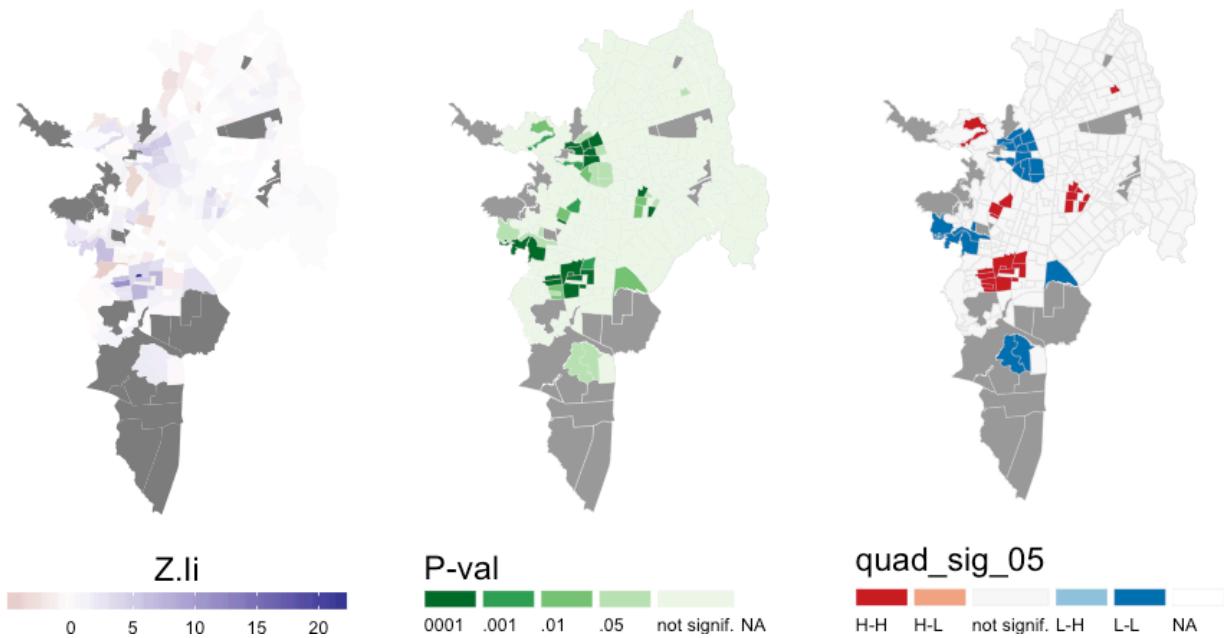


Figura 4.27: Mapas LISA para la matriz  $W_d$  dende los residuos del modelo lineal para del área de copa

#### 4.1.3.3. Modelo espacial área de copa

A continuación se presentan los resultados del modelado de los 4 tipos de modelos geoespaciales para el área de copa.

```
#SAR

sar.mod.log.area_copa.wq<-lagsarlm(formula = as.formula(lm.mod.area_copa),
                                       data = lm.mod.area_copa$model,
                                       listw = W_queen,
                                       zero.policy = T,
                                       tol.solve=1.0e-30)

sar.sm<-summary(sar.mod.log.area_copa.wq,
                  Nagelkerke=T)

sar.sm

## Call:
## lagsarlm(formula = as.formula(lm.mod.area_copa), data = lm.mod.area_copa$model,
##           listw = W_queen, zero.policy = T, tol.solve = 1e-30)
## 
## Residuals:
##       Min        1Q     Median        3Q       Max
## -0.248371 -0.030762  0.005022  0.041072  0.134767
## 
## Type: lag
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value
## (Intercept)               0.614048  0.051644 11.8901
## superior_postgrado.mxn   0.294621  0.023802 12.3780
## densidad_poblacion.mxn -0.124616  0.020126 -6.1918
```

```

## con_alguna_limitacion.porcentaje.mxn -0.010718  0.024028 -0.4461
## afro.porcentaje.mxn                  0.015215  0.020383  0.7465
## cuarto.porcentaje.mxn              -0.113944  0.030393 -3.7490
## area_ep.mxn                      0.113248  0.026952  4.2019
##
##                                     Pr(>|z|)
## (Intercept)                   < 2.2e-16
## superior_postgrado.mxn      < 2.2e-16
## densidad_poblacion.mxn      5.947e-10
## con_alguna_limitacion.porcentaje.mxn 0.6555402
## afro.porcentaje.mxn          0.4553948
## cuarto.porcentaje.mxn       0.0001775
## area_ep.mxn                 2.647e-05
##
## Rho: 0.20203, LR test value: 11.788, p-value: 0.00059614
## Asymptotic standard error: 0.059876
##      z-value: 3.3742, p-value: 0.00074018
## Wald statistic: 11.386, p-value: 0.00074018
##
## Log likelihood: 464.8042 for lag model
## ML residual variance (sigma squared): 0.0032968, (sigma: 0.057418)
## Nagelkerke pseudo-R-squared: 0.62599
## Number of observations: 324
## Number of parameters estimated: 9
## AIC: -911.61, (AIC for lm: -901.82)
## LM test for residual autocorrelation
## test value: 0.5245, p-value: 0.46893

# SEM

sem.mod.log.area_copa.wq<-errorsarlm(formula = as.formula(lm.mod.area_copa) ,
                                         data = lm.mod.area_copa$model,

```

```

listw = W_queen,
zero.policy = T,tol.solve=1.0e-30)

sem.sm<-summary(sem.mod.log.area_copa.wq,Nagelkerke=T)

sem.sm

##

## Call:

## errorsarlm(formula = as.formula(lm.mod.area_copa), data = lm.mod.area_copa$model,
##   listw = W_queen, zero.policy = T, tol.solve = 1e-30)
## 

## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2425228 -0.0282977  0.0045877  0.0403829  0.1265353
## 

## Type: error

## Coefficients: (asymptotic standard errors)

##                               Estimate Std. Error z value
## (Intercept)               0.78118860 0.01359311 57.4695
## superior_postgrado.mxn    0.30293183 0.02426334 12.4852
## densidad_poblacion.mxn   -0.13519457 0.02119774 -6.3778
## con_alguna_limitacion.porcentaje.mxn -0.01216501 0.02543709 -0.4782
## afro.porcentaje.mxn        -0.00099588 0.02319354 -0.0429
## cuarto.porcentaje.mxn     -0.13028543 0.03262446 -3.9935
## area_ep.mxn                0.13015543 0.02877820  4.5227
## 
##                               Pr(>|z|)

## (Intercept) < 2.2e-16
## superior_postgrado.mxn < 2.2e-16
## densidad_poblacion.mxn 1.797e-10
## con_alguna_limitacion.porcentaje.mxn 0.6325

```

```

## afro.porcentaje.mxn          0.9658
## cuarto.porcentaje.mxn       6.511e-05
## area_ep.mxn                 6.105e-06
##
## Lambda: 0.21436, LR test value: 6.4661, p-value: 0.010995
## Asymptotic standard error: 0.080871
##      z-value: 2.6507, p-value: 0.0080327
## Wald statistic: 7.0262, p-value: 0.0080327
##
## Log likelihood: 462.1432 for error model
## ML residual variance (sigma squared): 0.0033481, (sigma: 0.057862)
## Nagelkerke pseudo-R-squared: 0.6198
## Number of observations: 324
## Number of parameters estimated: 9
## AIC: -906.29, (AIC for lm: -901.82)

```

#SD

```

sd.mod.log.area_copa.wq<-lagsarlm(formula = as.formula(lm.mod.area_copa),
                                     data = lm.mod.area_copa$model,
                                     listw = W_queen,
                                     zero.policy = T,
                                     tol.solve=1.0e-30,
                                     type = "mixed")
sd.sm<-summary(sd.mod.log.area_copa.wq,
                Nagelkerke=T)
sd.sm

##
## Call:

```

```

## lagsarlm(formula = as.formula(lm.mod.area_copa), data = lm.mod.area_copa$model,
##           listw = W_queen, type = "mixed", zero.policy = T, tol.solve = 1e-30)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2043468 -0.0354340  0.0023694  0.0416533  0.1261893
##
## Type: mixed
## Coefficients: (asymptotic standard errors)

##                                     Estimate Std. Error z value
## (Intercept)                 0.6645922  0.0674240 9.8569
## superior_postgrado.mxn     0.2865849  0.0244089 11.7410
## densidad_poblacion.mxn    -0.0878170  0.0255961 -3.4309
## con_alguna_limitacion.porcentaje.mxn 0.0072678  0.0266239  0.2730
## afro.porcentaje.mxn        -0.0405391  0.0400062 -1.0133
## cuarto.porcentaje.mxn     -0.0617448  0.0415622 -1.4856
## area_ep.mxn                0.1338318  0.0300406  4.4550
## lag.superior_postgrado.mxn 0.0403785  0.0511699  0.7891
## lag.densidad_poblacion.mxn -0.1030583  0.0444954 -2.3162
## lag.con_alguna_limitacion.porcentaje.mxn -0.0597724  0.0462771 -1.2916
## lag.afro.porcentaje.mxn     0.1164235  0.0513221  2.2685
## lag.cuarto.porcentaje.mxn  -0.1232108  0.0675687 -1.8235
## lag.area_ep.mxn            -0.1034271  0.0417974 -2.4745
##
##                                     Pr(>|z|)
## (Intercept) < 2.2e-16
## superior_postgrado.mxn < 2.2e-16
## densidad_poblacion.mxn 0.0006016
## con_alguna_limitacion.porcentaje.mxn 0.7848675
## afro.porcentaje.mxn      0.3109068

```

```

## cuarto.porcentaje.mxn          0.1373848
## area_ep.mxn                   8.388e-06
## lag.superior_postgrado.mxn   0.4300497
## lag.densidad_poblacion.mxn   0.0205497
## lag.con_alguna_limitacion.porcentaje.mxn 0.1964890
## lag.afro.porcentaje.mxn      0.0232996
## lag.cuarto.porcentaje.mxn    0.0682293
## lag.area_ep.mxn              0.0133428
##
## Rho: 0.17513, LR test value: 4.8204, p-value: 0.028125
## Asymptotic standard error: 0.081498
##      z-value: 2.1488, p-value: 0.031647
## Wald statistic: 4.6175, p-value: 0.031647
##
## Log likelihood: 472.2662 for mixed model
## ML residual variance (sigma squared): 0.0031546, (sigma: 0.056166)
## Nagelkerke pseudo-R-squared: 0.64283
## Number of observations: 324
## Number of parameters estimated: 15
## AIC: -914.53, (AIC for lm: -911.71)
## LM test for residual autocorrelation
## test value: 1.2688, p-value: 0.26

```

### #SLX

```

slx.mod.log.area_copa.wq<-lmSLX(formula = as.formula(lm.mod.area_copa) ,
                                    data = lm.mod.area_copa$model,
                                    listw = W_queen,
                                    zero.policy = T)

slx.sm<-summary(slx.mod.log.area_copa.wq)

```

```
slx.sm
```

```
##  
## Call:  
## lm(formula = y ~ x - 1, weights = weights)  
##  
## Residuals:  
##       Min        1Q    Median        3Q       Max  
## -0.182288 -0.034947  0.001153  0.042134  0.130952  
##  
## Coefficients:  
##                                     Estimate Std. Error t value  
## x(Intercept)                      0.805190  0.019782 40.702  
## xsuperior_postgrado.mxn          0.287317  0.025135 11.431  
## xdensidad_poblacion.mxn         -0.090655  0.026359 -3.439  
## xcon_alguna_limitacion.porcentaje.mxn 0.006775  0.027434  0.247  
## xafro.porcentaje.mxn            -0.039136  0.041216 -0.950  
## xcuarto.porcentaje.mxn          -0.066535  0.042834 -1.553  
## xarea_ep.mxn                   0.130202  0.030934  4.209  
## xlag.superior_postgrado.mxn     0.103958  0.044022  2.362  
## xlag.densidad_poblacion.mxn     -0.130585  0.043688 -2.989  
## xlag.con_alguna_limitacion.porcentaje.mxn -0.069458  0.047652 -1.458  
## xlag.afro.porcentaje.mxn        0.124034  0.052871  2.346  
## xlag.cuarto.porcentaje.mxn      -0.146498  0.068425 -2.141  
## xlag.area_ep.mxn                -0.083521  0.042322 -1.973  
##                                     Pr(>|t|)  
## x(Intercept)                      < 2e-16 ***  
## xsuperior_postgrado.mxn          < 2e-16 ***  
## xdensidad_poblacion.mxn          0.000663 ***
```

```

## xcon_alguna_limitacion.porcentaje.mxn      0.805116
## xafro.porcentaje.mxn                      0.343078
## xcuarto.porcentaje.mxn                   0.121367
## xarea_ep.mxn                            3.36e-05 ***
## xlag.superior_postgrado.mxn            0.018817 *
## xlag.densidad_poblacion.mxn          0.003022 **
## xlag.con_alguna_limitacion.porcentaje.mxn 0.145953
## xlag.afro.porcentaje.mxn              0.019607 *
## xlag.cuarto.porcentaje.mxn            0.033051 *
## xlag.area_ep.mxn                      0.049328 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05792 on 311 degrees of freedom
## Multiple R-squared:  0.9948, Adjusted R-squared:  0.9946
## F-statistic:  4599 on 13 and 311 DF,  p-value: < 2.2e-16

AIC(slx.mod.log.area_copa.wq)

## [1] -911.712

```

Al comparar los resultados de las métricas de ajuste se identifica al modelo SD o mixto(variables con retardo y autocorrelación) con el mejor AIC. El modelo SD ha logrado eliminar la autocorrelación espacial global en los residuos, al igual que SAR y SEM (ver figura 4.28). Aunque persiste la no normalidad de los residuos y la heterocedasticidad como los muestran los test y las gráficas diagnósticas, el error cometido disminuye y los coeficientes pueden considerarse más confiables. El  $\rho$  y  $\lambda$  de las estimaciones con términos de autorregresivos es significativo, lo que implica que la inclusión de los retardos funciona correctamente, aunque solo en los coeficientes de la variable retardada `cuarto.porcentaje` es significativo, aunque esta no tiene un coeficiente significativo como variable sin retardo, sugiriendo que interesa su efecto sobre el área de copa en los sectores vecinos a los focos

Cuadro 4.2: Metricas de ajuste para los modelos de área de copa

medidasfit	OLS	SLX	SAR	SEM	SD
Globla Moran'I	0.09117	0.07976	-0.01715	-0.01031	-0.00695
GMI p-value	0.00313	0.00816	0.65831	0.58286	0.54450
Shapiro-Wilk	0.98196	0.99004	0.98107	0.97801	0.98845
SW p-value	0.00043	0.02653	0.00028	0.00007	0.01121
Breusch-Pagan	21.62587	34.78295	15.28709	14.33666	32.35172
BP p-value	0.00142	0.00051	0.01814	0.02609	0.00122
Media Residuos	0.00000	0.00000	0.00000	0.00000	0.00000
MSE	0.00345	0.00322	0.00330	0.00335	0.00315
adj-Rsquare	0.60479	0.99461	NA	NA	NA
Nagelkerke pseudo-R-squared	NA	NA	0.62599	0.61980	0.64283
AIC	-901.82032	-911.71201	-911.60831	-906.28644	-914.53237
Log likelihood	458.91016	469.85601	464.80415	462.14322	472.26619

de mayor concentración de viviendas tipo cuarto, pero no es importante en el resto de sectores. Como era de esperarse el porcentaje de área de espacios verdes es significativo en las variables sin retardo, pero no genera ningún efecto de derrame en la cobertura de copa de los sectores vecinos. A pesar de que los test de normalidad y heterocedasticidad no son exitosos, las gráficas diagnósticas muestran que los problemas se presentan en los valores extremos.

#### 4.1.3.4. Modelo espacial porcentaje de cobertura de área de copa

El mismo ejercicio se aplica al porcentaje de cobertura de copa.

#SAR

```
sar.cobertura.ap.wq<-lagsarlm(formula = as.formula(lm.best.cobertura.ap),
                                 data = lm.best.cobertura.ap$model,
                                 listw = W_queen,
                                 zero.policy = T,
                                 tol.solve=1.0e-30)

sar.sm<-summary(sar.cobertura.ap.wq, Nagelkerke=T)
```

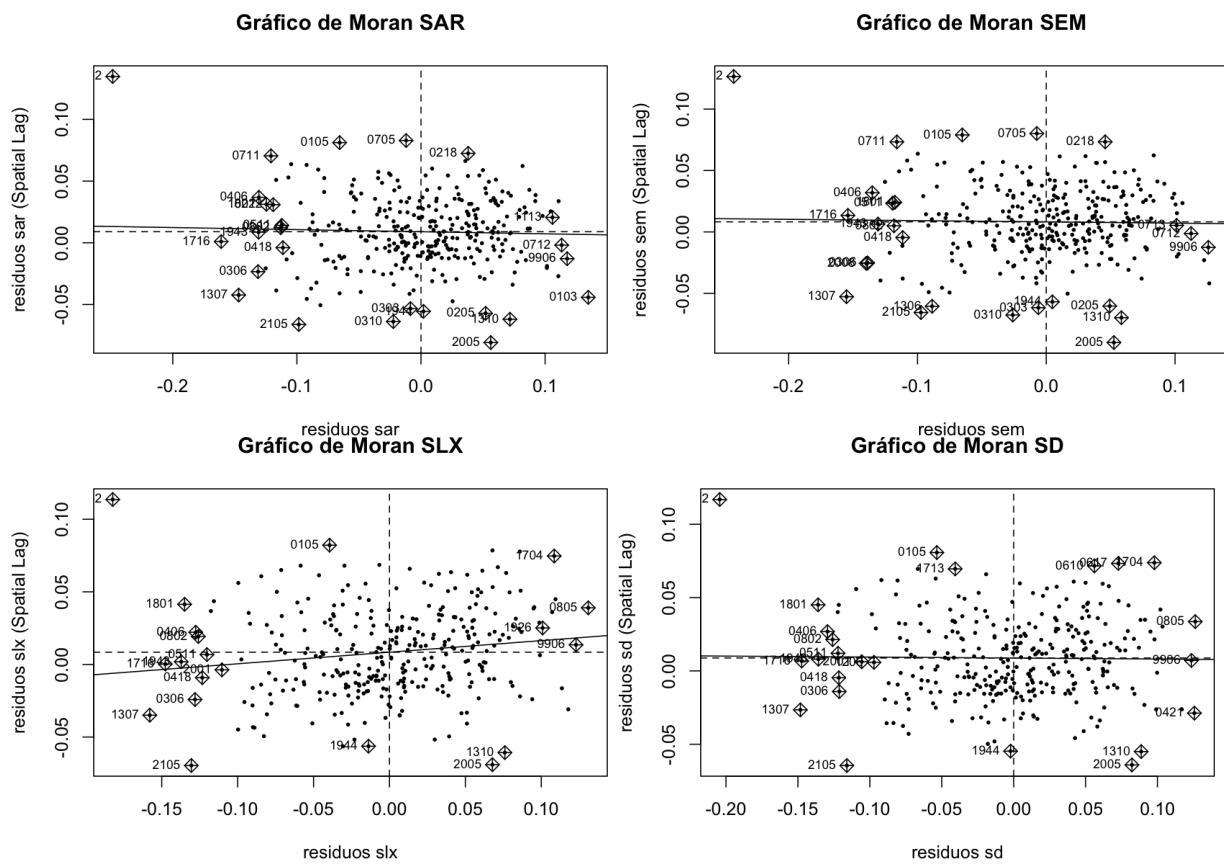


Figura 4.28: Gráfico de Moran para los residuos de los modelos espaciales de área de copa usando  $W_q$

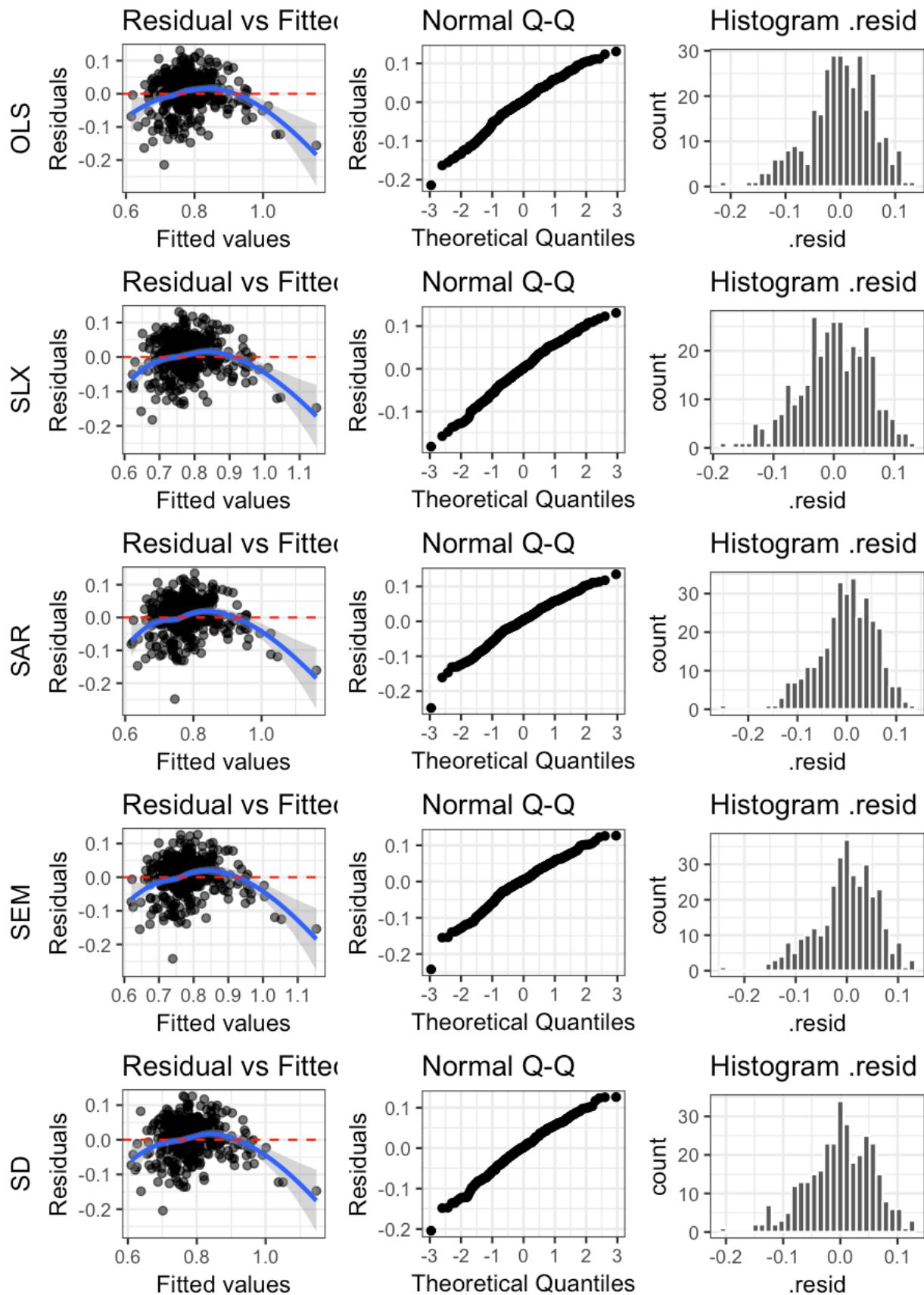


Figura 4.29: Diagnóstico comparativo entre modelos

```
sar.sm
```

```
##
```

```
## Call:
```

```
## lagsarlm(formula = as.formula(lm.best.cobertura.ap), data = lm.best.cobertura.ap$mode
```

```
##      listw = W_queen, zero.policy = T, tol.solve = 1e-30)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -0.270499 -0.049562 -0.010666  0.035328  0.624252
```

```
##
```

```
## Type: lag
```

```
## Coefficients: (asymptotic standard errors)
```

```
##                               Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)           0.030066  0.010654  2.8220  0.004773
```

```
## superior_postgrado.porcentaje.mxn 0.246343  0.027300  9.0234 < 2.2e-16
```

```
##
```

```
## Rho: 0.37521, LR test value: 28.579, p-value: 8.9958e-08
```

```
## Asymptotic standard error: 0.064408
```

```
##      z-value: 5.8255, p-value: 5.6938e-09
```

```
## Wald statistic: 33.937, p-value: 5.6938e-09
```

```
##
```

```
## Log likelihood: 285.1216 for lag model
```

```
## ML residual variance (sigma squared): 0.0097906, (sigma: 0.098947)
```

```
## Nagelkerke pseudo-R-squared: 0.50267
```

```
## Number of observations: 324
```

```
## Number of parameters estimated: 4
```

```
## AIC: -562.24, (AIC for lm: -535.66)
```

```
## LM test for residual autocorrelation
```

```
## test value: 0.19089, p-value: 0.66218
```

```

# SEM

sem.cobertura.ap.wq<-errorsarlm(formula = as.formula(lm.best.cobertura.ap),
                                    data = lm.best.cobertura.ap$model,
                                    listw = W_queen,
                                    zero.policy = T,
                                    tol.solve=1.0e-30)

sem.sm<-summary(sem.cobertura.ap.wq,
                  Nagelkerke=T)

sem.sm

## Call:
## errorsarlm(formula = as.formula(lm.best.cobertura.ap), data = lm.best.cobertura.ap$)
##   listw = W_queen, zero.policy = T, tol.solve = 1e-30)
## 
## Residuals:
##       Min        1Q     Median        3Q       Max
## -0.281664 -0.052052 -0.012380  0.034632  0.625256
## 
## Type: error
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.070011  0.013596  5.1495 2.612e-07
## superior_postgrado.porcentaje.mxn 0.334060  0.028075 11.8989 < 2.2e-16
## 
## Lambda: 0.39903, LR test value: 26.77, p-value: 2.2916e-07
## Asymptotic standard error: 0.071019
##      z-value: 5.6187, p-value: 1.9242e-08
## Wald statistic: 31.57, p-value: 1.9242e-08

```

```

## 
## Log likelihood: 284.2172 for error model
## ML residual variance (sigma squared): 0.0098055, (sigma: 0.099023)
## Nagelkerke pseudo-R-squared: 0.49988
## Number of observations: 324
## Number of parameters estimated: 4
## AIC: -560.43, (AIC for lm: -535.66)

```

#SD

```

sd.cobertura.ap.wq<-lagsarlm(formula = as.formula(lm.best.cobertura.ap),
                                 data = lm.best.cobertura.ap$model,
                                 listw = W_queen, zero.policy = T,
                                 tol.solve=1.0e-30,
                                 type = "mixed")
sd.sm<-summary(sd.cobertura.ap.wq,Nagelkerke=T)
sd.sm

```

##

## Call:

```
## lagsarlm(formula = as.formula(lm.best.cobertura.ap), data = lm.best.cobertura.ap$model)
```

```
##      listw = W_queen, type = "mixed", zero.policy = T, tol.solve = 1e-30)
```

##

## Residuals:

	Min	1Q	Median	3Q	Max
--	-----	----	--------	----	-----

##	-0.275705	-0.051349	-0.011907	0.033539	0.621785
----	-----------	-----------	-----------	----------	----------

##

## Type: mixed

## Coefficients: (asymptotic standard errors)

	Estimate	Std. Error	z value
--	----------	------------	---------

```

## (Intercept)          0.032144  0.010945  2.9370
## superior_postgrado.porcentaje.mxn    0.278081  0.044417  6.2607
## lag.superior_postgrado.porcentaje.mxn -0.051197  0.058161 -0.8803
##                                         Pr(>|z|)
## (Intercept)          0.003314
## superior_postgrado.porcentaje.mxn    3.833e-10
## lag.superior_postgrado.porcentaje.mxn  0.378714
##
## Rho: 0.40097, LR test value: 27.538, p-value: 1.5403e-07
## Asymptotic standard error: 0.070765
##      z-value: 5.6663, p-value: 1.4592e-08
## Wald statistic: 32.107, p-value: 1.4592e-08
##
## Log likelihood: 285.5239 for mixed model
## ML residual variance (sigma squared): 0.0097234, (sigma: 0.098607)
## Nagelkerke pseudo-R-squared: 0.5039
## Number of observations: 324
## Number of parameters estimated: 5
## AIC: -561.05, (AIC for lm: -535.51)
## LM test for residual autocorrelation
## test value: 1.633, p-value: 0.20129

```

**#SLX**

```

slx.cobertura.ap.wq<-lmSLX(formula = as.formula(lm.best.cobertura.ap),
                               data = lm.best.cobertura.ap$model,
                               listw = W_queen,
                               zero.policy = T)

```

```

slx.sm<-summary(slx.cobertura.ap.wq)
slx.sm

```

```

## Call:
## lm(formula = y ~ x - 1, weights = weights)

## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.29497 -0.05404 -0.01494  0.04472  0.66363 

## Coefficients:
##                               Estimate Std. Error t value
## x(Intercept)                  0.05912   0.01080  5.472
## xsuperior_postgrado.porcentaje.mxn    0.29297   0.04733  6.190
## xlag.superior_postgrado.porcentaje.mxn  0.07542   0.05570  1.354
##                               Pr(>|t|)    
## x(Intercept)                  8.99e-08 ***
## xsuperior_postgrado.porcentaje.mxn    1.84e-09 ***
## xlag.superior_postgrado.porcentaje.mxn     0.177
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Residual standard error: 0.1051 on 321 degrees of freedom
## Multiple R-squared:  0.808,  Adjusted R-squared:  0.8062
## F-statistic: 450.4 on 3 and 321 DF,  p-value: < 2.2e-16

AIC(slx.cobertura.ap.wq)

```

```
## [1] -535.5097
```

Al comparar los resultados de las métricas de ajuste se identifica al modelo SAR con el mejor rendimiento en el ajuste del modelo. El coeficiente de correlación  $\rho$  es alto y muy significativo, lo que nos dice que es una mejora la inclusión de las características

Cuadro 4.3: Metricas de ajuste para los modelos de porcentaje de área de copa

medidasfit	OLS	SLX	SAR	SEM	SD
Globla Moran'I	0.19436	0.19945	0.00734	-0.01131	-0.00863
GMI p-value	0.00000	0.00000	0.37996	0.59509	0.56441
Shapiro-Wilk	0.90492	0.90851	0.89936	0.89238	0.89685
SW p-value	0.00000	0.00000	0.00000	0.00000	0.00000
Breusch-Pagan	16.82734	16.80794	16.22458	15.21537	17.19946
BP p-value	0.00004	0.00022	0.00006	0.00010	0.00018
Media Residuos	0.00000	0.00000	0.00000	0.00000	0.00000
MSE	0.01100	0.01094	0.00979	0.00981	0.00972
adj-Rsquare	0.45512	0.80624	NA	NA	NA
Nagelkerke pseudo-R-squared	NA	NA	0.50267	0.49988	0.50390
AIC	-535.66434	-535.50971	-562.24323	-560.43438	-561.04776
Log likelihood	270.83217	271.75485	285.12162	284.21719	285.52388

espaciales en los datos, y se reflejados en el índice de Akaike que tiene una mejora visible (de -535,6643429 a -562,2432322). Consistentemente la variable de **estudios superiores** en la población refleja el patrón de agrupamiento de la cobertura de copa teniendo la mayor importancia para la estimación; pero es poco significativa como variable retardada, y es por esto que los modelos de error espacial SEM y el mixto (SD) no son mejores que el autorregresivo. Todos los modelos espaciales logran reducir la autocorrelación en los residuos (ver figura 4.30) lo que hace más confiables los coeficientes estimados, pero en el modelo SAR se mantiene una importancia en el efecto de la densidad de población como determinante para la reducción de la cobertura de copa.

## 4.2. Acceso a espacios verdes

En el caso de los EV contamos con una gran variedad de medidas sobre el acceso en relación con la distancia, el área disponible y diferentes formulaciones para aproximarse al concepto de acceso a un servicio ambiental (Figura 4.32). Para acotar el alcance de este trabajo, nos concentramos en dos métricas: el porcentaje de área de espacio verde de un sector censal (`area_ep.porcentaje`), para aproximarse desde la idea de los beneficios principalmente

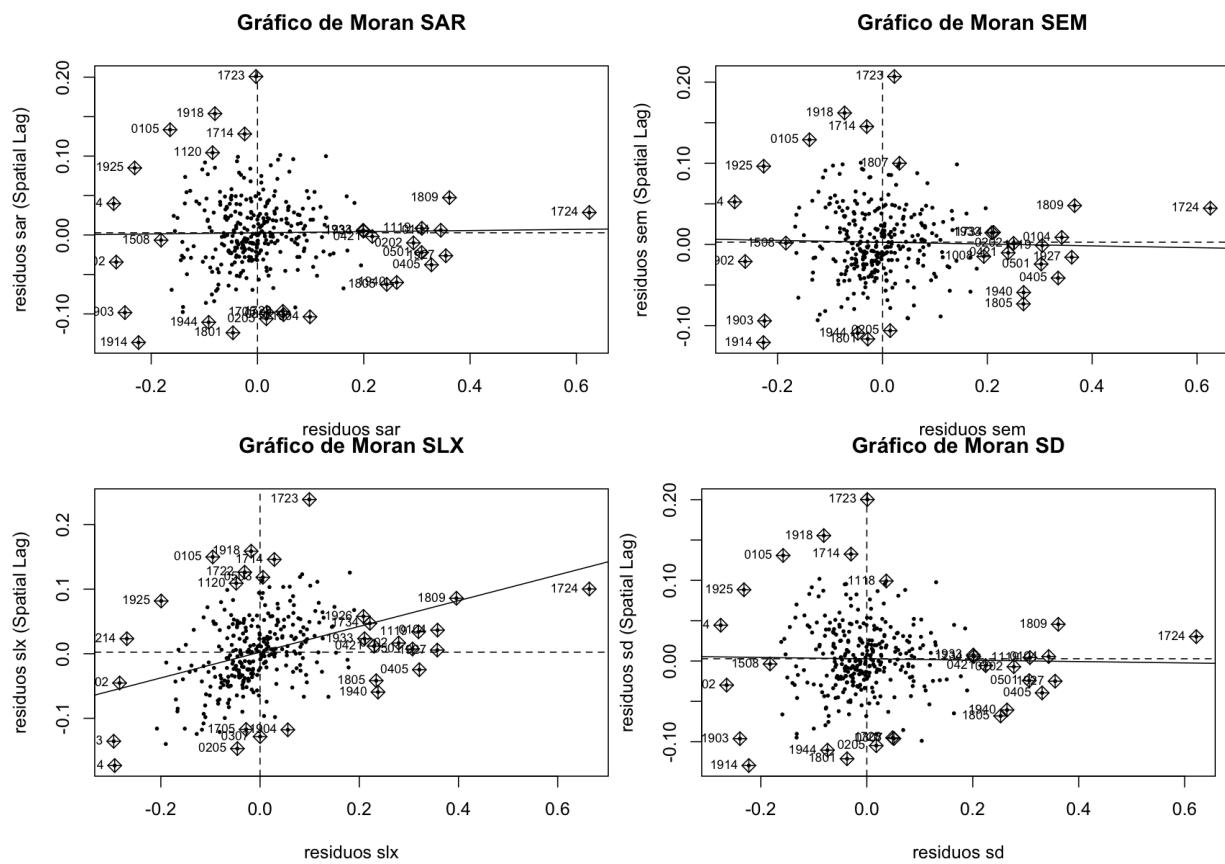


Figura 4.30: Gráfico de Moran para los residuos de los modelos espaciales del porcentaje de área de copa usando  $W_q$

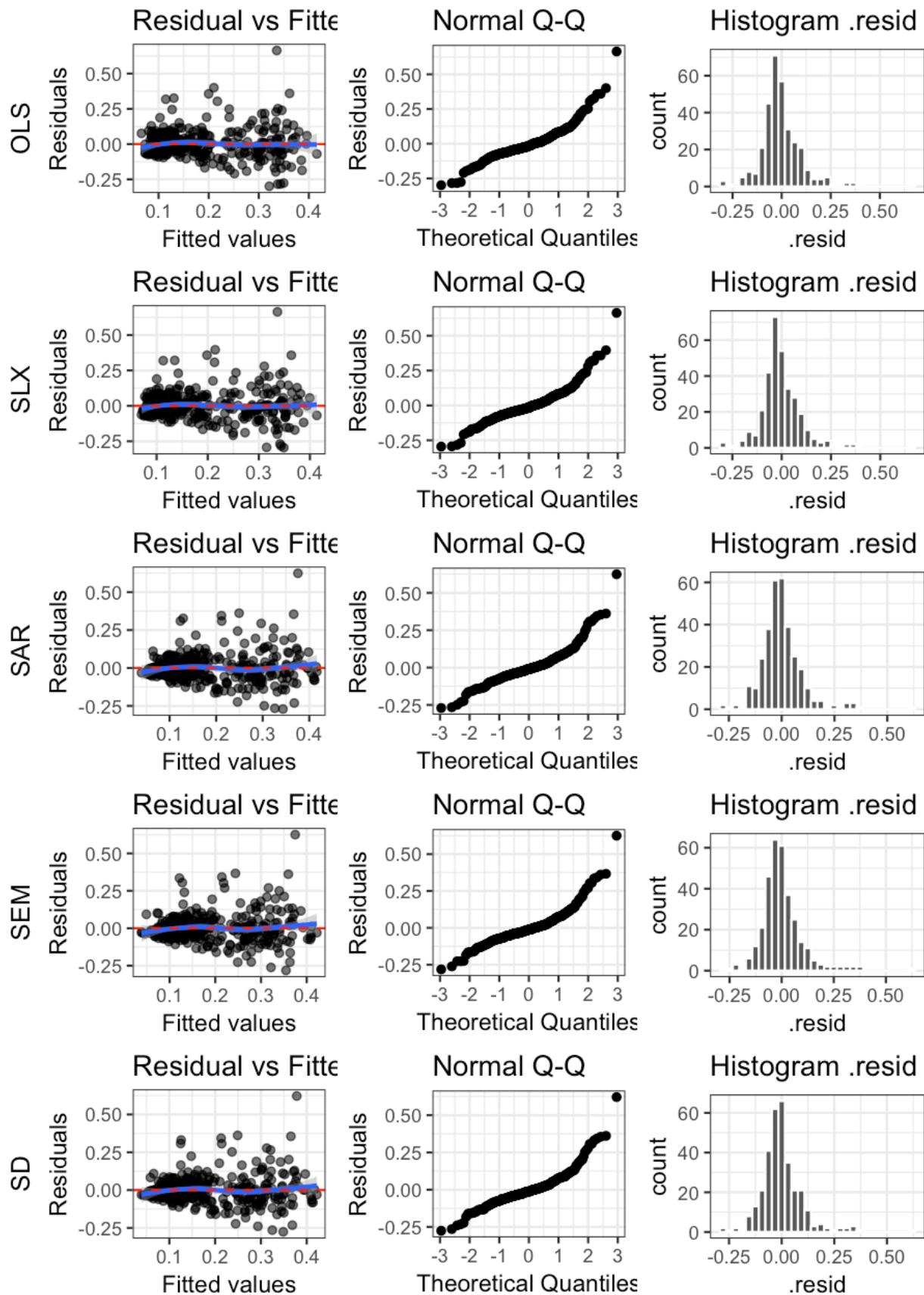


Figura 4.31: Diagnóstico comparativo entre modelos de porcentaje de copa

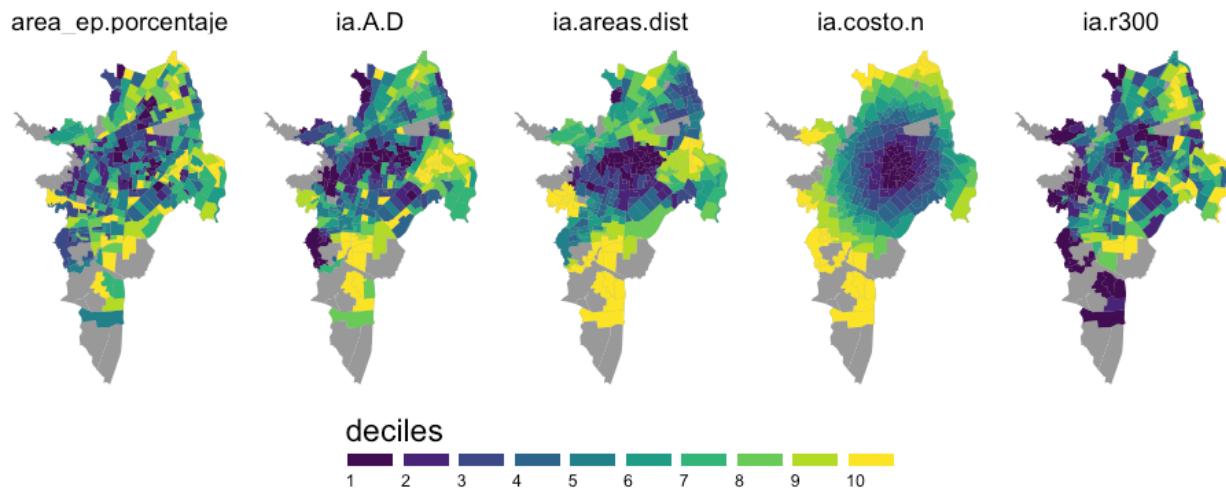


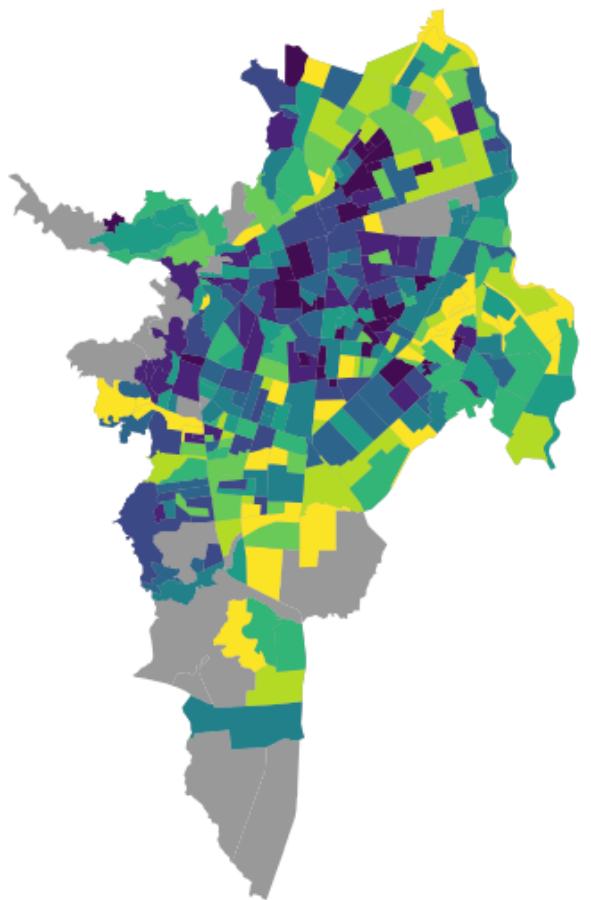
Figura 4.32: Metricas de acceso a espacio verdes

a nivel local, y la razón área disponible distancia (ia.areas.dist), que tiende a formar agrupaciones de SU al rededor de las sectore donde se ubican espacios verdes, ya sea por número o por tamaño, que contemplan el fenómeno del acceso o beneficio como un proceso acotado por la distancia escogida de 1 kilómetro como radio de búsqueda; considerándola una distancia caminable para viajar en una ciudad como Cali. (Figura 4.33)

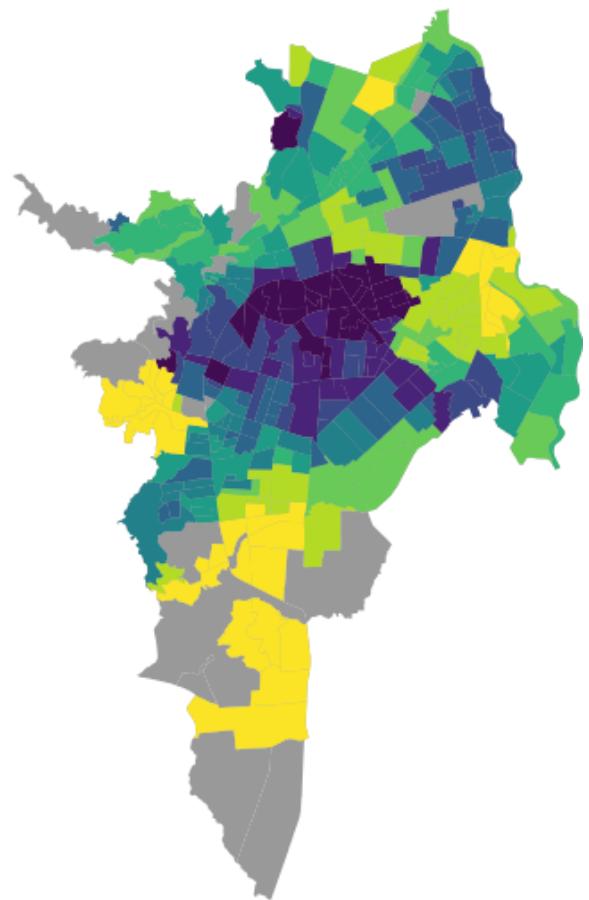
#### 4.2.1. Correlaciones y distribuciones bivariadas

Las variables de población y su correlación con los indicadores de acceso seleccionados nos sirven para seleccionar una vez más las variables independientes para usar en los modelos de regresión lineal. Las figuras 4.34 y 4.35 resumen los resultados del cálculo de los coeficientes de Pearson y Spearman respectivamente. Como se observa, esta relación es muy débil, y en todos las variables (y para ambos coeficientes de correlación) es inferior a 0.3, un valor considerado muy bajo para incluir alguna de estas variables para que tenga éxito una aproximación lineal o no lineal al predecir o ajustar valores. Sin embargo, y como parte del proceso para indagar sobre el efecto en la estimación de parámetros de los modelos geoestadísticos, seleccionaremos las variables con mayor correlación: `densidad_poblacion,con_alguna_limitacion.porcentaje` mejor relacionadas con el índice de acceso `ia.areas.dist` y `ningun_estudio.porcentaje` para

area\_ep.porcentaje



ia.areas.dist



deciles



Figura 4.33: Metricas de acceso a espacio verdes seleccionadas

## Pearson entre acceso a EV y variables de población

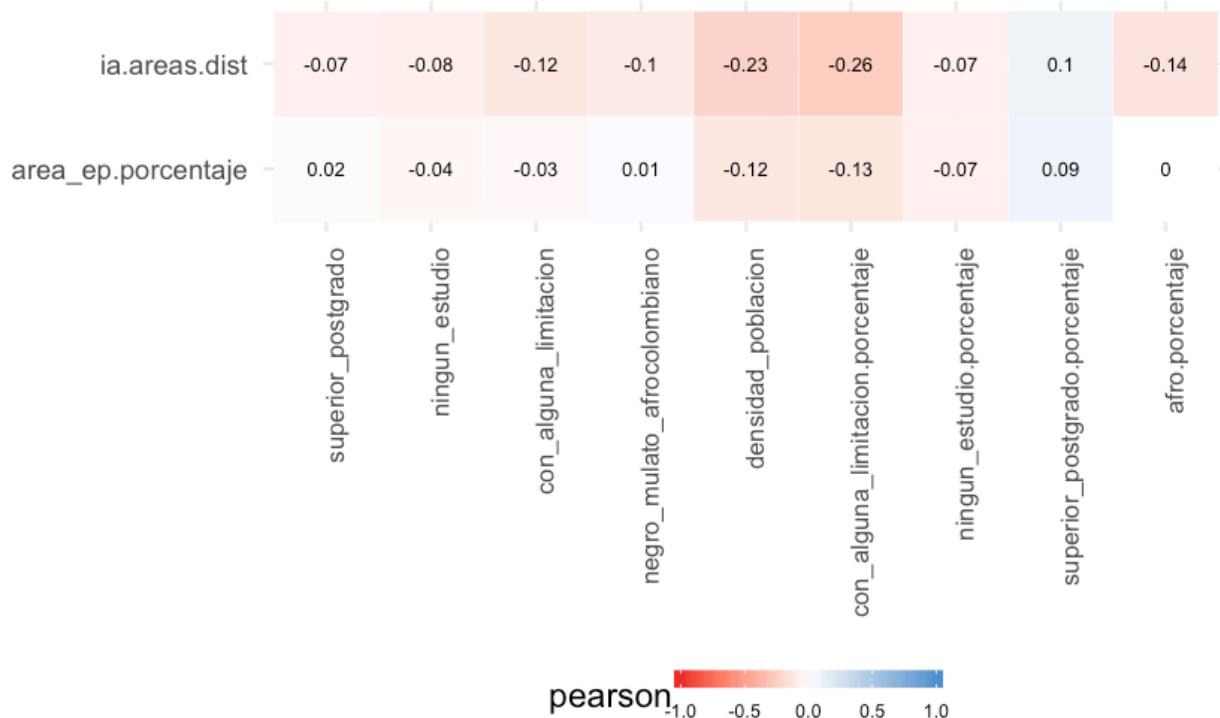


Figura 4.34: Coeficiente Pearson entre acceso a EV y variables de población

`area_ep.porcentaje`.

Resulta interesante ver, al igual que en el modelado de la cobertura de copa, si otras variables no poblacionales explican los resultados de los índices de acceso seleccionados. El conjunto de variables sobre el uso de los predios y sus coeficientes de correlación se muestran en las figuras 4.36 y 4.37. De nuevo las correlaciones son bajas, aparentemente poco explicativas de los índices de acceso. Las variables de uso de los predios que mejor se relacionan con los índices son: `unidad_economica.porcentaje` y el `cuarto.porcentaje`.

El último bloque de variables indaga sobre las áreas y proporciones de las manzanas de cada sector censal y la vocación como pública o privada de los espacios dentro de un sector urbano. La figura 4.38 y 4.39 muestran que el área media de las manzanas (`area_media_manzana`) de los sectores urbanos se relaciona de forma positiva con ambos índices de acceso, mucho más fuertemente que las variables exploradas hasta el momento.

## Spearman entre acceso a EV y variables de población

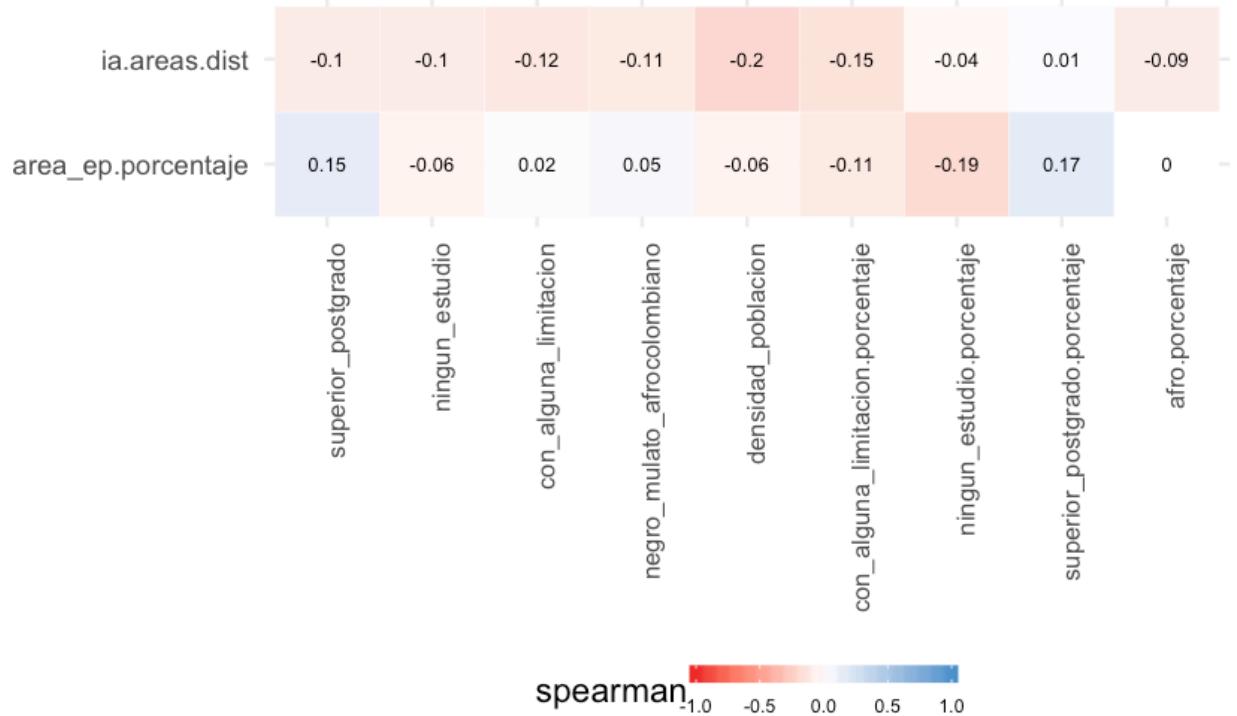


Figura 4.35: Coeficiente Spearman entre acceso a EV y variables de población

## Coeficiente Pearson entre dependiente EV e independientes uso de predios

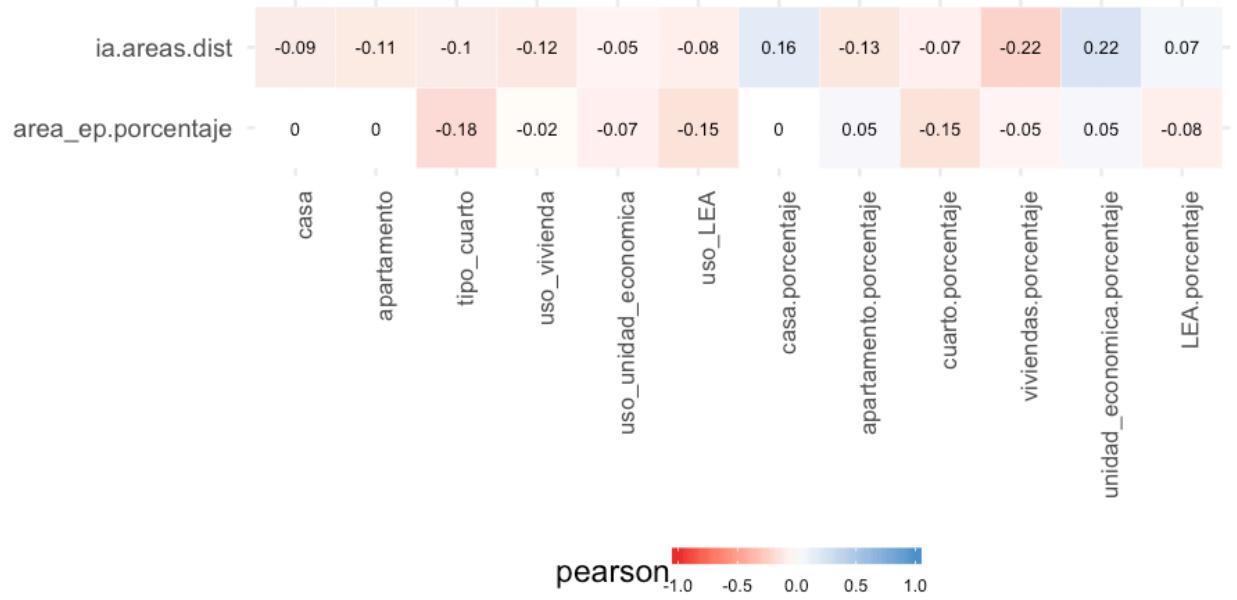


Figura 4.36: Coeficiente Pearson entre acceso a EV y variables de uso de los predios

## Coeficiente Spearman entre dependiente EV e independientes uso de predios

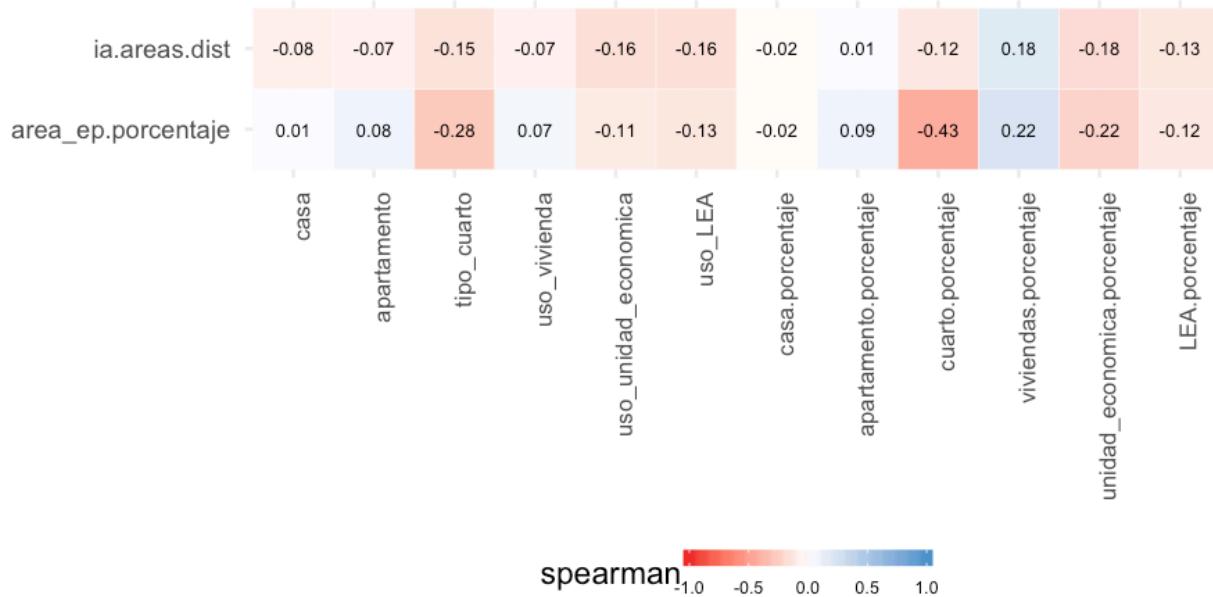


Figura 4.37: Coeficiente Spearman entre acceso a EV y variables de uso de los predios

Aunque parece haber una fuerte correlación de los indicadores de acceso con las áreas privadas, públicas y del sector urbano, estas hacen parte de los cálculos que generan estos índices, produciendo en efecto ficticio en la correlación, razón por la que no haremos uso de ellas en la modelación.

### 4.2.2. Modelos de regresión lineal EV

Para los espacios verdes usaremos todos los términos seleccionados con los coeficientes de correlación para luego ver la significancia de las variables en el modelo y elegir el modelo con mejor ajuste usando criterio de información de Akaike (AIC) seleccionando los coeficientes significativos. Para estos índices de acceso no usaremos variantes transformadas de la variable dependiente, sólo se aplica una normalización a los datos.

Para el índice de acceso de porcentaje de área de espacios verdes en un sector urbano el modelo a ajusta el siguiente:

### Coeficiente Pearson entre dependientes EV e independientes físicas de predios y SU

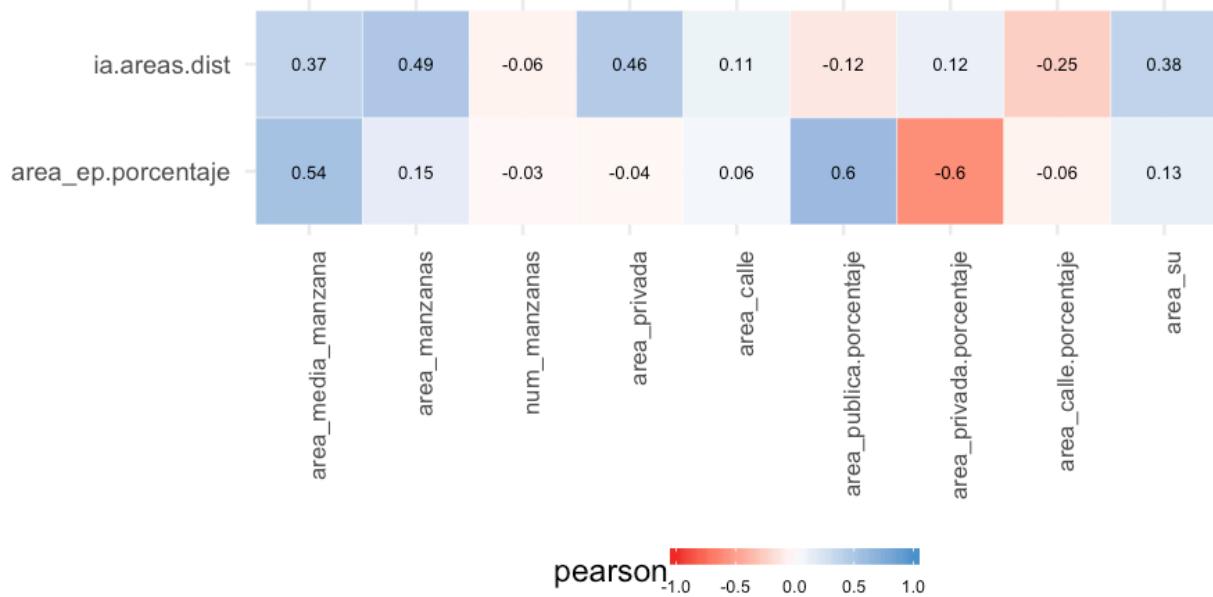


Figura 4.38: Coeficiente Pearson entre acceso a EV y variables sobre aspectos físicos de las manzanas y SU

### Coeficiente Spearman entre dependientes EV e independientes físicas de predios y SU

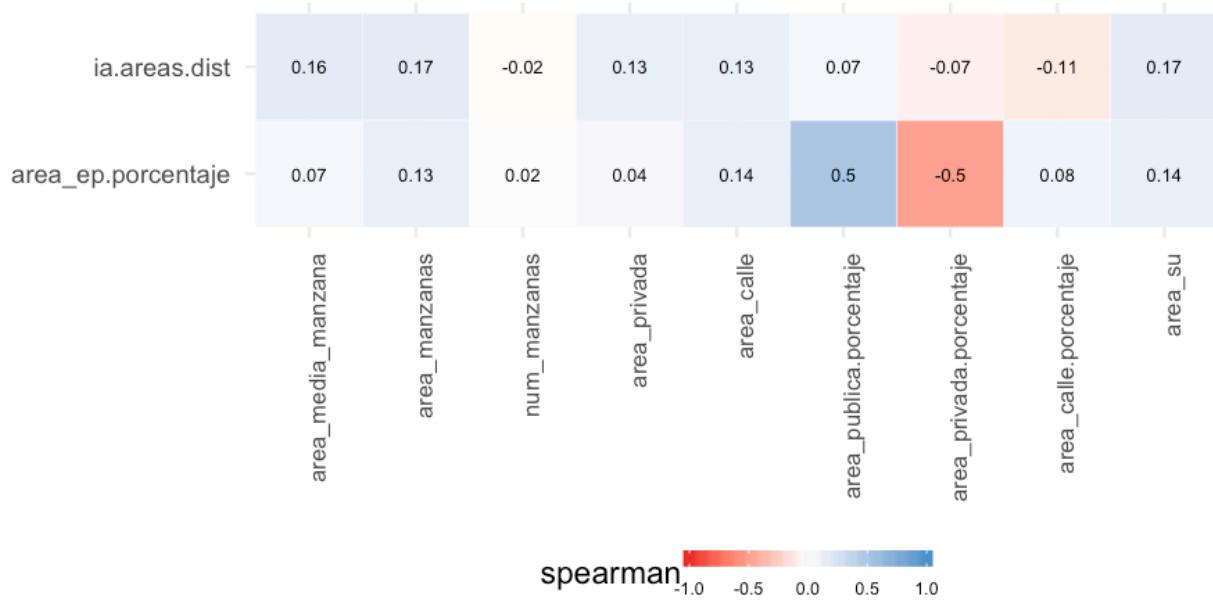


Figura 4.39: Coeficiente Spearman entre acceso a EV y variables sobre aspectos físicos de las manzanas y SU

```

dependiente <- "area_ep.porcentaje"
independientes <- independientes.EV
# max normalizado
var_names<-c(dependiente,names(regresion.EV[,independientes]))
regresion.EV.mn<-max_nomalization(regresion.EV,var_names)
lm.area_ep.ptje<-crear_lm_from_df(regresion.EV.mn)
summary(lm.area_ep.ptje)

##
## Call:
## lm(formula = form, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.26917 -0.04344 -0.01571  0.02577  0.69572
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.038434  0.016805   2.287  0.02284
## area_media_manzana.mxn      0.760522  0.070424  10.799 < 2e-16
## cuarto.porcentaje.mxn     -0.126650  0.048389  -2.617  0.00928
## unidad_economica.porcentaje.mxn -0.000400  0.033502  -0.012  0.99048
## densidad_poblacion.mxn      0.001654  0.026925   0.061  0.95105
## con_alguna_limitacion.porcentaje.mxn 0.002491  0.032620   0.076  0.93918
## ningun_estudio.porcentaje.mxn    0.055332  0.043902   1.260  0.20845
##
## (Intercept) *
## area_media_manzana.mxn ***
## cuarto.porcentaje.mxn **

```

```

## unidad_economica.porcentaje.mxn
## densidad_poblacion.mxn
## con_alguna_limitacion.porcentaje.mxn
## ningun_estudio.porcentaje.mxn
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07843 on 322 degrees of freedom
## Multiple R-squared: 0.3062, Adjusted R-squared: 0.2933
## F-statistic: 23.69 on 6 and 322 DF, p-value: < 2.2e-16

```

Los resultados muestran que las variables cuarto.porcentaje area\_media\_manzana son significativas.

```

##
## Call:
## lm(formula = area_ep.porcentaje.mxn ~ cuarto.porcentaje.mxn +
##      area_media_manzana.mxn, data = regresion.EV.mn)
##
## Residuals:
##    Min     1Q   Median     3Q    Max
## -0.26869 -0.04398 -0.01599  0.02522  0.69949
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)             0.049241  0.005703  8.635 2.66e-16 ***
## cuarto.porcentaje.mxn -0.093855  0.037643 -2.493  0.0132 *
## area_media_manzana.mxn  0.744371  0.065197 11.417 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

area\_ep.porcentaje.mx ~ cuarto.porcentaje.mx + area\_media\_manzana.mx

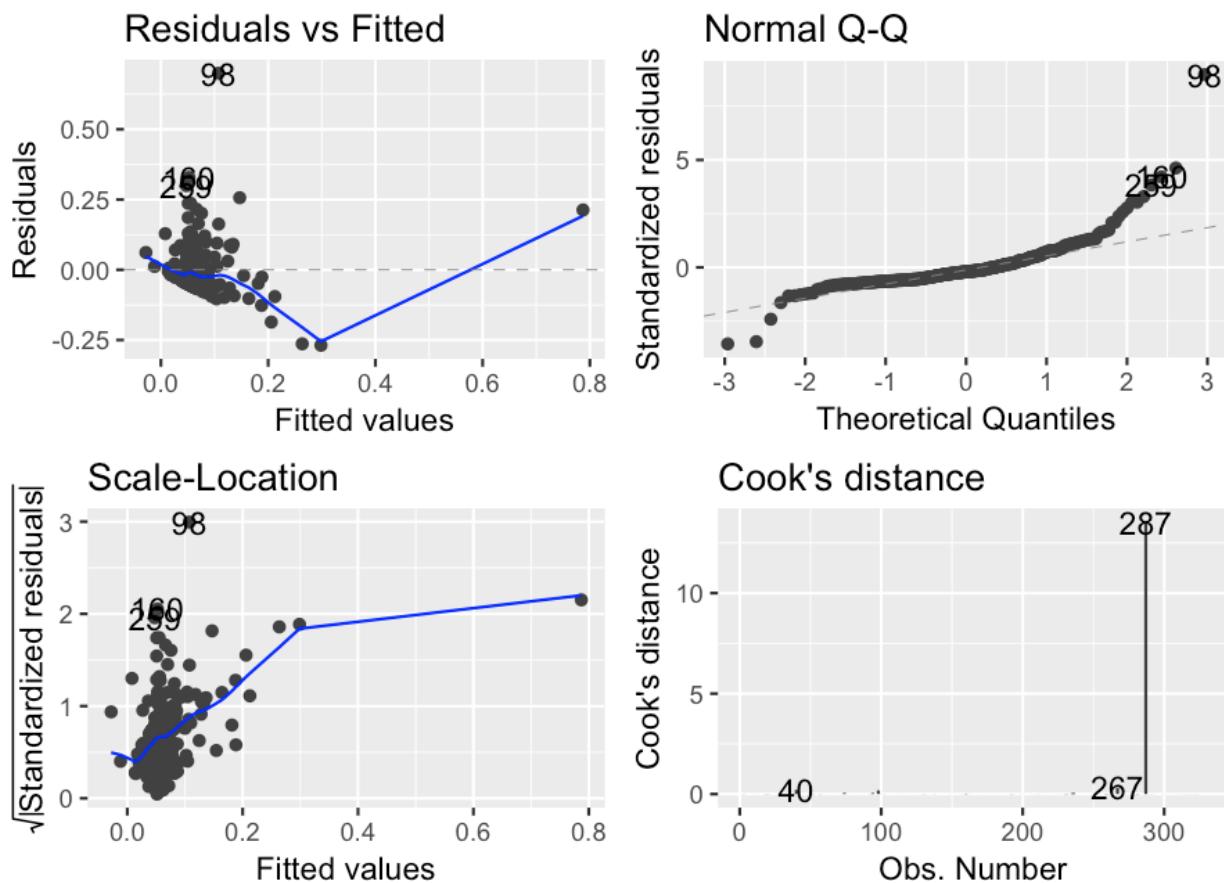


Figura 4.40: Gráficas diagnósticas para el análisis de regresión lineal de porcentaje de área de espacio verde

```
##
## Residual standard error: 0.07825 on 326 degrees of freedom
## Multiple R-squared:  0.3008, Adjusted R-squared:  0.2966
## F-statistic: 70.14 on 2 and 326 DF,  p-value: < 2.2e-16
```

Los gráficos 4.40 muestran los resultados del ajuste en relación con los residuos y 4.41 espacialmente.

Para el índice de acceso `ia.areas.dist` el modelo con todos los términos es el siguiente:

```
dependiente <- "ia.areas.dist"
independientes <- independientes.EV
```

`area_ep.porcentaje.mxn ~ cuarto.porcentaje.mxn + area_media_manzana.mxn`

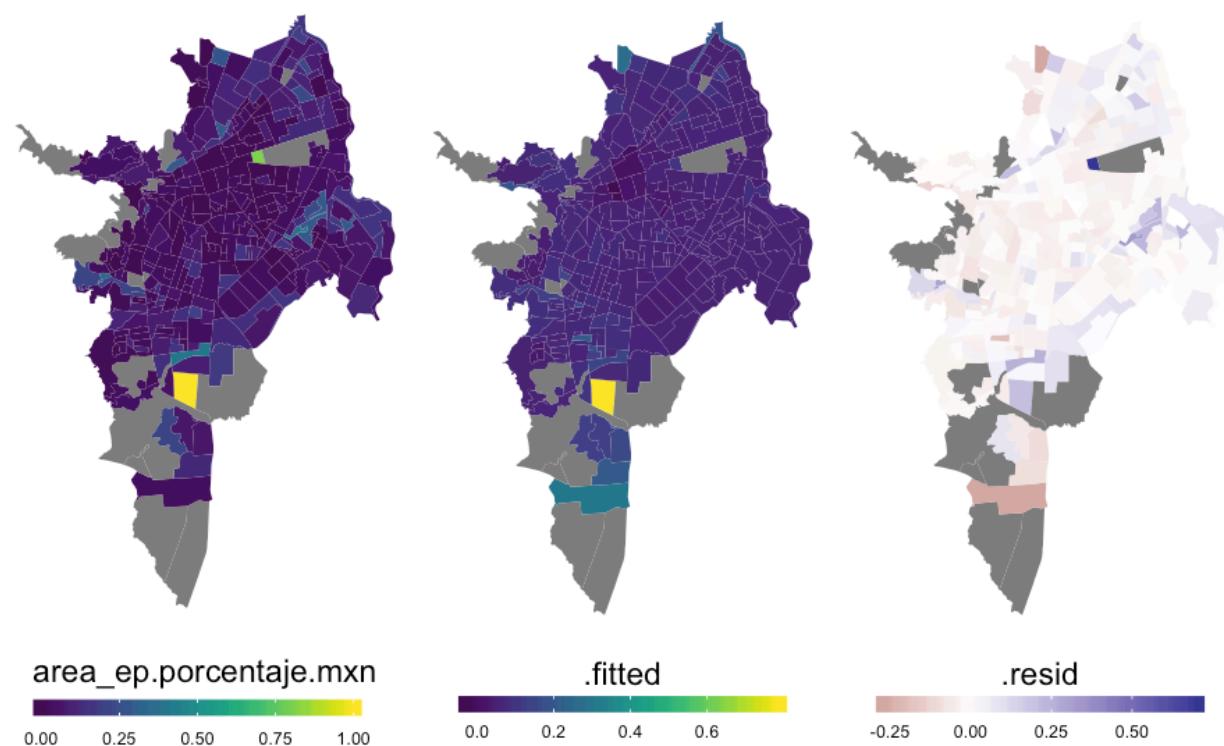


Figura 4.41: Variable dependiente, valor ajustado por el modelo y residuos del OLS para ‘area ep.porcentaje’ normalizada

```

# max normalizado

var_names<-c(dependiente,names(regresion.EV[,independientes]))

regresion.EV.mn<-max_nomalization(regresion.EV,var_names)

lm.ia.areas.dist<-crear_lm_from_df(regresion.EV.mn)

summary(lm.ia.areas.dist)

##

## Call:

## lm(formula = form, data = df)

##

## Residuals:

##      Min       1Q   Median       3Q      Max
## -0.29642 -0.03451 -0.01574  0.00990  0.82343

##

## Coefficients:

##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)               0.09203   0.02050  4.489 9.97e-06
## area_media_manzana.mxn    0.43288   0.08591  5.039 7.82e-07
## cuarto.porcentaje.mxn   -0.15932   0.05903 -2.699  0.00732
## unidad_economica.porcentaje.mxn  0.11736   0.04087  2.872  0.00435
## densidad_poblacion.mxn   -0.04920   0.03285 -1.498  0.13511
## con_alguna_limitacion.porcentaje.mxn -0.13427   0.03979 -3.374  0.00083
## ningun_estudio.porcentaje.mxn     0.14126   0.05355  2.638  0.00875
## 

## (Intercept) ***
## area_media_manzana.mxn ***
## cuarto.porcentaje.mxn **
## unidad_economica.porcentaje.mxn **
## densidad_poblacion.mxn

```

```

## con_alguna_limitacion.porcentaje.mxn ***
## ningun_estudio.porcentaje.mxn      **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09567 on 322 degrees of freedom
## Multiple R-squared:  0.2074, Adjusted R-squared:  0.1926
## F-statistic: 14.04 on 6 and 322 DF,  p-value: 3.342e-14

```

Solo la densidad de población no es significativa. Consideraremos el modelo sin simplificar para los ajustes geoestadístico. Los gráficos 4.42 muestran los resultados del ajuste del modelo con todas las variables en relación con los residuos y 4.43 espacialmente.

```

##
## Call:
## lm(formula = ia.areas.dist.mxn ~ cuarto.porcentaje.mxn + unidad_economica.porcentaje.mxn +
##     area_media_manzana.mxn + ningun_estudio.porcentaje.mxn +
##     con_alguna_limitacion.porcentaje.mxn, data = regresion.EV.mn)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31569 -0.03598 -0.01588  0.00771  0.82467
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.08111   0.01920   4.225 3.11e-05
## cuarto.porcentaje.mxn      -0.14696   0.05856  -2.510 0.012578
## unidad_economica.porcentaje.mxn  0.13706   0.03877   3.535 0.000467
## area_media_manzana.mxn        0.45626   0.08464   5.391 1.36e-07
## ningun_estudio.porcentaje.mxn  0.11501   0.05070   2.268 0.023970

```

Cuadro 4.4: Resumen métricas de ajuste OLS Indice contenedor y de acceso área-distancia

medidasfit	%Area EV	Area-Distancia
Shapiro-Wilk	0.76782	0.57265
SW p-value	0.00000	0.00000
Breusch-Pagan	12.98572	57.30864
BP p-value	0.00151	0.00000
Media Residuos	0.00000	0.00000
MSE	0.00607	0.00896
adj-Rsquare	0.29656	0.19261
AIC	-737.84913	-601.61394
Log likelihood	372.92456	308.80697

```

## con_alguna_limitacion.porcentaje.mxn -0.14154      0.03957  -3.577  0.000401
##
## (Intercept) ***
## cuarto.porcentaje.mxn *
## unidad_economica.porcentaje.mxn ***
## area_media_manzana.mxn ***
## ningun_estudio.porcentaje.mxn *
## con_alguna_limitacion.porcentaje.mxn ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09586 on 323 degrees of freedom
## Multiple R-squared:  0.2019, Adjusted R-squared:  0.1895
## F-statistic: 16.34 on 5 and 323 DF,  p-value: 2.254e-14

```

La tabla 4.4 resume las métricas de ajuste de ambos modelos de acceso a EV.

$\text{ia.areas.dist.mxn} \sim \text{cuarto.porcentaje.mxn} + \text{unidad_economica.porcentaje.mxn} +$   
 $\text{area_media_manzana.mxn} + \text{ningun_estudio.porcentaje.mxn} +$   
 $\text{con_alguna_limitacion.porcentaje.mxn}$

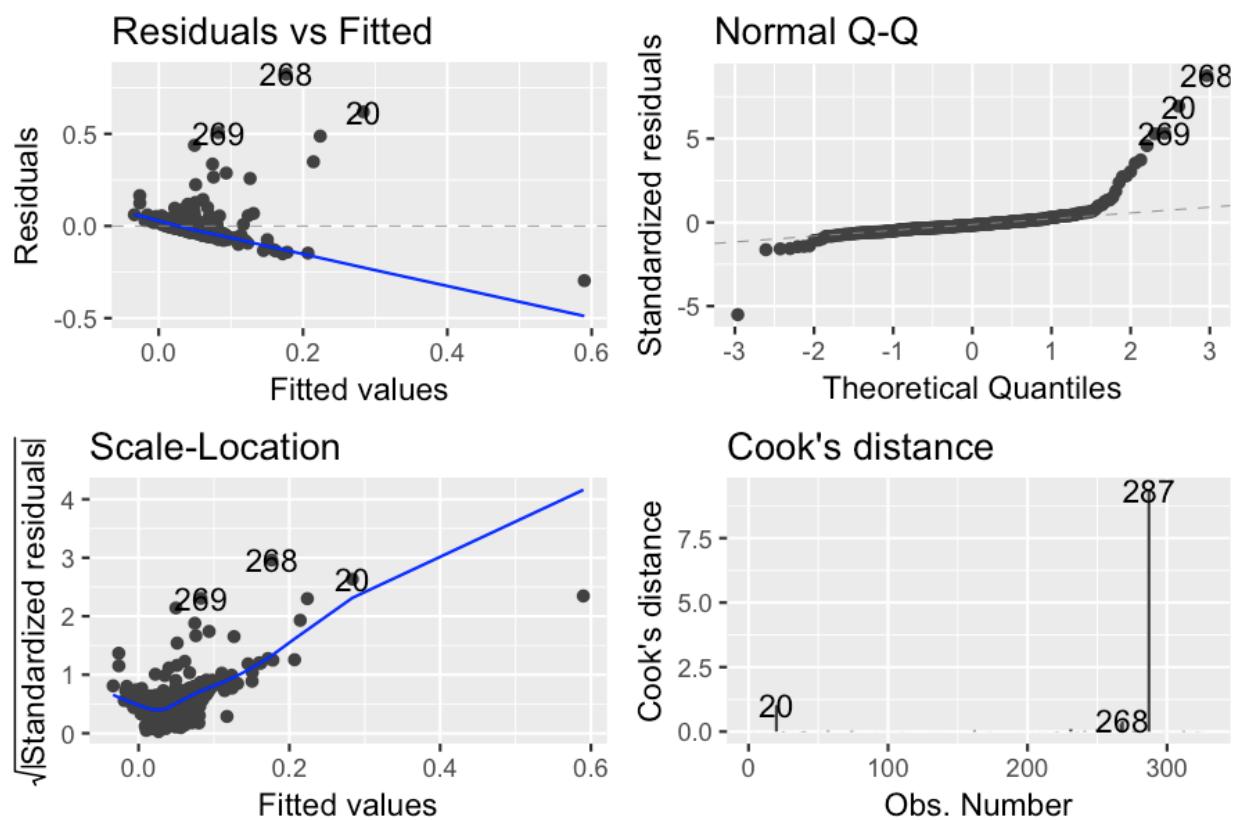


Figura 4.42: Gráficas diagnósticas para el análisis de regresión lineal de 'ia.areas.dist'

```
ia.Areas.dist.mxn ~ area_media_manzana.mxn + cuarto.porcentaje.mxn +  
    unidad_economica.porcentaje.mxn + densidad_poblacion.mxn +  
    con_alguna_limitacion.porcentaje.mxn + ningun_estudio.porcentaje.mxn
```

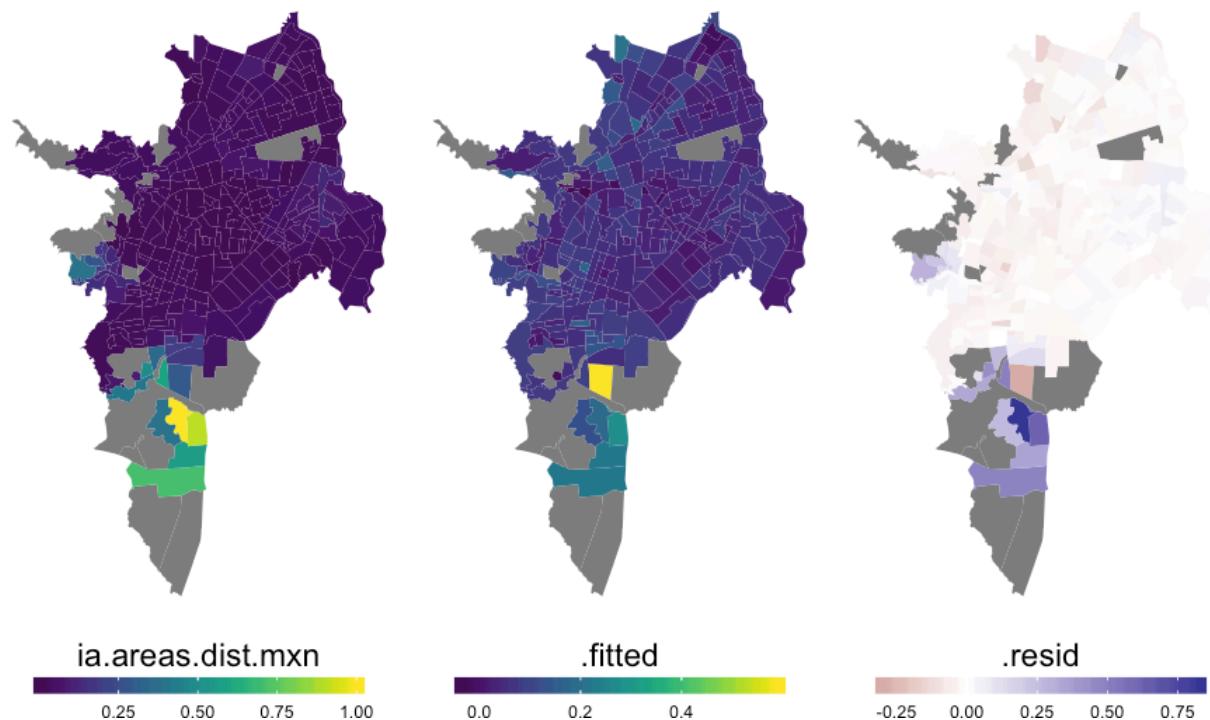


Figura 4.43: Variable dependiente, valor ajustado por el modelo y residuos del OLS para 'ia.Areas.dist' normalizada

### 4.2.3. Modelado espacial de espacios verdes

El proceso de ajuste de los modelos geoestadísticos para el análisis de espacios verdes hace uso de los mismos elementos metodológicos usados para la cobertura de copa. Se construyen dos matrices de velocidad usando un kernel de vecindad Queen  $W_q$  y otro con base en un radio de búsqueda de 1 kilómetro  $W_d$ . El primer paso es evaluar cuál de las matrices captura mejor la autocorrelación de los residuos de los modelos lineales y de las variables dependientes. Seguidamente se compara los diferentes modelos espaciales para seleccionar el de mejor ajuste y finalmente se evalúa la significancia de las variables y el valor de los coeficientes de la regresión y la mejora en el ajuste con relación a los modelos lineales.

#### 4.2.3.1. Matrices de vecindad

Las dos matrices de vecindad construidas con los SU seleccionados para el análisis de regresión lineal. Las dos matrices resultantes se muestran en la figura 4.44

#### 4.2.3.2. Autocorrelación espacial

##### 4.2.3.2.1. Variables dependientes

Comparemos el efecto de cada una de las matrices sobre el indicador `area_ep.porcentaje`

```
moran.test(regresion.EV$area_ep.porcentaje,
            listw=W_dist1000.inv.ev,
            zero.policy = TRUE)
```

```
##
## Moran I test under randomisation
##
## data: regresion.EV$area_ep.porcentaje
```

**Queen**



**Distancia inversa a 1km**

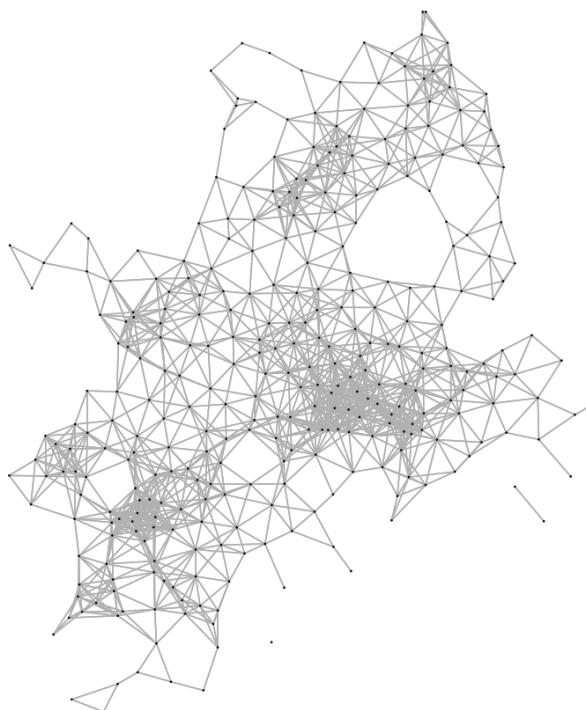


Figura 4.44: Matrices de vecindad del análisis espacial de espacios verdes

```

## weights: W_dist1000.inv.ev
##
## Moran I statistic standard deviate = 2.0602, p-value = 0.01969
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.0537662589     -0.0030769231    0.0007612826

moran.test(regresion.EV$area_ep.porcentaje,
            listw=W_queen.ev)

##
## Moran I test under randomisation
##
## data: regresion.EV$area_ep.porcentaje
## weights: W_queen.ev
##
## Moran I statistic standard deviate = 3.3938, p-value = 0.0003447
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.104622145     -0.003048780    0.001006547

moran.test(regresion.EV$ia.areas.dist,
            listw=W_dist1000.inv.ev,
            zero.policy = TRUE)

##
## Moran I test under randomisation
##
## data: regresion.EV$ia.areas.dist

```

```

## weights: W_dist1000.inv.ev
##
## Moran I statistic standard deviate = 24.519, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.6777872466    -0.0030769231     0.0007711172

moran.test(regresion.EV$ia.areas.dist,
            listw=W_queen.ev)

##
## Moran I test under randomisation
##
## data: regresion.EV$ia.areas.dist
## weights: W_queen.ev
##
## Moran I statistic standard deviate = 24.624, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.783140286    -0.003048780     0.001019399

```

La matriz  $W_q$  captura con mayor fuerza la autocorrelación espacial de `area_ep.porcentaje`. Al repetir el test para `ai.areas.dist` una vez más  $W_q$  captura con mayor intensidad la autocorrelación espacial del indicador, que es mucho mayor que en `area_ep.porcentaje`, posiblemente por representar una característica local del acceso. Hay que anotar aquí que el cálculo del índice `ia.areas.dist` en su construcción usa una distancia de radio de búsqueda de 1 kilómetro; en su definición el indicador está influenciado por sus vecinos por lo que se forman grupos o clusters alrededor de ciertos sectores urbanos. Resulta pues interesante no sea  $W_d$  la que capture mejor el agrupamiento.

Para indagar sobre los patrones espaciales de los dos indicadores usando  $W_q$  se muestran los mapas de LISA en la figura 4.45. Se aprecia que se forman cluster alrededor de tres zonas en el caso del porcentaje de área de espacio verde y dos para el indicador de la relación areas-distancia, coincidentes con el anterior. Ahí se encuentran equipamientos de ciudad como un cementerio de gran tamaño, las universidades, zonas conservadas de riberas de ríos. El grupo que se forma al oriente de la ciudad es donde se encuentra la laguna del Pondaje.

#### 4.2.3.2.2. Residuos de los OLS

Examinemos ahora los residuos de cada modelo lineal seleccionado. Aplicamos el test de Moran primero al modelo de `area_ep.porcentaje` con ambas matrices de vecindad. La figura 4.46 muestra el mismo resultado gráficamente.

```
lm.morantest(lm.area_ep.ptje.sel,
              W_queen.ev,
              alternative="two.sided",
              zero.policy = T)

##
## Global Moran I for regression residuals
##
## data:
## model: lm(formula = area_ep.porcentaje.mxn ~ cuarto.porcentaje.mxn
## + area_media_manzana.mxn, data = regresion.EV.mn)
## weights: W_queen.ev
##
## Moran I statistic standard deviate = 3.5892, p-value = 0.0003316
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
```

Mapas LISA - area\_ep.porcentaje -  $W_q$

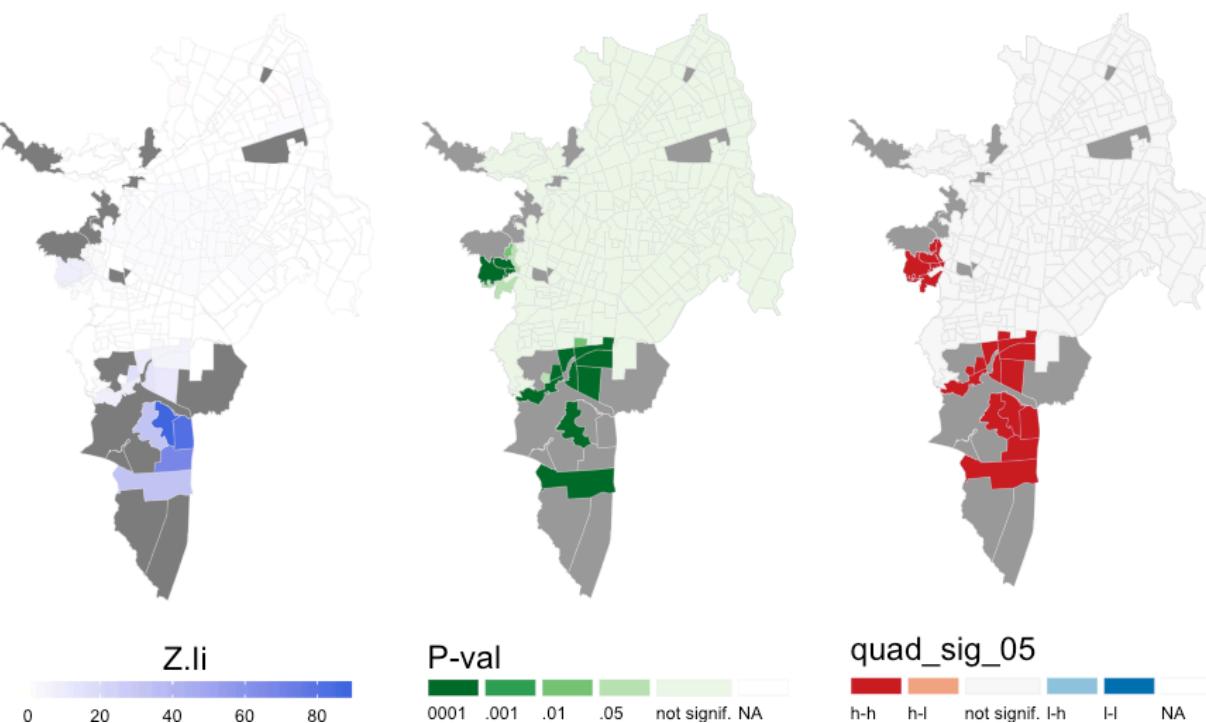
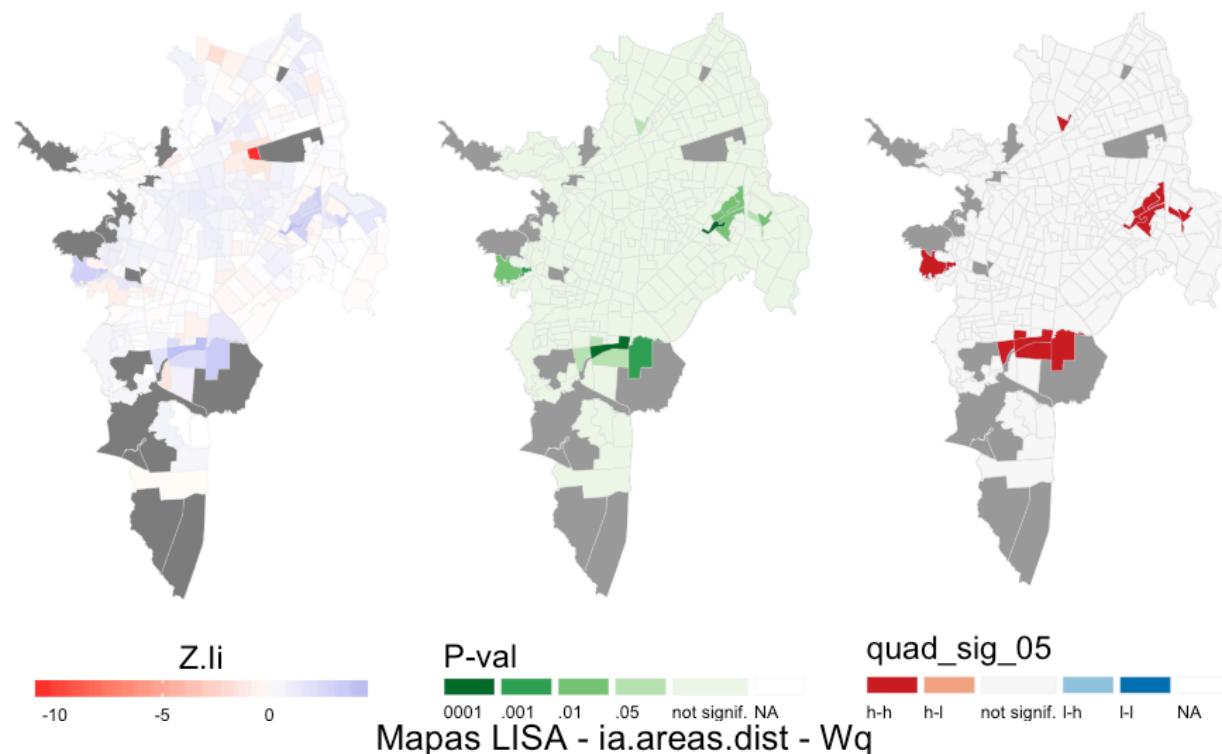


Figura 4.45: Mapas LISA para la matriz  $W_q$  de ambos indicadores de acceso a EV

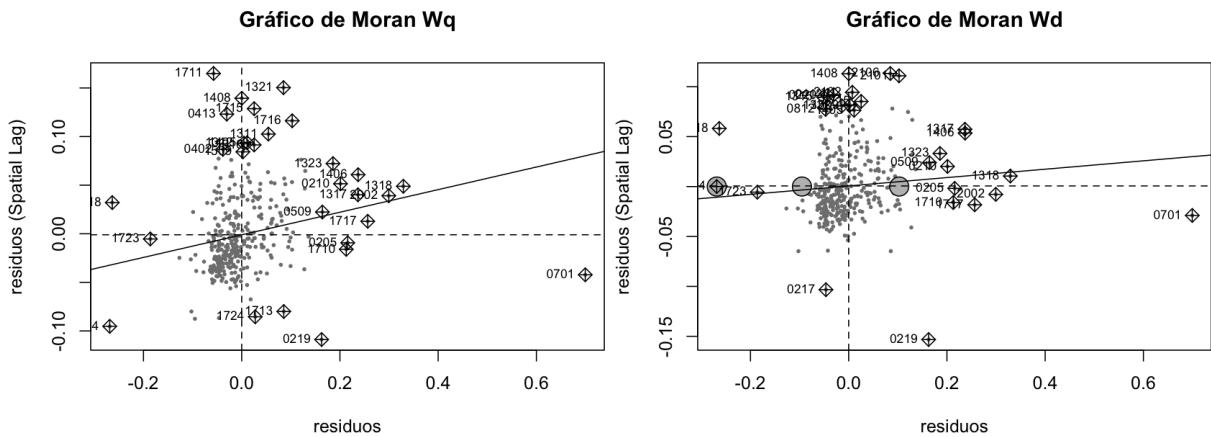


Figura 4.46: Gráfico de Moran para los residuos del modelo lineal para el porcentaje de área de espacio verde

```
##          0.116355498      -0.005017782       0.001143508
```

```
lm.morantest(lm.area_ep.ptje.sel,
              W_dist1000.inv.ev,
              alternative="two.sided",
              zero.policy = T)

##
## Global Moran I for regression residuals
##
## data:
## model: lm(formula = area_ep.porcentaje.mxn ~ cuarto.porcentaje.mxn
## + area_media_manzana.mxn, data = regresion.EV.mn)
## weights: W_dist1000.inv.ev
##
## Moran I statistic standard deviate = 1.579, p-value = 0.1143
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
##          0.0420095663     -0.0044075679     0.0008641599
```

Al igual que con la variable dependiente,  $W_q$  captura con mayor intensidad autocorrelación en los residuos. Los resultados de aplicar el test a `ia.areas.dist` confirman que la matriz de vecindad que usaremos para el ajuste es  $W_q$ .

```

lm.morantest(lm.ia.areas.dist.sel,
              W_queen.ev,
              alternative="two.sided",
              zero.policy = T)

## Global Moran I for regression residuals
## data:
## model: lm(formula = form, data = df)
## weights: W_queen.ev
## Moran I statistic standard deviate = 17.805, p-value < 2.2e-16
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
##      0.585432955     -0.010137536     0.001118915

lm.morantest(lm.ia.areas.dist.sel,
              W_dist1000.inv.ev,
              alternative="two.sided",
              zero.policy = T)

## Global Moran I for regression residuals
## data:

```

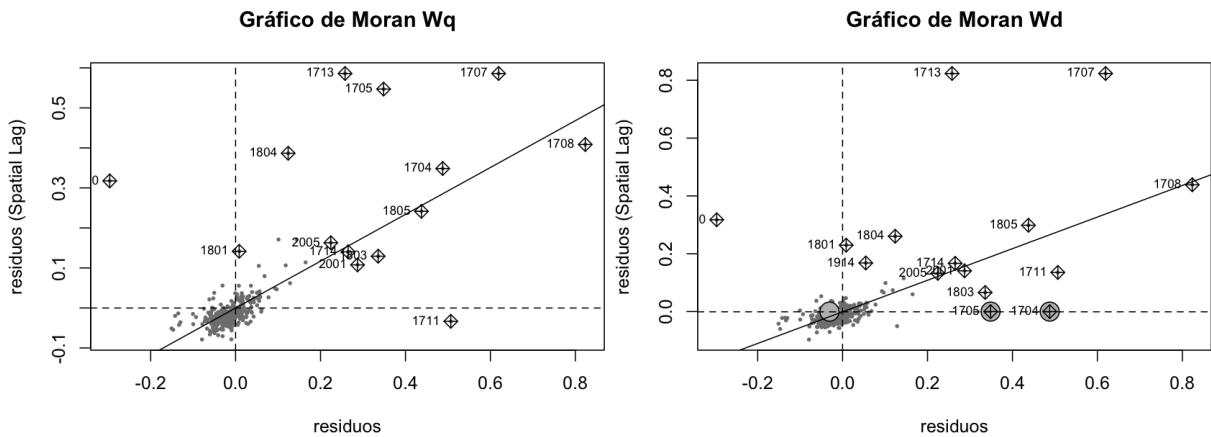


Figura 4.47: Gráfico de Moran para los residuos del modelo lineal para el indicador ‘ia.areas.dist’

```

## model: lm(formula = form, data = df)
## weights: W_dist1000.inv.ev
##
## Moran I statistic standard deviate = 19.219, p-value < 2.2e-16
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
##      0.5466198707    -0.0088593824   0.0008353882

```

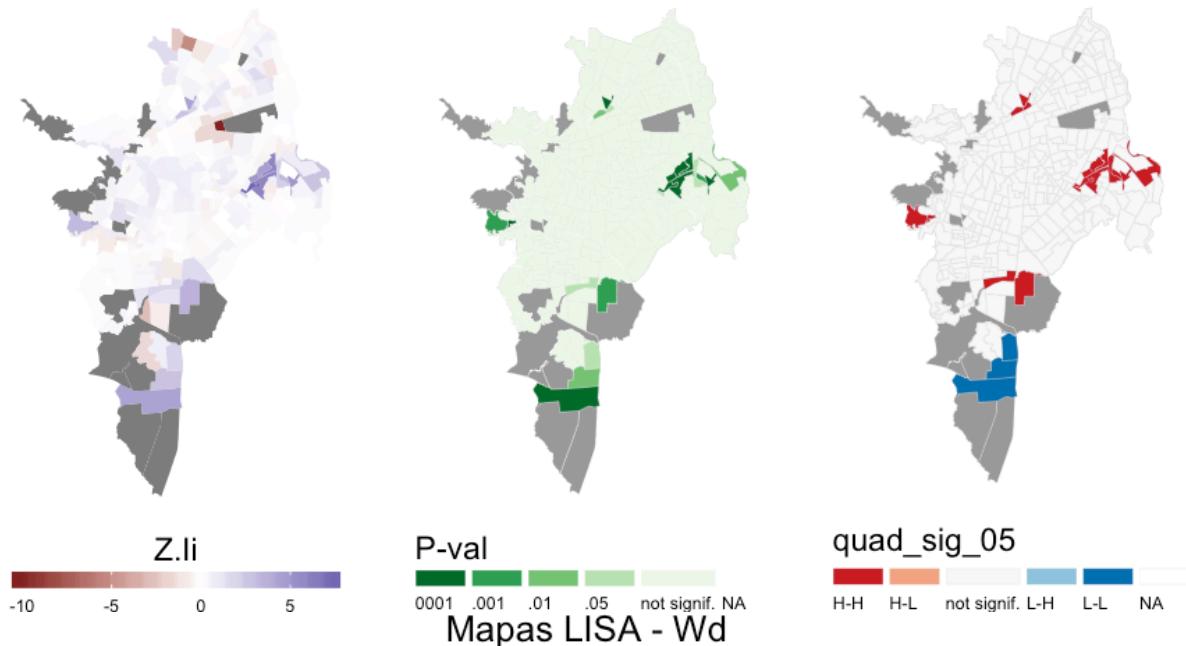
Los mapas de LISA para los residuos de ambos modelos con la matriz  $W_q$  se muestran a continuación.

#### 4.2.3.3. Ajuste de modelos espaciales EV

Probamos los 4 tipos de modelos con la matriz  $W_q$  que resultó capturar mejor la asociación espacial en los datos y comparamos sus resultados.

##### 4.2.3.3.1. Porcentaje de espacio verde

Mapas LISA - Wd  
Residuos:  
area\_ep.porcentaje.mxn ~ cuarto.porcentaje.mxn + area\_media\_manzana.mxn



Residuos:  
ia.Areas.dist.mxn ~ area\_media\_manzana.mxn + cuarto.porcentaje.mxn +  
unidad\_economica.porcentaje.mxn + densidad\_poblacion.mxn +  
con\_alguna\_limitacion.porcentaje.mxn + ningun\_estudio.porcentaje.mxn

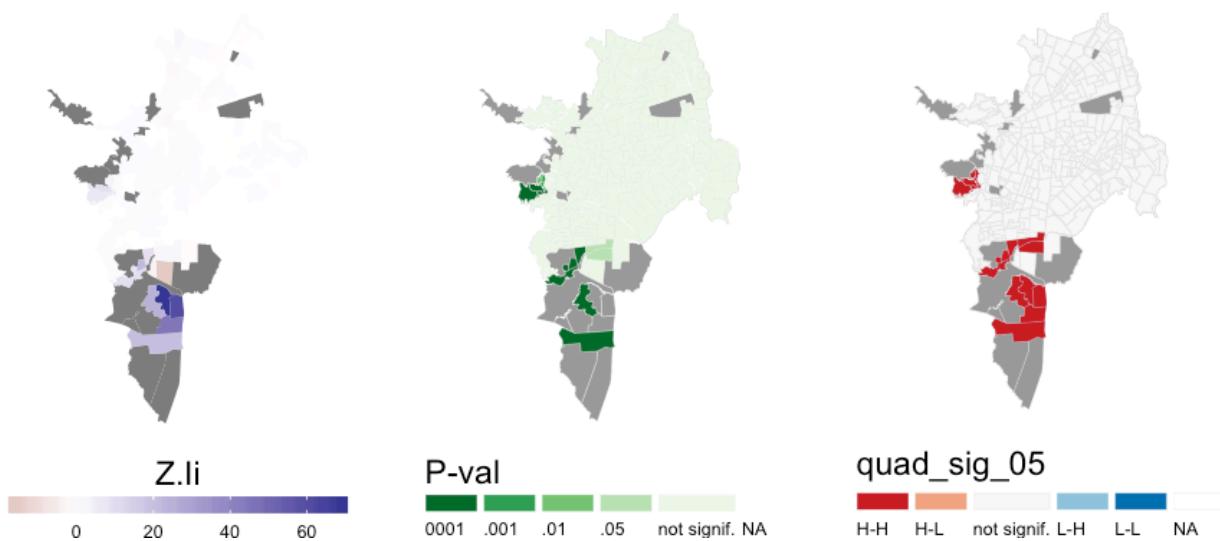


Figura 4.48: Mapas LISA para la matriz  $W_q$  de los residuos de los modelos lineales para los indicadores de acceso a EV

#SAR

```
sar.areas_ep<-lagsarlm(formula = as.formula(lm.area_ep.ptje.sel),
                         data = lm.area_ep.ptje.sel$model,
                         listw = W_queen.ev,
                         zero.policy = T,
                         tol.solve=1.0e-30)

sar.sm<-summary(sar.areas_ep,Nagelkerke=T)
sar.sm

##  
## Call:  
## lagsarlm(formula = as.formula(lm.area_ep.ptje.sel), data = lm.area_ep.ptje.sel$model  
##   listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30)  
##  
## Residuals:  
##       Min        1Q     Median        3Q       Max
## -0.273098 -0.040970 -0.014876  0.019167  0.707957
##  
## Type: lag  
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.0388741  0.0073986  5.2542 1.486e-07
## cuarto.porcentaje.mxn -0.0807630  0.0376324 -2.1461  0.03186
## area_media_manzana.mxn  0.7335412  0.0654545 11.2069 < 2.2e-16
##  
## Rho: 0.15309, LR test value: 4.5439, p-value: 0.033036
## Asymptotic standard error: 0.072979
```

```

##      z-value: 2.0977, p-value: 0.03593
## Wald statistic: 4.4004, p-value: 0.03593
##
## Log likelihood: 375.1965 for lag model
## ML residual variance (sigma squared): 0.0059579, (sigma: 0.077187)
## Nagelkerke pseudo-R-squared: 0.31043
## Number of observations: 329
## Number of parameters estimated: 5
## AIC: -740.39, (AIC for lm: -737.85)
## LM test for residual autocorrelation
## test value: 5.7788, p-value: 0.016221

```

#### # SEM

```

sem.areas_ep<-errorsarlm(formula = as.formula(lm.area_ep.ptje.sel),
                           data = lm.area_ep.ptje.sel$model,
                           listw = W_queen.ev, zero.policy = T,
                           tol.solve=1.0e-30)

sem.sm<-summary(sem.areas_ep,
                  Nagelkerke=T)

sem.sm

```

##

## Call:

```
## errorsarlm(formula = as.formula(lm.area_ep.ptje.sel), data = lm.area_ep.ptje.sel$mode
```

```
##      listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30)
```

##

## Residuals:

	Min	1Q	Median	3Q	Max
--	-----	----	--------	----	-----

##	-0.279173	-0.039125	-0.014652	0.021290	0.708489
----	-----------	-----------	-----------	----------	----------

##

```

## Type: error
## Coefficients: (asymptotic standard errors)
##
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)           0.0491056  0.0069003 7.1165 1.107e-12
## cuarto.porcentaje.mxn -0.0979094  0.0415838 -2.3545  0.01855
## area_media_manzana.mxn  0.7712382  0.0653415 11.8032 < 2.2e-16
##
## Lambda: 0.25902, LR test value: 10.215, p-value: 0.0013934
## Asymptotic standard error: 0.078665
##      z-value: 3.2928, p-value: 0.00099206
## Wald statistic: 10.842, p-value: 0.00099206
##
## Log likelihood: 378.0319 for error model
## ML residual variance (sigma squared): 0.0058066, (sigma: 0.076201)
## Nagelkerke pseudo-R-squared: 0.32222
## Number of observations: 329
## Number of parameters estimated: 5
## AIC: -746.06, (AIC for lm: -737.85)

```

#SD

```

sd.areas_ep<-lagsarlm(formula = as.formula(lm.area_ep.ptje.sel),
                        data = lm.area_ep.ptje.sel$model,
                        listw = W_queen.ev,
                        zero.policy = T,
                        tol.solve=1.0e-30,
                        type = "mixed")

sd.sm<-summary(sd.areas_ep,
                Nagelkerke=T)

sd.sm

```

```

##
## Call:
## lagsarlm(formula = as.formula(lm.area_ep.ptje.sel), data = lm.area_ep.ptje.sel$model,
##           listw = W_queen.ev, type = "mixed", zero.policy = T, tol.solve = 1e-30)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.281818 -0.041196 -0.015641  0.025087  0.701126
##
## Type: mixed
## Coefficients: (asymptotic standard errors)
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.0377326  0.0086946  4.3398 1.426e-05
## cuarto.porcentaje.mxn -0.1366936  0.0544000 -2.5128  0.011979
## area_media_manzana.mxn  0.7929675  0.0665592 11.9137 < 2.2e-16
## lag.cuarto.porcentaje.mxn  0.0985071  0.0829144  1.1881  0.234811
## lag.area_media_manzana.mxn -0.3647870  0.1326357 -2.7503  0.005954
##
## Rho: 0.26168, LR test value: 10.614, p-value: 0.0011222
## Asymptotic standard error: 0.078409
##      z-value: 3.3374, p-value: 0.0008458
## Wald statistic: 11.138, p-value: 0.0008458
##
## Log likelihood: 379.4362 for mixed model
## ML residual variance (sigma squared): 0.0057556, (sigma: 0.075866)
## Nagelkerke pseudo-R-squared: 0.32798
## Number of observations: 329
## Number of parameters estimated: 7
## AIC: -744.87, (AIC for lm: -736.26)

```

```

## LM test for residual autocorrelation
## test value: 1.9206, p-value: 0.16579

#SLX

slx.areas_ep<-lmSLX(formula = as.formula(lm.area_ep.ptje.sel),
                      data = lm.area_ep.ptje.sel$model,
                      listw = W_queen.ev,
                      zero.policy = T)

slx.sm<-summary(slx.areas_ep)
slx.sm

##
## Call:
## lm(formula = y ~ x - 1, weights = weights)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.26922 -0.04573 -0.01615  0.02428  0.69203
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## x(Intercept)            0.050863  0.007944  6.402 5.37e-10 ***
## xcuarto.porcentaje.mxn -0.133837  0.056072 -2.387  0.0176 *
## xarea_media_manzana.mxn  0.776764  0.068460 11.346 < 2e-16 ***
## xlag.cuarto.porcentaje.mxn  0.066635  0.085226  0.782  0.4349
## xlag.area_media_manzana.mxn -0.158851  0.124959 -1.271  0.2046
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

Cuadro 4.5: Metricas de ajuste para los modelos de porcentaje de área de EV

medidasfit	OLS	SLX	SAR	SEM	SD
Globla Moran'I	0.11636	0.12094	0.04015	-0.00817	-0.00602
GMI p-value	0.00014	0.00008	0.09378	0.56220	0.53618
Shapiro-Wilk	0.76782	0.77525	0.76040	0.75813	0.76159
SW p-value	0.00000	0.00000	0.00000	0.00000	0.00000
Breusch-Pagan	12.98572	14.59961	13.97239	11.07822	13.54161
BP p-value	0.00151	0.00561	0.00092	0.00393	0.00891
Media Residuos	0.00000	0.00000	0.00000	0.00000	0.00000
MSE	0.00607	0.00602	0.00596	0.00581	0.00576
adj-Rsquare	0.29656	0.52434	NA	NA	NA
Nagelkerke pseudo-R-squared	NA	NA	0.31043	0.32222	0.32798
AIC	-737.84913	-736.25817	-740.39306	-746.06370	-744.87233
Log likelihood	372.92456	374.12908	375.19653	378.03185	379.43617

```
## Residual standard error: 0.0782 on 324 degrees of freedom
## Multiple R-squared:  0.5316, Adjusted R-squared:  0.5243
## F-statistic: 73.53 on 5 and 324 DF,  p-value: < 2.2e-16

AIC(slx.areas_ep)

## [1] -736.2582
```

Al comparar los resultados de las métricas de ajuste se identifica al modelo SEM con el mejor AIC. El modelo SEM logra eliminar la autocorrelación espacial global en los residuos, al igual que SD (ver figura 4.49). Aunque persiste la no normalidad de los residuos y la heterocedasticidad como los muestran los test y las gráficas diagnósticas, el error cometido disminuye y los coeficientes pueden considerarse más confiables. El  $\lambda$  de las estimaciones con términos autorregresivos es significativo, esto sugiere que no es necesario plantear efectos distintivos de la variables dependiente rezagada, y que es posible que ese efecto sea por otras variables no tenidas en cuenta: el agrupamiento espacial observado en la variable dependiente se explica simplemente por el patrón geográfico de variables independientes medidas y no medidas, pero no genera ningún efecto de derrame en el acceso a espacios verdes de los sectores vecinos. A pesar de que los test de normalidad y heterocedasticidad

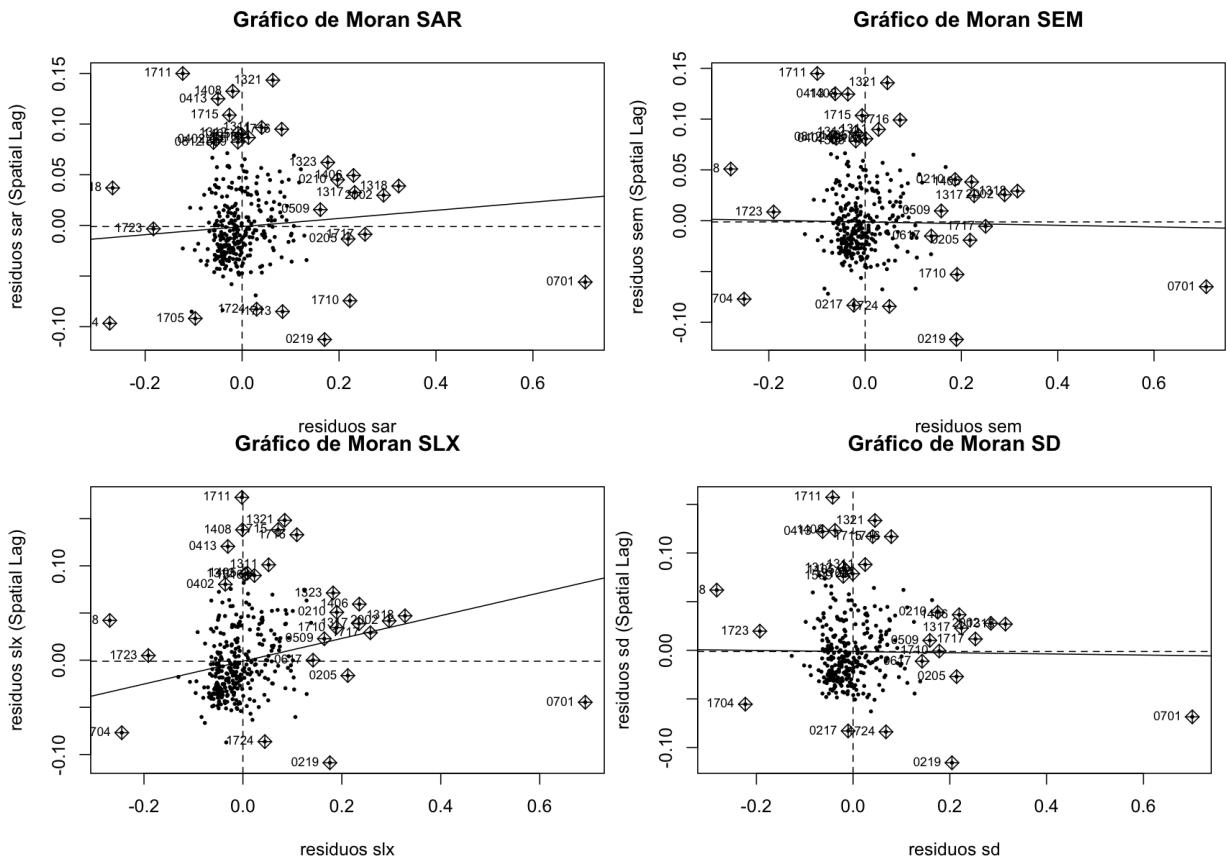


Figura 4.49: Gráfico de Moran para los residuos de los modelos espaciales de porcentaje de área de EV  $W_q$

no son exitosos, las gráficas diagnósticas muestran que los problemas se presentan en los valores extremos.

#### 4.2.3.3.2. Índice de acceso área-distancia

#SAR

```
sar.areas.dist<-lagsarlm(formula = as.formula(lm.ia.areas.dist.sel),
  data = lm.ia.areas.dist.sel$model,
  listw = W_queen.ev, zero.policy = T,tol.solve=1.0e-30)

sar.sm<-summary(sar.areas.dist,Nagelkerke=T)
```

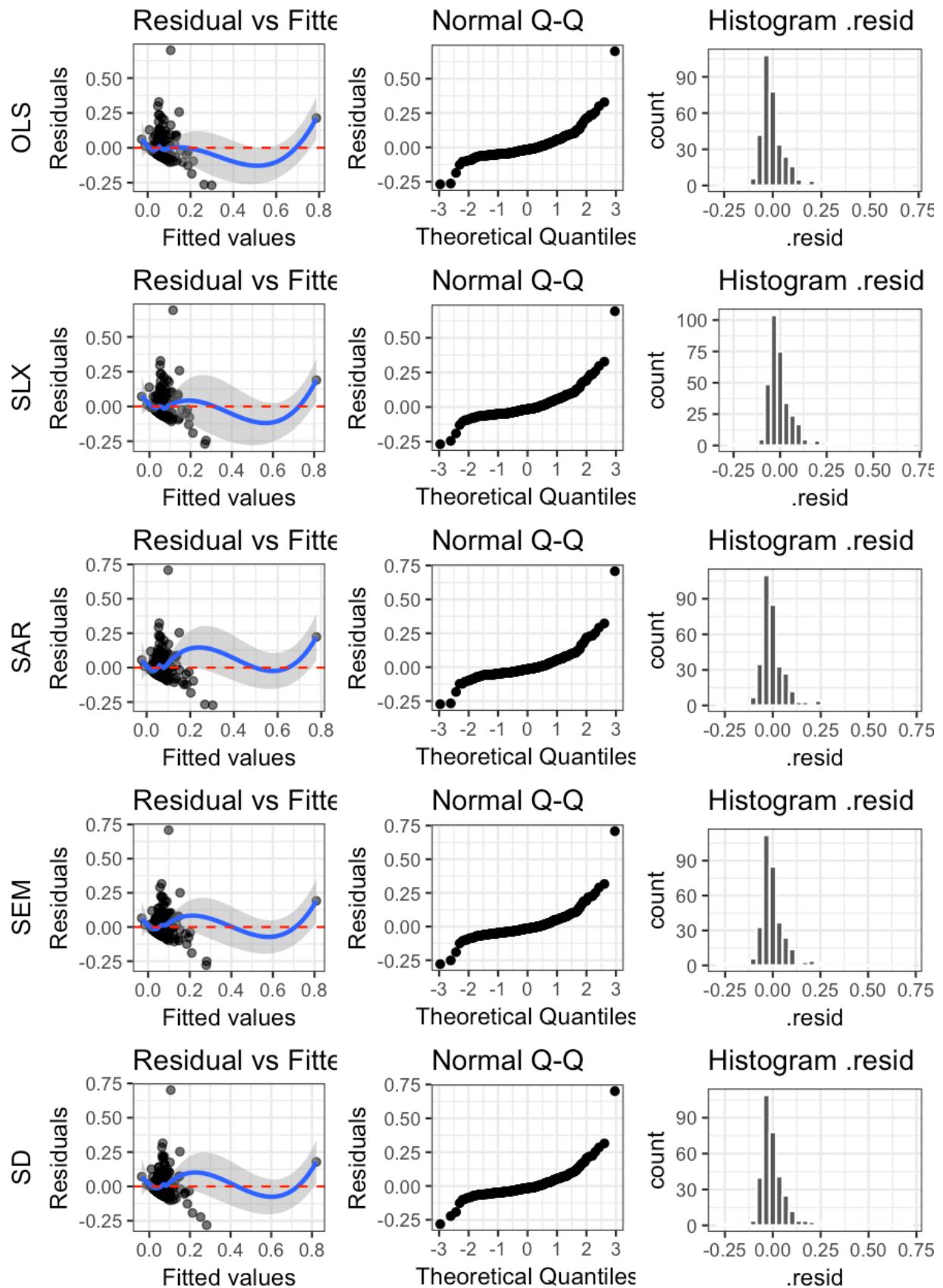


Figura 4.50: Diagnóstico comparativo entre modelos  
170

```
sar.sm
```

```
##  
## Call:  
## lagsarlm(formula = as.formula(lm.ia.areas.dist.sel), data = lm.ia.areas.dist.sel$mo  
##      listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30)  
##  
## Residuals:  
##       Min        1Q     Median        3Q       Max  
## -0.2780579 -0.0122299 -0.0041903  0.0053696  0.4480791  
##  
## Type: lag  
## Coefficients: (asymptotic standard errors)  
##  
##                                     Estimate Std. Error z value  
## (Intercept)                 0.0222944  0.0116285  1.9172  
## area_media_manzana.mxn      0.0403663  0.0474079  0.8515  
## cuarto.porcentaje.mxn      -0.0478921  0.0325380 -1.4719  
## unidad_economica.porcentaje.mxn  0.0543715  0.0227131  2.3938  
## densidad_poblacion.mxn      -0.0096933  0.0181712 -0.5334  
## con_alguna_limitacion.porcentaje.mxn -0.0447170  0.0219950 -2.0331  
## ningun_estudio.porcentaje.mxn      0.0417468  0.0295132  1.4145  
##  
##                                     Pr(>|z|)  
## (Intercept)                 0.05521  
## area_media_manzana.mxn      0.39451  
## cuarto.porcentaje.mxn      0.14105  
## unidad_economica.porcentaje.mxn  0.01667  
## densidad_poblacion.mxn      0.59373  
## con_alguna_limitacion.porcentaje.mxn  0.04205  
## ningun_estudio.porcentaje.mxn      0.15721  
##
```

```

## Rho: 0.81871, LR test value: 325.5, p-value: < 2.22e-16
## Asymptotic standard error: 0.032947
##      z-value: 24.85, p-value: < 2.22e-16
## Wald statistic: 617.5, p-value: < 2.22e-16
##
## Log likelihood: 471.5556 for lag model
## ML residual variance (sigma squared): 0.0027799, (sigma: 0.052725)
## Nagelkerke pseudo-R-squared: 0.70529
## Number of observations: 329
## Number of parameters estimated: 9
## AIC: -925.11, (AIC for lm: -601.61)
## LM test for residual autocorrelation
## test value: 90.542, p-value: < 2.22e-16

```

#### # SEM

```

sem.areas.dist<-errorsarlm(formula = as.formula(lm.ia.areas.dist.sel),
                             data = lm.ia.areas.dist.sel$model,
                             listw = W_queen.ev, zero.policy = T,tol.solve=1.0e-10)
sem.sm<-summary(sem.areas.dist,Nagelkerke=T)
sem.sm

```

```
##
```

```
## Call:
```

```
## errorsarlm(formula = as.formula(lm.ia.areas.dist.sel), data = lm.ia.areas.dist.sel$mo
##      listw = W_queen.ev, zero.policy = T, tol.solve = 1e-30)
```

```
##
```

```
## Residuals:
```

	Min	1Q	Median	3Q	Max
##	-0.2739716	-0.0130918	-0.0043041	0.0044020	0.4534021
##					

```

## Type: error

## Coefficients: (asymptotic standard errors)

##                                     Estimate Std. Error z value
## (Intercept)                 0.0716311  0.0256815 2.7892
## area_media_manzana.mxn     -0.0985923  0.0471135 -2.0927
## cuarto.porcentaje.mxn      0.0016561  0.0444938  0.0372
## unidad_economica.porcentaje.mxn 0.0644677  0.0292013 2.2077
## densidad_poblacion.mxn     -0.0168933  0.0227898 -0.7413
## con_alguna_limitacion.porcentaje.mxn -0.0381619  0.0242516 -1.5736
## ningun_estudio.porcentaje.mxn      0.0180572  0.0426374  0.4235
##                                     Pr(>|z|)
## (Intercept)                  0.005284
## area_media_manzana.mxn       0.036380
## cuarto.porcentaje.mxn        0.970310
## unidad_economica.porcentaje.mxn 0.027266
## densidad_poblacion.mxn       0.458532
## con_alguna_limitacion.porcentaje.mxn 0.115584
## ningun_estudio.porcentaje.mxn      0.671925
##
## Lambda: 0.8581, LR test value: 320.95, p-value: < 2.22e-16
## Asymptotic standard error: 0.02916
##      z-value: 29.428, p-value: < 2.22e-16
## Wald statistic: 865.98, p-value: < 2.22e-16
##
## Log likelihood: 469.2818 for error model
## ML residual variance (sigma squared): 0.0027409, (sigma: 0.052354)
## Nagelkerke pseudo-R-squared: 0.70119
## Number of observations: 329
## Number of parameters estimated: 9

```

```

## AIC: -920.56, (AIC for lm: -601.61)

#SD

sd.areas.dist<-lagsarlm(formula = as.formula(lm.ia.areas.dist.sel),
                         data = lm.ia.areas.dist.sel$model,
                         listw = W_queen.ev, zero.policy = T,tol.solve=1.0e-3
                         type = "mixed")

sd.sm<-summary(sd.areas.dist,Nagelkerke=T)
sd.sm

## Call:
## lagsarlm(formula = as.formula(lm.ia.areas.dist.sel), data = lm.ia.areas.dist.sel$mode
##           listw = W_queen.ev, type = "mixed", zero.policy = T, tol.solve = 1e-30)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.2816380 -0.0138278 -0.0016630  0.0081623  0.4186494
##
## Type: mixed
## Coefficients: (asymptotic standard errors)
##                                     Estimate Std. Error z value
## (Intercept)                 -0.01785763  0.01768939 -1.0095
## area_media_manzana.mxn      0.02917490  0.05038592  0.5790
## cuarto.porcentaje.mxn     -0.00296323  0.04346976 -0.0682
## unidad_economica.porcentaje.mxn 0.06684156  0.02812408  2.3767
## densidad_poblacion.mxn     -0.00518664  0.02222968 -0.2333
## con_alguna_limitacion.porcentaje.mxn -0.04284384  0.02317612 -1.8486
## ningun_estudio.porcentaje.mxn      0.03467732  0.04144553  0.8367

```

```

## lag.area_media_manzana.mxn          0.73768946  0.10528513  7.0066
## lag.cuarto.porcentaje.mxn         -0.08865463  0.06965991 -1.2727
## lag.unidad_economica.porcentaje.mxn 0.01171474  0.04316099  0.2714
## lag.densidad_poblacion.mxn        0.02325558  0.03739340  0.6219
## lag.con_alguna_limitacion.porcentaje.mxn -0.00053905  0.04274059 -0.0126
## lag.ningun_estudio.porcentaje.mxn    0.08424458  0.06092012  1.3829
##
##                                         Pr(>|z|)
## (Intercept)                         0.31273
## area_media_manzana.mxn              0.56257
## cuarto.porcentaje.mxn              0.94565
## unidad_economica.porcentaje.mxn   0.01747
## densidad_poblacion.mxn             0.81551
## con_alguna_limitacion.porcentaje.mxn 0.06451
## ningun_estudio.porcentaje.mxn     0.40276
## lag.area_media_manzana.mxn         2.442e-12
## lag.cuarto.porcentaje.mxn         0.20313
## lag.unidad_economica.porcentaje.mxn 0.78607
## lag.densidad_poblacion.mxn        0.53400
## lag.con_alguna_limitacion.porcentaje.mxn 0.98994
## lag.ningun_estudio.porcentaje.mxn  0.16670
##
## Rho: 0.7088, LR test value: 213.39, p-value: < 2.22e-16
## Asymptotic standard error: 0.041984
##      z-value: 16.883, p-value: < 2.22e-16
## Wald statistic: 285.02, p-value: < 2.22e-16
##
## Log likelihood: 499.9153 for mixed model
## ML residual variance (sigma squared): 0.0024821, (sigma: 0.049821)
## Nagelkerke pseudo-R-squared: 0.75196

```

```

## Number of observations: 329
## Number of parameters estimated: 15
## AIC: -969.83, (AIC for lm: -758.44)
## LM test for residual autocorrelation
## test value: 43.178, p-value: 4.9987e-11

#SLX

slx.areas.dist<-lmSLX(formula = as.formula(lm.ia.areas.dist.sel),
                       data = lm.ia.areas.dist.sel$model,
                       listw = W_queen.ev, zero.policy = T)

slx.sm<-summary(slx.areas.dist)
slx.sm

## Call:
## lm(formula = y ~ x - 1, weights = weights)
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32193 -0.02981 -0.00631  0.01504  0.58292
## Coefficients:
##                               Estimate Std. Error t value
## x(Intercept)                0.032537  0.026513  1.227
## xarea_media_manzana.mxn     0.316684  0.070266  4.507
## xcuarto.porcentaje.mxn    -0.049288  0.065134 -0.757
## xunidad_economica.porcentaje.mxn  0.063757  0.042176  1.512
## xdensidad_poblacion.mxn     0.012842  0.033313  0.385
## xcon_alguna_limitacion.porcentaje.mxn -0.059156  0.034691 -1.705

```

```

## xningun_estudio.porcentaje.mxn      0.076228  0.062044  1.229
## xlag.area_media_manzana.mxn        1.509318  0.142478 10.593
## xlag.cuarto.porcentaje.mxn       -0.259880  0.103035 -2.522
## xlag.unidad_economica.porcentaje.mxn 0.177529  0.062806  2.827
## xlag.densidad_poblacion.mxn       0.003132  0.056084  0.056
## xlag.con_alguna_limitacion.porcentaje.mxn -0.205932  0.063491 -3.243
## xlag.ningun_estudio.porcentaje.mxn   0.349373  0.089964  3.883
##
##                                         Pr(>|t|)
## x(Intercept)                         0.220648
## xarea_media_manzana.mxn             9.27e-06 ***
## xcuarto.porcentaje.mxn            0.449787
## xunidad_economica.porcentaje.mxn  0.131613
## xdensidad_poblacion.mxn           0.700131
## xcon_alguna_limitacion.porcentaje.mxn 0.089134 .
## xningun_estudio.porcentaje.mxn     0.220131
## xlag.area_media_manzana.mxn        < 2e-16 ***
## xlag.cuarto.porcentaje.mxn        0.012151 *
## xlag.unidad_economica.porcentaje.mxn 0.005004 **
## xlag.densidad_poblacion.mxn       0.955503
## xlag.con_alguna_limitacion.porcentaje.mxn 0.001307 **
## xlag.ningun_estudio.porcentaje.mxn 0.000125 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07472 on 316 degrees of freedom
## Multiple R-squared:  0.6236, Adjusted R-squared:  0.6081
## F-statistic: 40.27 on 13 and 316 DF,  p-value: < 2.2e-16

AIC(slx.areas.dist)

```

Cuadro 4.6: Metricas de ajuste para los modelos de áreas-distancia de EV

medidasfit	OLS	SLX	SAR	SEM	SD
Globla Moran'I	0.58543	0.53694	-0.18256	-0.17227	-0.12540
GMI p-value	0.00000	0.00000	1.00000	1.00000	0.99991
Shapiro-Wilk	0.57265	0.66619	0.51268	0.50161	0.60898
SW p-value	0.00000	0.00000	0.00000	0.00000	0.00000
Breusch-Pagan	57.30864	63.06072	20.13485	6.43007	81.81776
BP p-value	0.00000	0.00000	0.00262	0.37677	0.00000
Media Residuos	0.00000	0.00000	0.00000	0.00000	0.00000
MSE	0.00896	0.00536	0.00278	0.00274	0.00248
adj-Rsquare	0.19261	0.60809	NA	NA	NA
Nagelkerke pseudo-R-squared	NA	NA	0.70529	0.70119	0.75196
AIC	-601.61394	-758.43674	-925.11120	-920.56367	-969.83068
Log likelihood	308.80697	393.21837	471.55560	469.28183	499.91534

```
## [1] -758.4367
```

Al comparar los resultados de las métricas de ajuste se identifica al modelo SD con el mejor AIC. El modelo SD logra eliminar la autocorrelación espacial global en los residuos al exhibir un valor de significancia mucho mayor que 0.05, al igual que SEM y SAR (ver figura 4.51). Aunque persiste la no normalidad de los residuos y la heterocedasticidad como lo muestran los test y las gráficas diagnósticas, el error cometido disminuye y los coeficientes pueden considerarse más confiables. El  $\rho$  de las estimaciones con términos autorregresivos es significativo, esto sugiere que una parte de los efectos de derrame de las variables significativas (`lag.area_media_manzana`, `unidad_economica.porcentaje`) en el acceso a espacios verdes de los sectores vecinos explica bien los grupos que se forman. Es posible que estar cerca de un sector con manzanas grandes y que tal vez sus parque pueden ser más grandes explica la influencia positiva en el acceso p.e. de tener equipamientos de ciudad grandes en algún sector aledaño. En el modelo SD una parte de la mejora en el ajuste puede provenir de dimensiones no modeladas. Los test de normalidad y heterocedasticidad no son exitosos, las gráficas diagnósticas muestran que los problemas se presentan en los valores extremos.

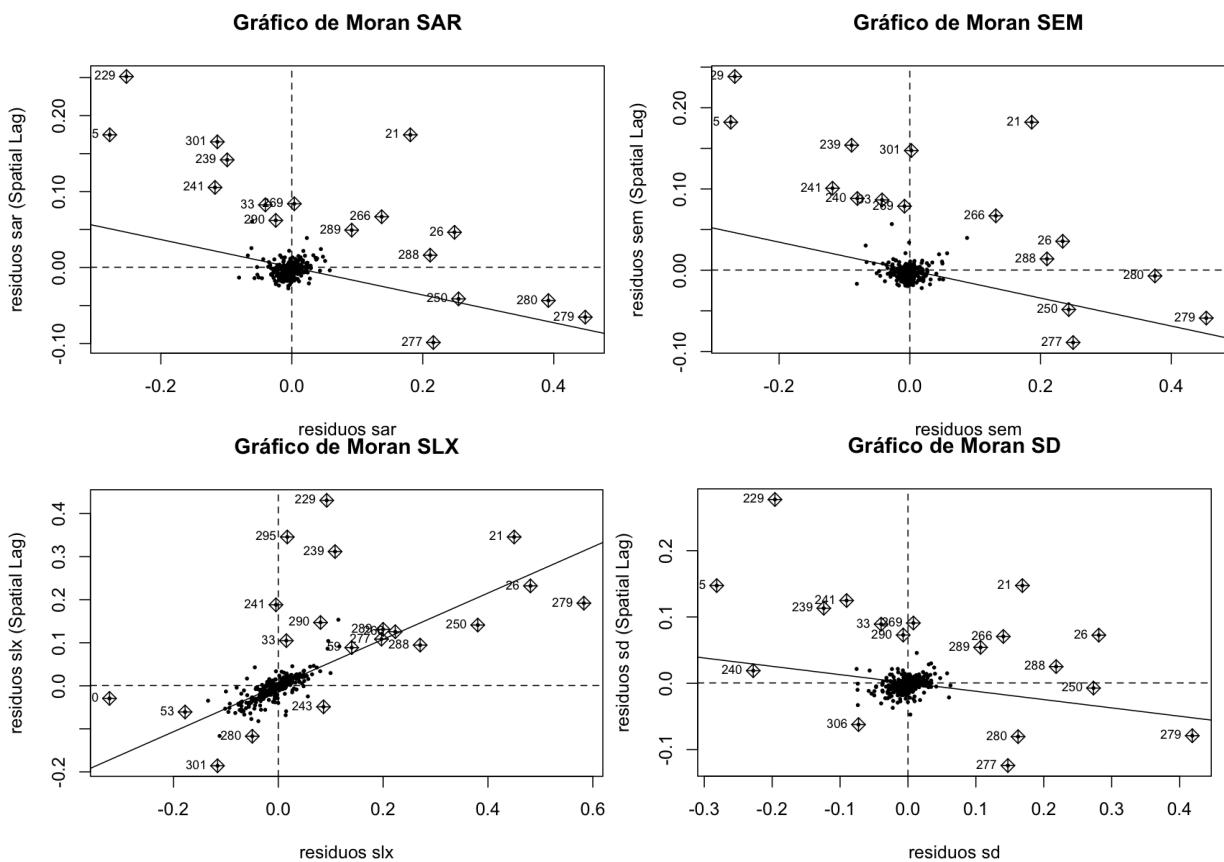


Figura 4.51: Gráfico de Moran para los residuos de los modelos espaciales de acceso área-distancia  $W_q$

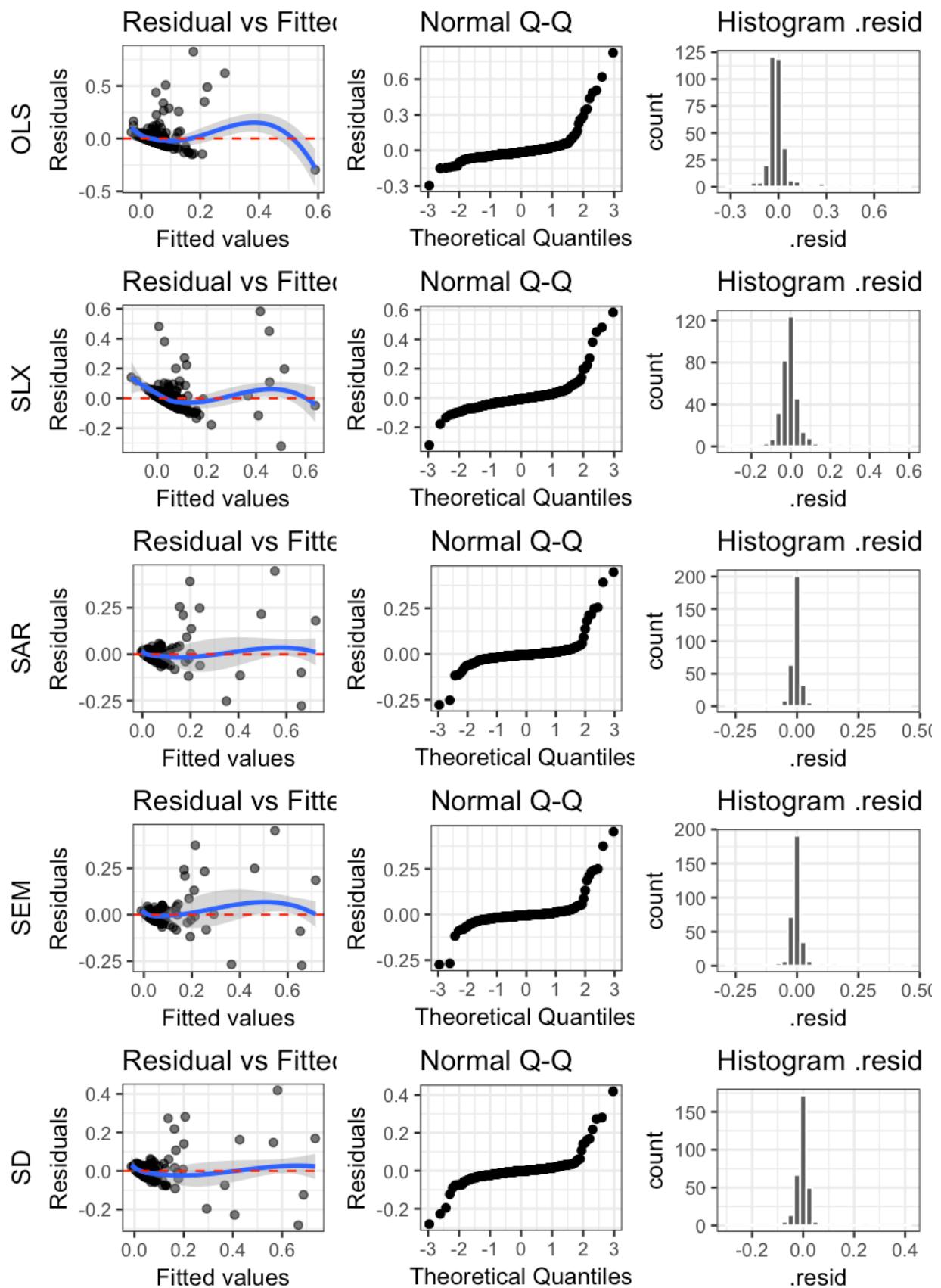


Figura 4.52: Diagnóstico comparativo entre modelos  
180

# **Capítulo 5**

## **Discusión**

Antes de iniciar con la discusión de los resultados es importante recordar que los modelos realizados son un intento por capturar evidencia empírica sobre el nivel explicativo de un conjunto de condiciones de la población y las condiciones de vida que provee la ciudad de Cali en lo que beneficios ambientales se refiere. El modelo no es la realidad ni prueba de causalidad, es un esfuerzo por cuantificar la correlación entre algunas de esas dimensiones y tener certeza de su potencial asociación. El uso de modelos espaciales indaga sobre los patrones espaciales de esas variables, lo que permite la identificación de zonas que agrupan características altas de esos servicios o de las condiciones de la población y de la estructura urbana. Los análisis no buscan hacer inferencia sobre la población, pues no son una muestra, son la población completa de árboles, espacio verde y los datos del censo. En este sentido los coeficientes y su representatividad son sobre la población, y los interpretaremos como la importancia relativa de esas variables sobre las otras.

### **5.1. Sobre la cobertura de copa**

Aunque el área de copa y el porcentaje de cobertura de copa configuran dos modelos individuales y son indicadores que expresan de diferente forma el nivel de acceso a servicios

del arbolado urbano, ambos son visiones complementarias sobre el mismo fenómeno, la interpretación no aísla los modelos, los relaciona.

A pesar de la persistencia de problemas en la estimación como la no normalidad de los residuos y la heterocedasticidad —esto debidos ya no la autocorrelación de la variable dependiente ni a la de los errores sino a posibles no linealidades entre los predictores y la cobertura de copa— las estimaciones mejoran con los modelos con estructura espacial. Inclusive parece que las matrices de vecindad construidas, aunque muy similares entre sí, muestran que la inclusión a priori de una estructura espacial es un mecanismo eficaz para identificar grupos y probar aspectos teóricos de la estructura de la autocorrelación si se profundiza e incluye criterios teóricos o conocimientos del desarrollo histórico de la ciudad en su formalización. Es posible que dichos problemas de la estimación provengan de dimensiones no incluidas en el modelo dado que los modelos SEM también tienen un rendimiento similar al de los mejor rankeados SAR y SD, escogidos como mejores candidatos para la estimación del porcentaje de cobertura de copa y área de copa, respectivamente. En este sentido traemos a colación aspectos como el importante número de árboles en los andes y separadores viales (figura 3.3), factor que no puede no ser capturado en las métricas construidas, pero que puede explicar el agrupamiento de sectores censales con buena cobertura de copa alrededor de la calle 5ta que se observa en el mapa de clusters del porcentaje de cobertura de copa (figura 4.20) o el bajo desarrollo de los individuos en sectores de calles estrechas y con pocas vías principales.

Así pues los coeficientes de los dos modelos ajustados se detallan en las tablas 5.1 y 5.2 para el modelo SD y 5.3 y 5.4 para el modelo SAR. En consecuencia con el ajuste de autocorrelación en ambos modelos (SAR y SD) el  $\rho$  es significativo y proporcional a las diferencias en el Moran's I Global de cada modelo. Es interesante que en ambos modelos la variable más significativa y de mayor valor en el coeficiente es **estudios superiores**. Cuando se formulan los modelos se excluye el uso de variables como porcentaje de población afro o personas sin estudios por tener una alta correlación negativa con personas con **estudios superiores o postgrado** o con su versión porcentual. Esto implica que los beneficios am-

Cuadro 5.1: Coeficientes del modelo SD de área de copa

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.665	0.067	9.857	0.000
superior_postgrado.mxn	0.287	0.024	11.741	0.000
densidad_poblacion.mxn	-0.088	0.026	-3.431	0.001
con_alguna_limitacion.porcentaje.mxn	0.007	0.027	0.273	0.785
afro.porcentaje.mxn	-0.041	0.040	-1.013	0.311
cuarto.porcentaje.mxn	-0.062	0.042	-1.486	0.137
area_ep.mxn	0.134	0.030	4.455	0.000
lag.superior_postgrado.mxn	0.040	0.051	0.789	0.430
lag.densidad_poblacion.mxn	-0.103	0.044	-2.316	0.021
lag.con_alguna_limitacion.porcentaje.mxn	-0.060	0.046	-1.292	0.196
lag.afro.porcentaje.mxn	0.116	0.051	2.268	0.023
lag.cuarto.porcentaje.mxn	-0.123	0.068	-1.823	0.068
lag.area_ep.mxn	-0.103	0.042	-2.474	0.013

Cuadro 5.2: Coeficiente de autocorrelación modelo SD de área de copa

\$\rho\$	Likelihood ratio	p-value
0.175	4.82	0.028

bientales son mejores en sectores con población mejor educada (figura 5.3), posiblemente una razón para preferir o habitar espacios con buena arborización o tal vez con andes con suficiente espacio y zonas blandas para arborizar el barrio. En cualquier caso, esto muestra un grado de segregación de los clusters de sectores con mayor arborización (figura 5.3) y la población afro (figura 5.1) y sin estudios (figura 5.2), ambos altamente correlacionadas, como se aprecia en los mapas de LISA y las matriz de correlación 4.3 y 4.4

La densidad de población es una variable que produce resultados inconsistentes: aunque

Cuadro 5.3: Coeficientes del modelo SAR de porcentaje de área de copa

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.030	0.011	2.822	0.005
superior_postgrado.porcentaje.mxn	0.246	0.027	9.023	0.000

Cuadro 5.4: Coeficiente de autocorrelación modelo SAR de área de copa

$\$\\rho\$$	Likelihood ratio	p-value
0.375	28.579	0

Mapas LISA - afro.porcentaje - Wd

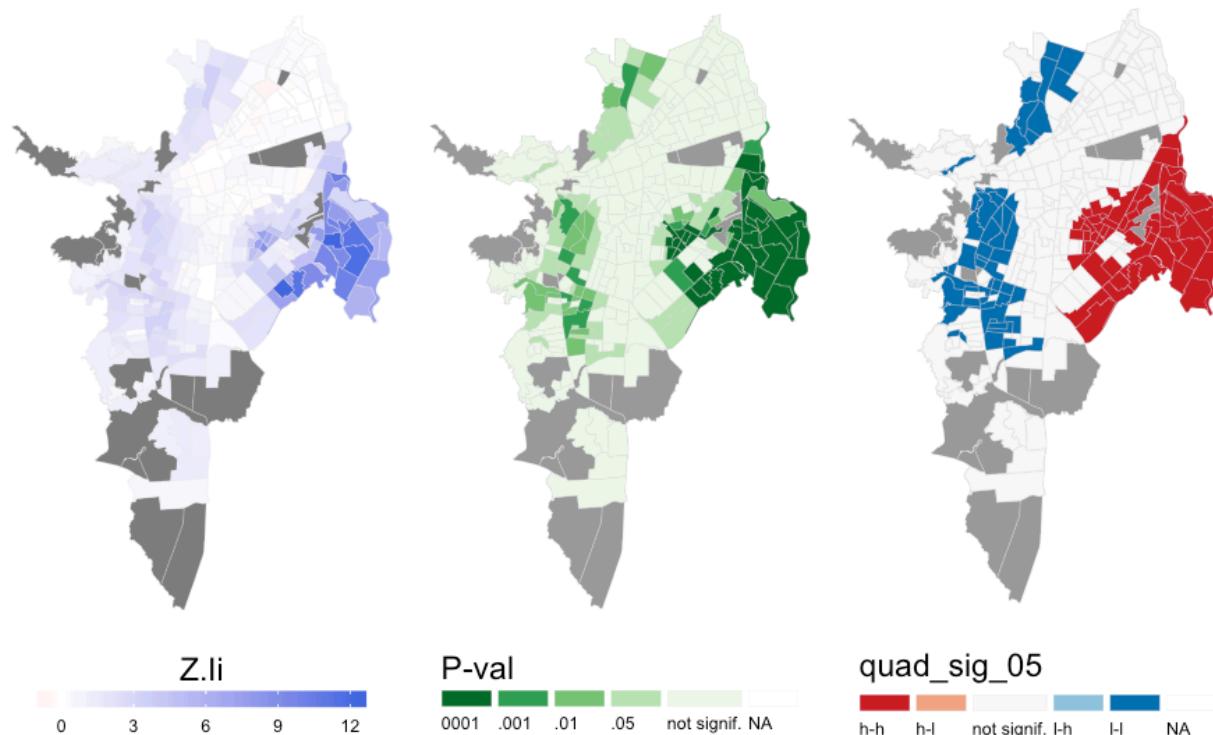


Figura 5.1: Mapas LISA - Porcentaje población Afro

Mapas LISA - ningun\_estudio.porcentaje - Wd

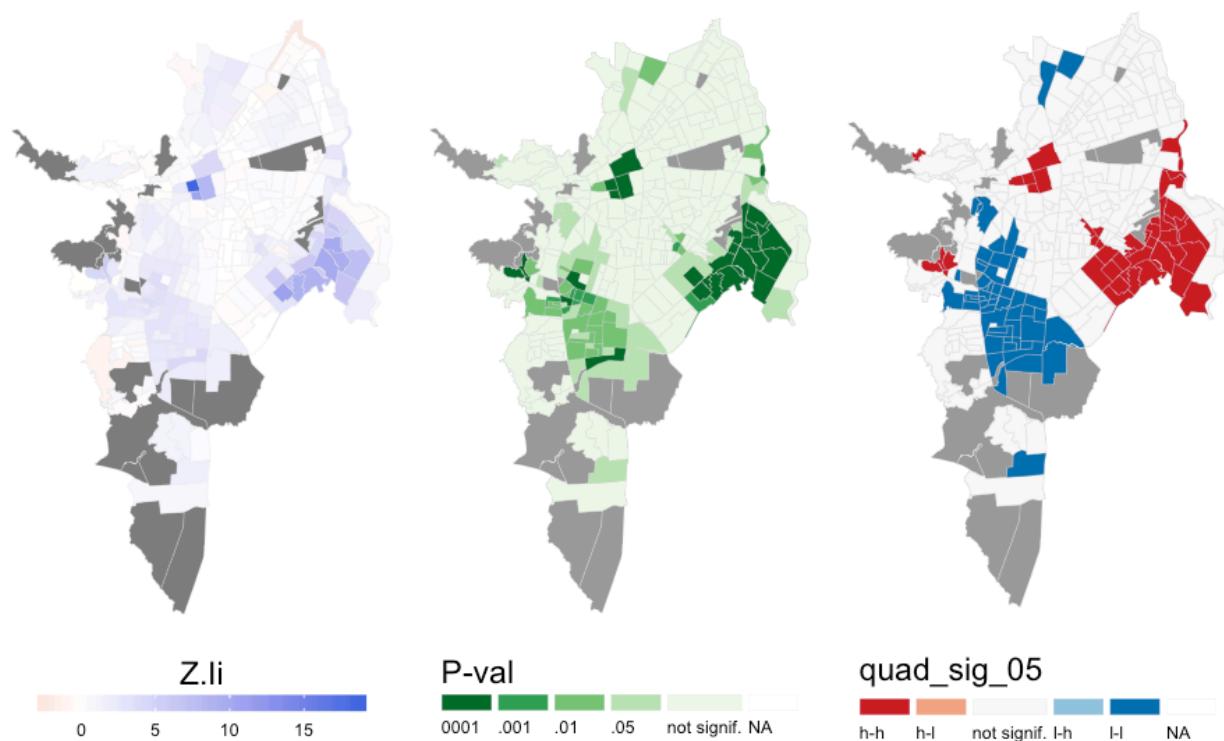


Figura 5.2: Mapas LISA - Porcentaje población sin estudios

Mapas LISA - superior\_postgrado.porcentaje - Wd

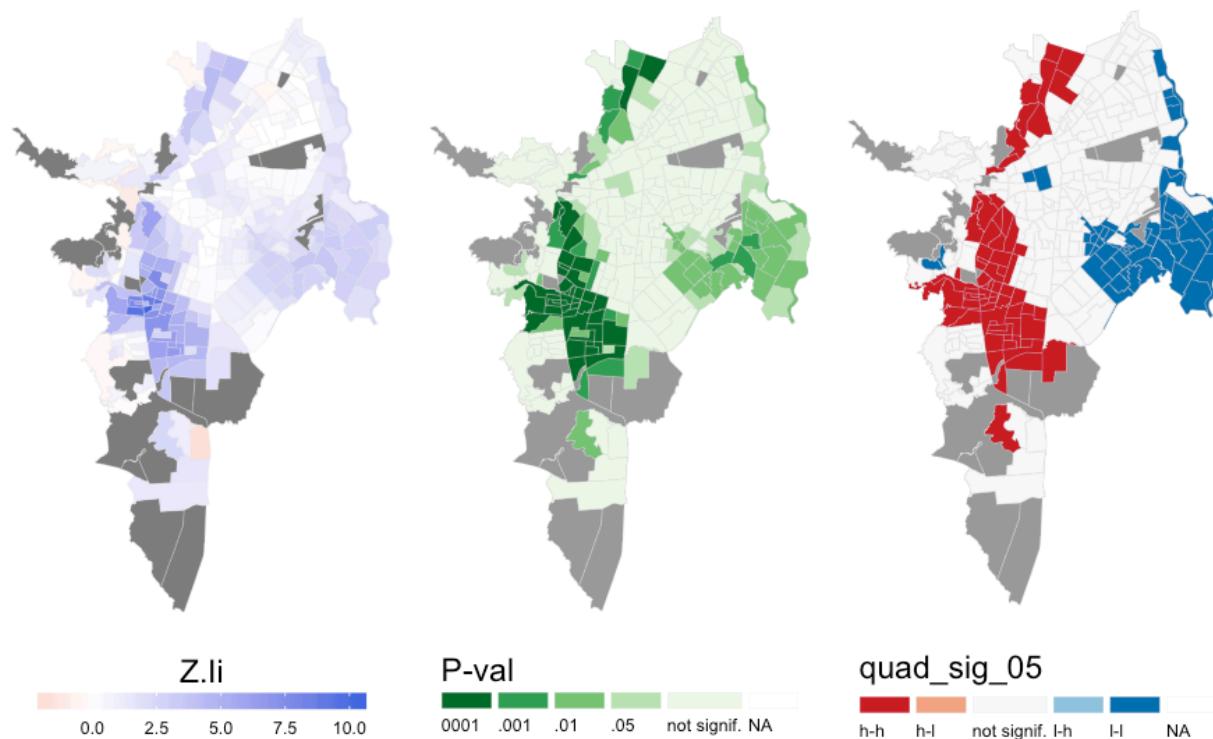


Figura 5.3: Mapas LISA - Porcentaje población con estudios superiores

## Mapas LISA - densidad\_poblacion - Wq

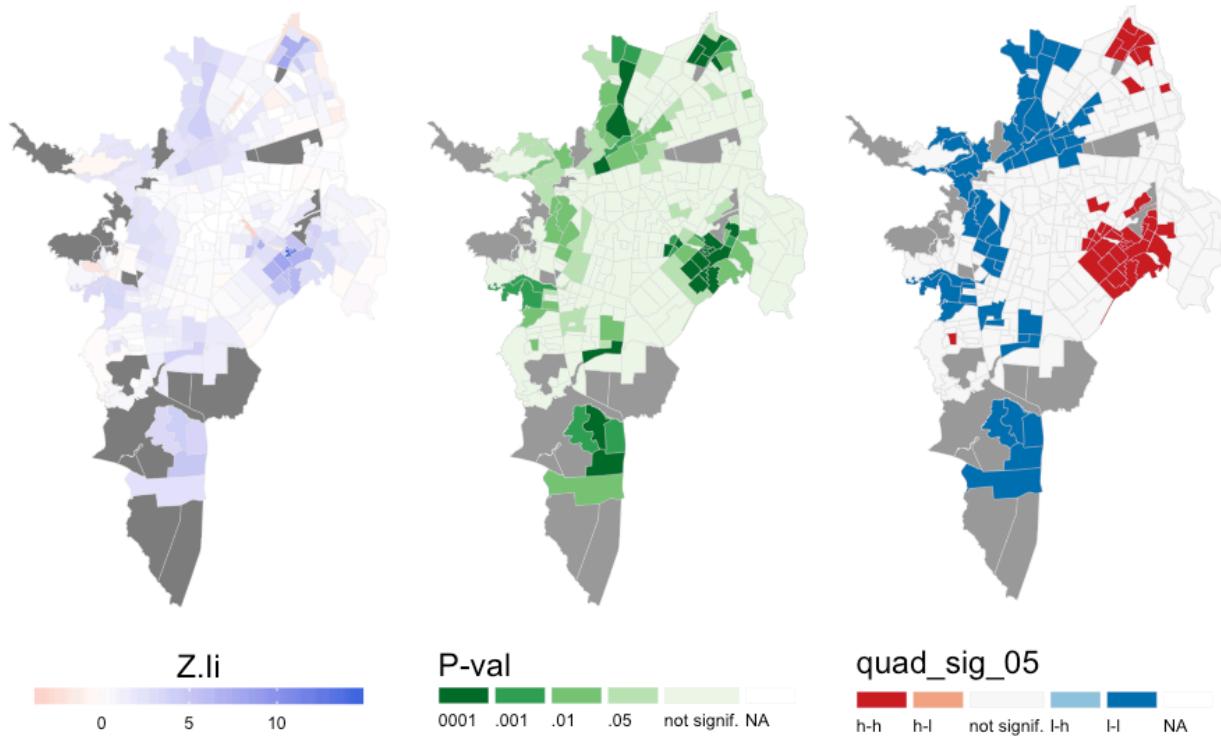


Figura 5.4: Mapas LISA - Densidad de Población

es significativo en ambos modelos, y tiene un coeficiente negativo para área de copa y positivo para el porcentaje de cobertura. Los patrones espaciales del área de copa y porcentaje no coinciden completamente, muestran dos patrones del mismo fenómeno que se ve alterado por la heterogeneidad en los tamaños de los sectores censales de un indicador absoluto a uno porcentual y que nos muestran concentraciones en zonas con cierto grado de complementariedad y continuidad con zonas de traslape de los sectores que los componen (figuras 4.17 y 4.20). Quizá esta situación sea la causa de la inconsistencia en el coeficientes de la densidad de población que toca ambos patrones de la cobertura pero no de forma simétrica.

En el modelo SD de las variables retardadas ninguna es significativa. El porcentaje de viviendas tipo cuarto no es significativo pues solo aporta como predictor para la zona del centro de la ciudad (ver figura 5.5) que ya es cubierto por la distribución espacial de la

### Mapas LISA - cuarto.porcentaje - Wq

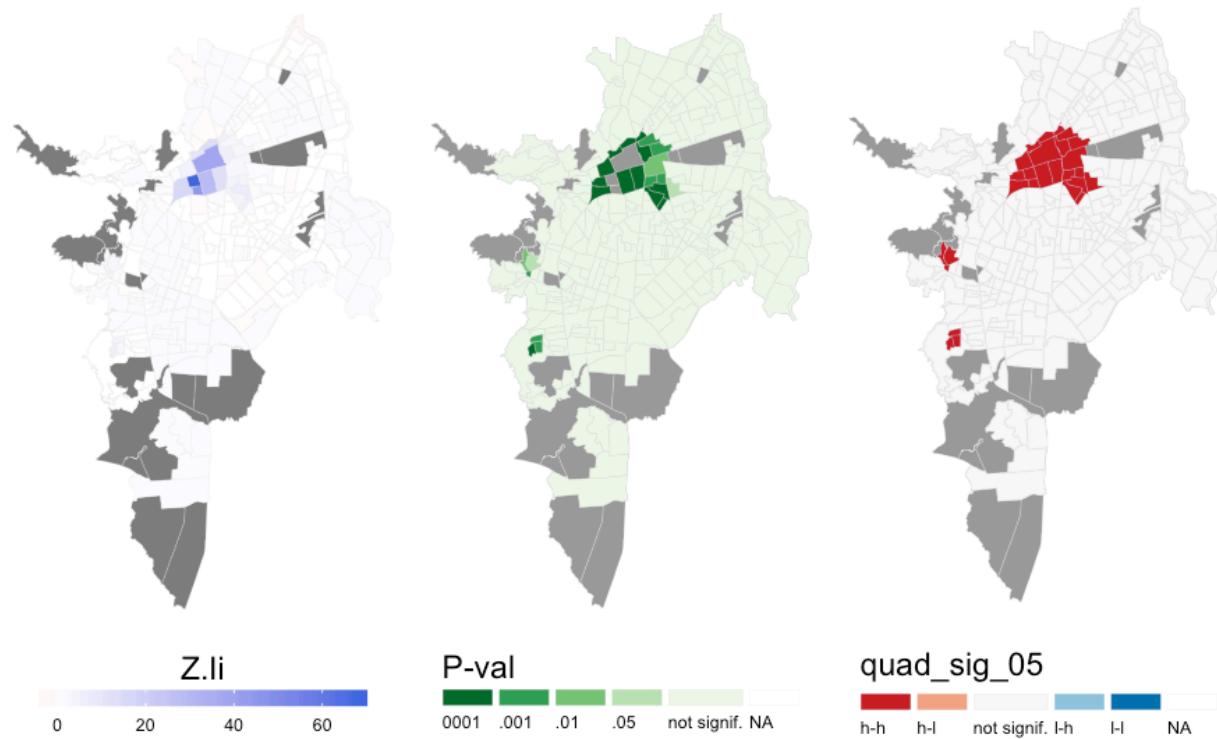


Figura 5.5: Mapas LISA - Porcentaje de viviendas tipo cuarto

variable sobre los estudios superiores. Es significativo el porcentaje de área de espacio público, con el segundo coeficiente más elevado explicando de forma consistente el gran número de árboles que se encuentran en estos lugares, pero que explica la cobertura en puntos muy específicos de la ciudad (figura 4.45). No es probable que tengan un efecto de difusión o derrame pues la creación de estas zonas verdes es algo que ocurre en el proceso de urbanización del sector y poco o nada cambia después de su creación.

Entre los posibles escenarios predichos en relación con la hipótesis planteada en este trabajo se encontró que efectivamente el patrón espacial de acceso a la educación de la población se correlaciona con el acceso a servicios ambientales del AU representado por el por el número de personas (y el porcentaje por SU) con estudios de educación superior, que en Colombia no es de acceso universal y la oferta educacional está cubierta principalmente por instituciones privadas. Es además fuerte la asociación negativa entre este indicador

de estatus social en zonas con prevalencia de población afro y población sin estudios, que además de la alta segregación laboral y geográfica que exhiben (Arroyo Mina et al., 2016; Mora and Arcila, 2014; Cerón and Escobar, 2014; Pacheco, 2013) está menos beneficiada de los servicios ambientales provisto por el arbolado urbano. Estos hallazgos coinciden con trabajos sobre ciudades en Estados Unidos como en (Schwarz et al., 2015) donde se prueba que la distribución espacial de árboles está sesgada por la distribución del ingreso, que a su vez también está relacionada con patrones de segregación racial.

Este tipo de desigualdades son una responsabilidad de los urbanizadores y de los gobiernos locales. Es de su interés y responsabilidad dialogar para construir y propiciar espacios para la siembra y desarrollo de los individuos arbóreos, pues es fundamental su agencia para reducir brechas sociales, crear una ciudad con espacio e infraestructura natural que sostenga los ecosistemas de los que depende la ciudad y la calidad de vida de sus ciudadanos. Ciudadanos para quienes lo ambiental viene cobrando más importancia como valor social, y que además se convierte en una estrategia para mitigar los cambios que pueda traer consigo el calentamiento global.

## 5.2. Sobre los espacios verdes

Análogamente a la cobertura arbórea, las métricas de acceso son aspectos complementarios sobre el mismo fenómeno de acceso aún beneficio; una representa un beneficio local, para el que vive en ese sector (`area_ep.porcentaje`) y la otra una forma (`ia.areas.dist`) que involucra un grupos de sectores beneficiarios, y que mide la relación entre el área del espacio verde y la distancia desde el centroide del sector urbano para luego sumar los valores de los espacios en un rango de distancia: intenta cuantificar la relación de costo de traslado por beneficio en área. El modelo de mejor rendimiento en la estimación medido usando el AIC fue el SEM para `area_ep.porcentaje` y SD para `ia.areas.dist`. Los coeficientes de los dos modelos ajustados se detallan en la tablas 5.5 y 5.6 para el modelo SEM y 5.7 y 5.8 para el modelo SD.

Cuadro 5.5: Coeficientes del modelo SEM de porcentaje de área de espacio verde

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.049	0.007	7.116	0.000
cuarto.porcentaje.mxn	-0.098	0.042	-2.355	0.019
area_media_manzana.mxn	0.771	0.065	11.803	0.000

Cuadro 5.6: Coeficiente de autocorrelación modelo SEM de porcentaje de área de espacio verde

\$\lambda\$	Likelihood ratio	p-value
0.259	10.215	0.001

En el SEM todos los coeficientes son significativos y el valor de área media de manzana es la variable más influyente. El coeficiente de autocorrelación en el error es muy significativo, lo que sugiere que si existe información de patrones espaciales en el error pero en variables no modeladas. La importancia que tiene en el modelo el área media de manzana puede interpretarse como una característica estructural de barrio o sector que por tener manzanas más grandes las zonas destinadas para parque o espacio verde son en consecuencia más grandes y el beneficio mayor o/y que algunas manzanas que albergan equipamientos de ciudad, zonas verdes como riberas de ríos o lagunas determinan el alto valor del indicador al ser el promedio sensible a valores extremos. Lo cierto es que desde este punto de vista local no existe evidencia sobre una relación concluyente con ninguna variable poblacional, por lo que no puede decirse que se vulnere a ciertas comunidades o grupos diferenciados, al menos en lo que respecta a las variables que proveen los sistemas de consulta del censo del 2005. Sin embargo puede afirmarse que la distribución de los EV es desigual pues más de 203 de 329 sectores en la ciudad tiene menos del 5 % de espacio verde de área respecto del area total del sector (figura 5.7).

El indicador de acceso area-distancia fue mejor ajustado por el modelo SD, que tuvo como coeficientes significativos a `unidad_economica.porcentaje` con un valor bajo pero positivo al igual que `con_alguna_limitacion.porcentaje` que tuvo un valor bajo pero negativo y significativo. La zona que muestra el mapa de LISA el porcentaje de unidades

Cuadro 5.7: Coeficientes del modelo SD del indicador de acceso área-distancia

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.018	0.018	-1.010	0.313
area_media_manzana.mxn	0.029	0.050	0.579	0.563
cuarto.porcentaje.mxn	-0.003	0.043	-0.068	0.946
unidad_economica.porcentaje.mxn	0.067	0.028	2.377	0.017
densidad_poblacion.mxn	-0.005	0.022	-0.233	0.816
con_alguna_limitacion.porcentaje.mxn	-0.043	0.023	-1.849	0.065
ningun_estudio.porcentaje.mxn	0.035	0.041	0.837	0.403
lag.area_media_manzana.mxn	0.738	0.105	7.007	0.000
lag.cuarto.porcentaje.mxn	-0.089	0.070	-1.273	0.203
lag.unidad_economica.porcentaje.mxn	0.012	0.043	0.271	0.786
lag.densidad_poblacion.mxn	0.023	0.037	0.622	0.534
lag.con_alguna_limitacion.porcentaje.mxn	-0.001	0.043	-0.013	0.990
lag.ningun_estudio.porcentaje.mxn	0.084	0.061	1.383	0.167

Cuadro 5.8: Coeficiente de autocorrelación modelo SD del indicador de acceso área-distancia

\$\rho\$	Likelihood ratio	p-value
0.709	213.394	0

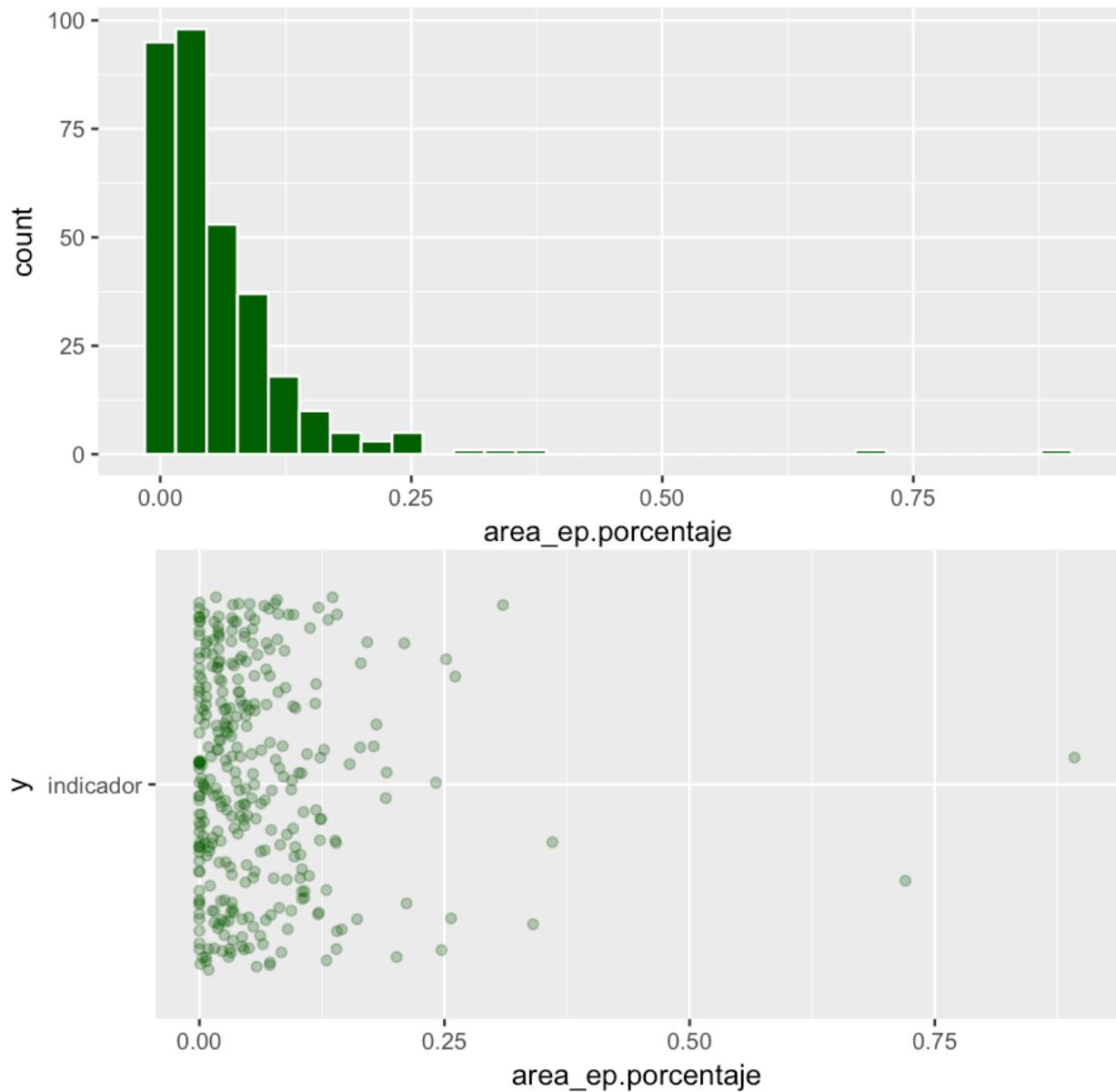


Figura 5.6: Distribucion de los indicadores de acceso local a EV

económicas en el sector (figura 5.9) coinciden de nuevo con equipamientos de ciudad como el Boulevard del Río mientras que las concentraciones de personas con limitación altas coinciden con poco espacio verde como bien indica el signo del coeficiente. La variable de área media de manzana retardada `lag.area_media_manzana` es significativa y es la que tiene el valor más alto entre los coeficientes, confirmando el efecto de derrame de los grandes espacio verdes que son parte del equipamiento de ciudad y que producen beneficios en un radio más amplio que el área de sector donde el disfrute exige que la gente se desplace, justo como lo propone el indicador de acceso. El coeficiente de autocorrelación espacial

*rho*. es muy alto y significativo, lo que indica que la introducción de autocorrelación espacial en la variable dependiente mejora el ajuste pues es evidente la existencia de grupos de sectores con valores similares.

A pesar de que se mantienen problemas como no normalidades en los errores, el ajuste de la variable dependiente presenta mejoras en las métricas de error y hace más confiables la estimación de los coeficientes. En cuanto al acceso a espacio verdes que involucran desplazamientos o una cobertura a grupos de sectores urbanos no puede decirse de forma irrefutable que está estrechamente relacionado con desventajas en la población como el estatus o la etnicidad, pero si marca que hay patrones espaciales en la población de limitaciones físicas o de salud que coinciden con niveles pobres de acceso.

La estrategia de construir equipamientos con espacios verdes para mejorar el acceso puede verse optimizada de este tipo de análisis, pues resalta las zonas donde el impacto de estas obras puede llegar a más personas si se ubican correctamente. Marca también la responsabilidad de los urbanizadores y las autoridades para generar proyectos de renovación urbana que integren y den relevancia al acceso a EV de manera razonada. Estos resultados apuntan a tener en cuenta aspectos de la planeación urbana al nivel del tamaño de las manzanas y la existencia de andenes que permitan alojar parques de mayor tamaño y espacio para que se desarrolle el potencial recreativo de esos espacios, así como las especies arbóreas.

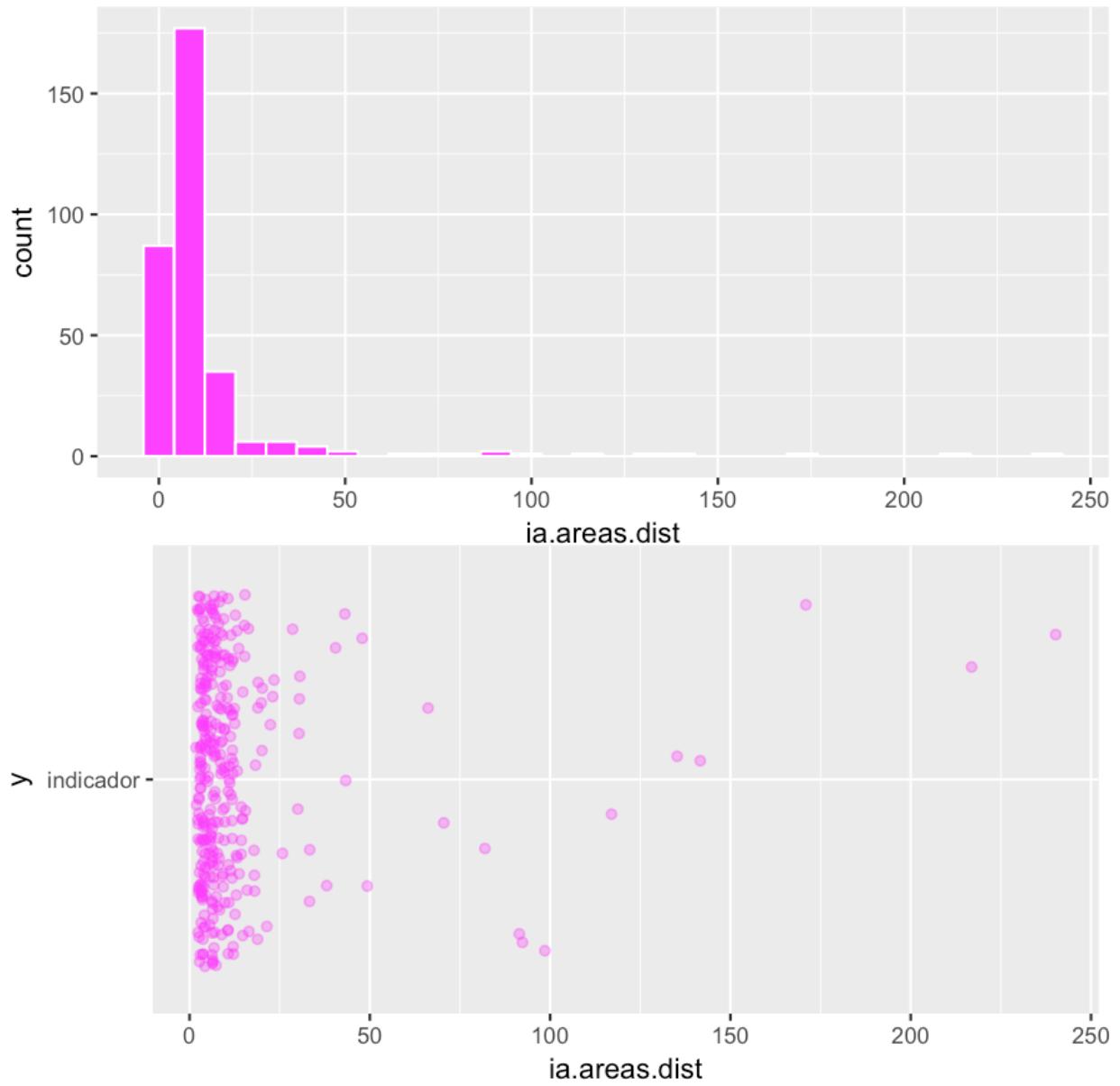


Figura 5.7: Distribucion de los indicadores de acceso a EV área-distancia

### Mapas LISA - area\_ep.porcentaje - Wq

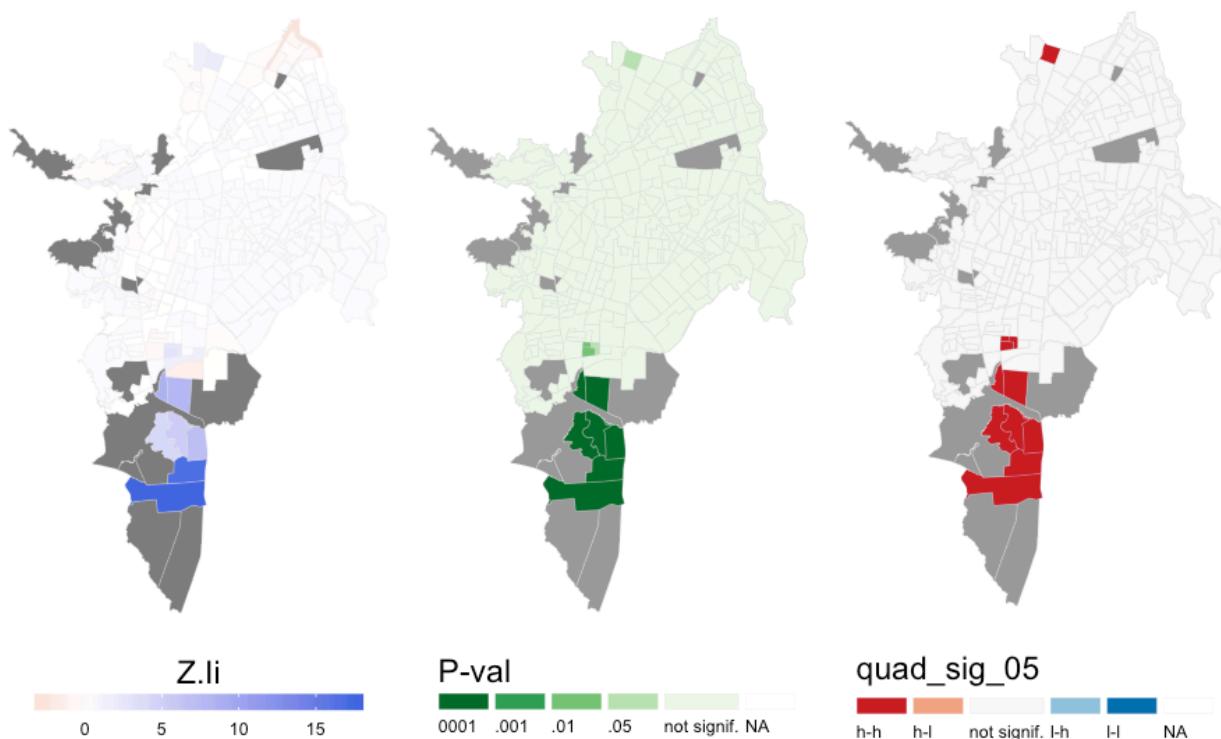
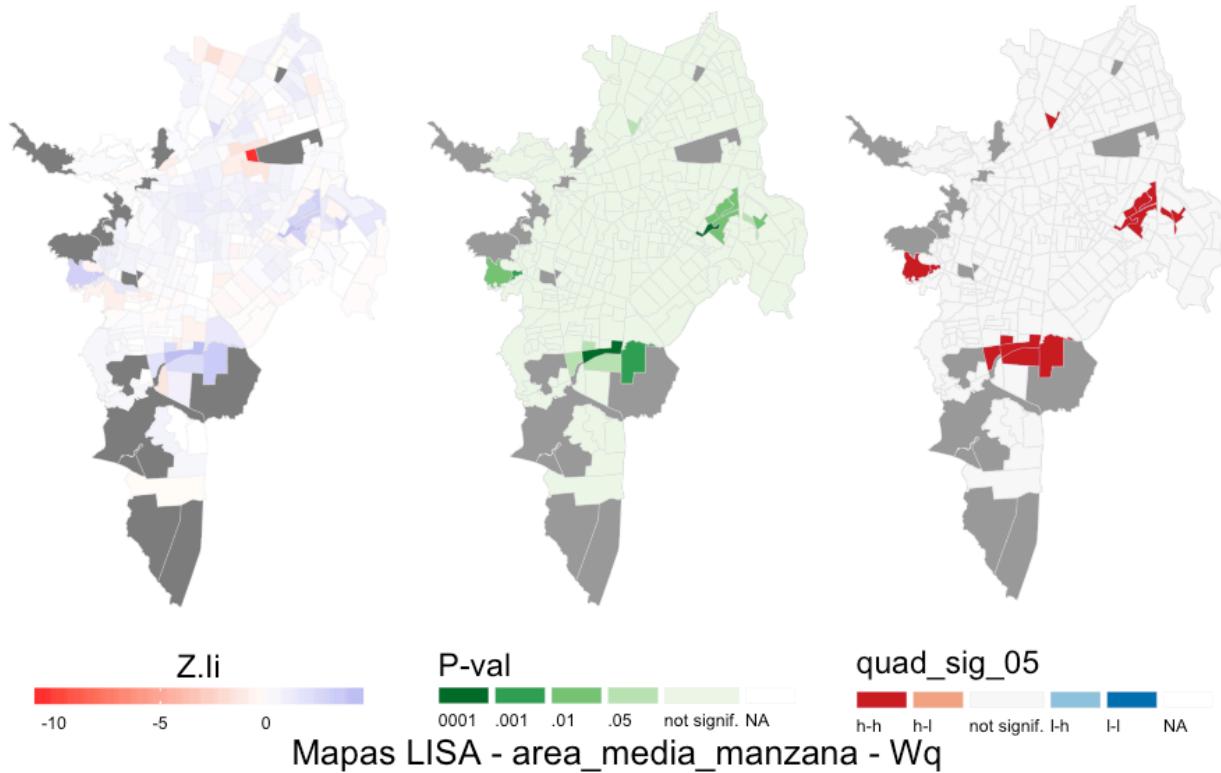


Figura 5.8: Mapas LISA - Porcentaje de espacios verdes y área media de manzana <sup>195</sup>

Mapas LISA - unidad\_economica.porcentaje - Wq

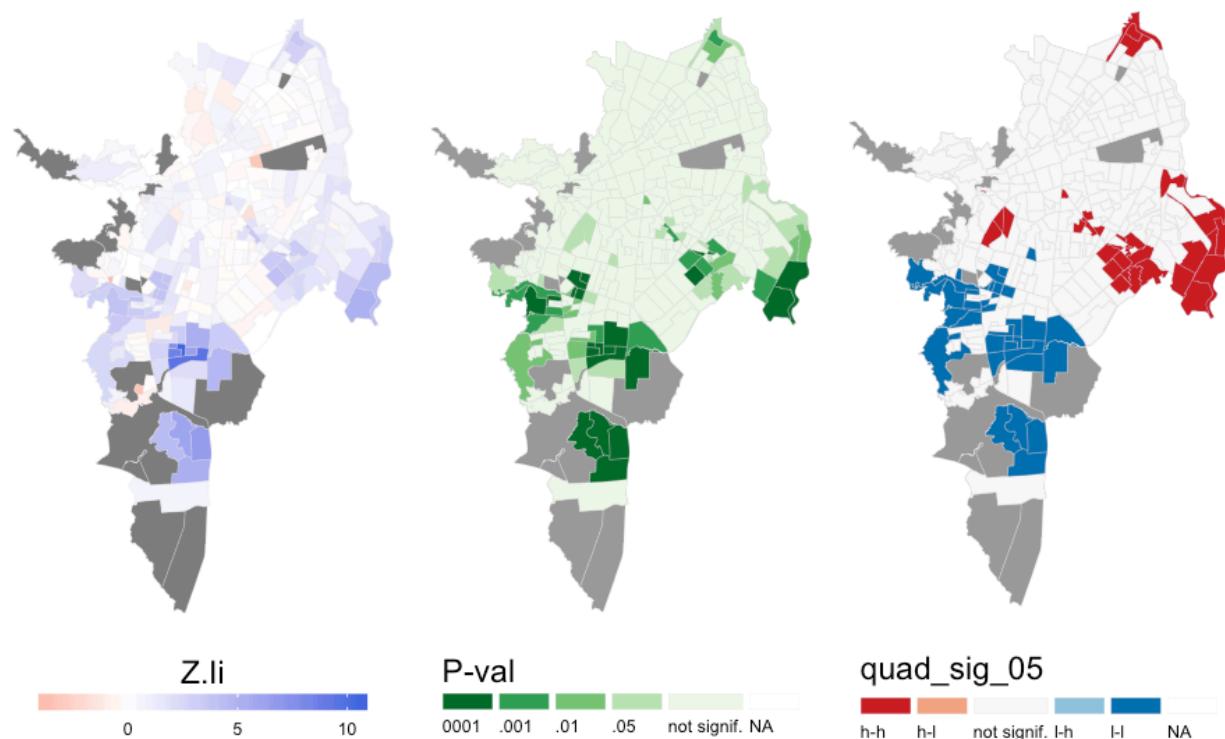
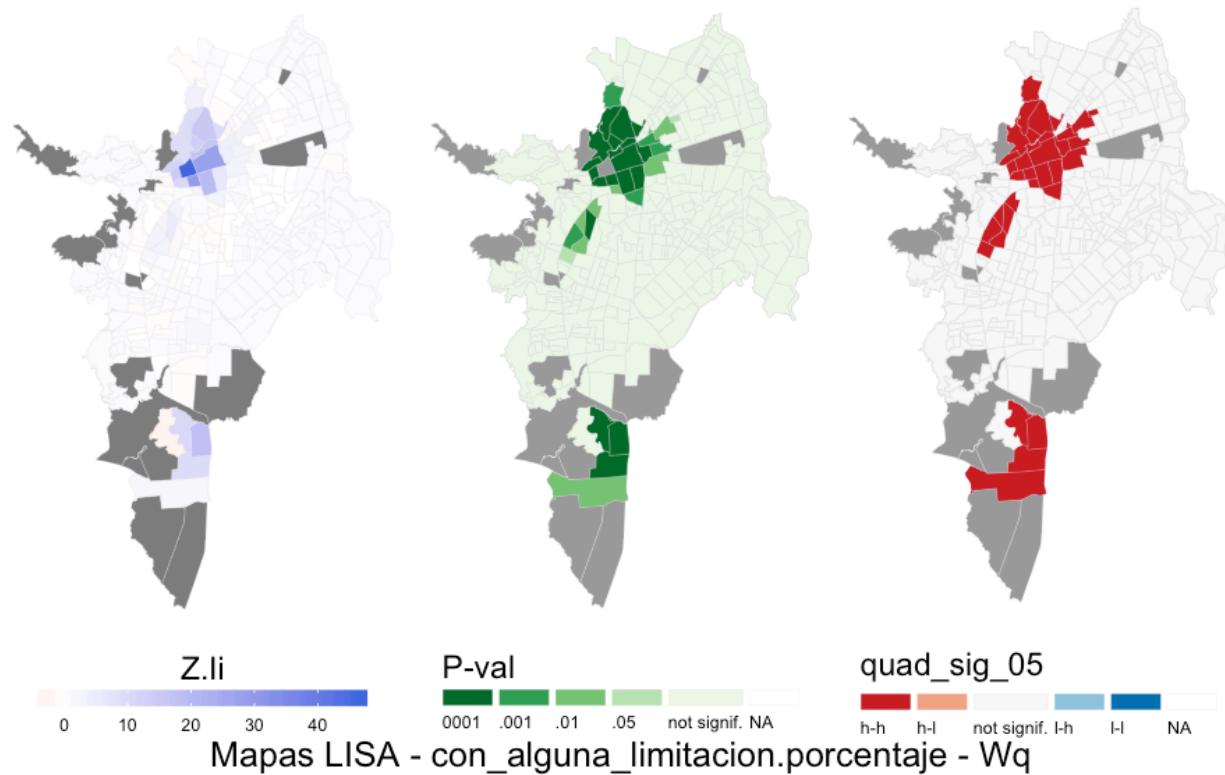


Figura 5.9: Mapas LISA - Porcentaje de unidades economicas y personas con alguna limitación <sup>196</sup>

# **Capítulo 6**

## **Conclusiones**

Quiero resaltar el uso de herramientas libres como *R*, Rstudio y los diferentes paquetes del ecosistema del CRAN, que favorecen la reproductibilidad y diseminación del conocimiento científico. El esfuerzo y la curva de aprendizaje de herramientas como Bookdown [R-bookdown] y knitr (Xie, 2018) para la generación de los contenidos y textos en diferentes formatos, la reproductibilidad del uso de scripts y la posibilidad de publicar un repositorio del resultado, del proceso y de los datos, no solo vale la pena como base para la construcción de flujos de trabajo eficientes en el manejo y divulgación de datos e información —geográfica—, sino que me parece un compromiso ineludible con la construcción de la ciencia de nuestro tiempo y de la ética investigativa.

En cuanto a lo metodológico en este trabajo se hizo esfuerzos por conjugar la exposición gráfica de los datos crudos y de los resultados estadísticos con la rigurosidad y capacidad explicativa de los modelos de regresión estadísticos y geoestadísticos. Entre las posibilidades que ofrece el código como herramienta cartográfica está la de producir y reproducir múltiples mapas, en lugar de uno solo con mucho detalle como respuestas a las limitaciones tecnológicas y de costos de la cartografía impresa, lo que redunda en mejor información para los análisis. Los métodos gráficos como los mapas de LISA permiten una interpretación más clara de los mecanismos de ajuste de los modelos espaciales y de la combinación

lineal de los términos. Estos mapas junto con los mapas temáticos de cada una de las variables, los gráficos de dispersión de las distribuciones multivariadas son una estrategia eficaz para lidiar con las dificultades para analizar conjuntos de datos multidimensionales y espaciales. Los métodos de ajuste espacial y la comparación entre distintos modelos permitieron seleccionar el resultado más confiable y eficiente de la estimación, dando firmeza a las conclusiones de los análisis.

En relación a lo temático este trabajo demuestra que en lo que respecta al acceso de beneficios del arbolado urbano en Santiago de Cali existe una fuerte relación de la variable de estatus social representada por el acceso a estudios superiores, que explica la variabilidad en los datos y los patrones espaciales de la cobertura de copa. Los modelos de regresión espacial confirman que los beneficios ambientales son mejores en sectores con población mejor educada, posiblemente una razón para preferir o habitar espacios con buena arborización o tal vez con condiciones suficientes de espacio y zonas blandas para arborizar el barrio y propiciar el desarrollo del potencial de los individuos arbóreos. En cualquier caso, esto muestra que los clusters de sectores urbanos con mayor arborización excluyen al grueso de la población afro y con bajos niveles de estudios, como prueban las altas correlaciones negativas entre los valores de estos grupos de población y la cobertura de copa. El mapa de los beneficios ambientales se suma pues a una serie de inequidades encontradas en la literatura que aborda problemas relacionados con la segregación racial y la baja empleabilidad de poblaciones afrocolombianas en la ciudad de Cali. Los casos de análisis espacial en relación al acceso de servicios ambientales fueron descritos ampliamente en la literatura y nos muestran una preocupación creciente sobre la integración y restauración de ecosistemas urbanos, una sofisticación metodológica y fundamentación teórica que fueron claves para aplicar exitosamente el análisis al contexto de Cali. Además coinciden en señalar problemas de segregación e inequidades ambientales en el continente americano.

En cuanto a los espacios verdes el ejercicio no es mucho más alentador. Este trabajo mostró cómo existe una importante cantidad de sectores urbanos con menos del 5 % del área del sector de espacio verde: una alta concentración del área disponible se encuentra unos pocos

sectores, que sirven para suplir las carencias de espacio verde en sus vecinos. Aunque el coeficiente de la dimensión poblacional representada por porcentaje de personas con alguna limitación física o mental, es bajo, coincide espacialmente con las zonas de bajo acceso a estos espacios. Sin embargo en la medición del acceso como un beneficio local las variables poblacionales no pasaron los criterios de selección para ser candidatas en el modelo de regresión. El área media de manzana de un sector fue muy representativa y significativa de la variabilidad de los datos, y señala indirectamente la responsabilidad de las autoridades y los urbanizadores que deben garantizar la provisión de los equipamientos de ciudad, los tamaños mínimos y deseados, y en general velar por el desarrollo la estructura ecológica del municipio y del casco urbano.

Analizar varios indicadores con variaciones en el cálculo y en las variables incorporadas en su definición permite construir una visión más compleja y detallada del acceso. Este trabajo se limitó a analizar 2 indicadores por beneficio ambiental, explorando las diferencia entre indicadores porcentuales que representan las diferencias entre medir el beneficio en el área total de un sector o sólo sobre su espacio público, para establecer una matiz entre lo que compete a las autoridades y el alcance de un beneficio que se irradia más allá de las fronteras de las calles y los límite administrativos. También explora dos experiencias distintas del acceso a espacio verdes, la local, al interior de un sector urbano y otra en la escala de las agrupaciones de sectores. Resultado de esta se puede pensar explorar complejizar estos indicadores para incorporar aspectos de la calidad de los espacio o incluir dimensiones poblacionales como la densidad o normalizarlos con respecto del número de habitantes.

Más interesante puede ser contar con dos conjuntos de datos separados en el tiempo tanto de las variables poblacionales y del arbolado urbano, para indagar sobre cómo la matriz de vecindad puede capturar las variaciones de la estructura de las características poblacionales de los individuos humanos y arbóreos; o analizar el impacto de la inclusión de nuevos espacio verdes a la topología de beneficios y estimar cambios en la estructura de grupos de sectores urbanos con acumulacion de desvenajas, por ejemplo.

Esto me lleva al último punto de las conclusiones y el primer momento dentro del proceso de esta investigación: la disponibilidad de los datos. Sin la existencias de los servicios de información geográfica de la Alcaldía de Cali ni la disposición de la cartografía del censo y los datos agregados a nivel de sector urbano del DANE, o la existencia de un censo arbóreo es imposible progresar en el entendimiento de nuestro desarrollo como ciudad, región o nación. Sin embargo, aún son escasas las variables del censo de población que están disponibles a este nivel, no existe un repositorio público de los datos del censo arbóreo, ni de las acciones de mantenimiento y estado del arbolado urbano a pesar de que no existe ningún impedimento técnico ni legal para que pueda estarlo. Los resultados dependen pues no solo de la disponibilidad sino de la calidad y nivel de actualización de los datos, que pueden y deben mejorar en los servicios de información que presta el el municipio.

# Bibliografía

Geoportal DANE.

Geoportal IDESC.

MAGNA-SIRGAS-CALI.

Plan de Ordenamiento Territorial - POT año 2014.

Sistema de consulta censo 2005.

Alanís, E., Jiménez, J., MORA, O., Canizales, P., and Rocha, L. (2014). Estructura y composición del arbolado urbano de un campus universitario del noroeste de México. *Revista Iberoamericana de Ciencias*, 1(7):93–101.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics*, 27(3):247–267.

Arroyo Mina, J. S., Pinzón Gutiérrez, L. F., Mora, J. J., Gómez Jaramillo, D. A., and Cendales, A. (2016). Afrocolombianos, discriminación y segregación espacial de la calidad del empleo para Cali. *Cuadernos de Economía*, 35(69):753–783.

Azócar, G., Romero, H., Sanhueza, R., Vega, C., Aguayo, M., and Muñoz, M. D. (2007). Urbanization patterns and their impacts on social restructuring of urban space in Chilean mid-cities: The case of Los Angeles, Central Chile. *Land Use Policy*, 24(1):199–211.

Bivand, R. (2017). *spdep: Spatial Dependence: Weighting Schemes, Statistics and Models*. R package version 0.7-4.

Bivand, R., Keitt, T., and Rowlingson, B. (2017). *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. R package version 1.2-16.

Bivand, R. and Rundel, C. (2017). *rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. R package version 0.3-26.

Bolund, P. and Hunhammar, S. (1999). Ecosystem services in urban areas. *Ecological economics*, 29(2):293–301.

Boone, C. G., Cadenasso, M. L., Grove, J. M., Schwarz, K., and Buckley, G. L. (2010). Landscape, vegetation characteristics, and group identity in an urban and suburban watershed: why the 60s matter. *Urban Ecosystems*, 13(3):255–271.

Braverman, I. (2008). Everybody loves trees: Policing American cities through street trees. *Duke Envtl. L. & Pol'y F.*, 19:81.

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, pages 1287–1294.

CEPAL. REDATAM.

Cerdà, M. O. (2011). Origen y evolución del movimiento de justicia ambiental. *Ecología política*, (41):17–24.

Cerón, W. L. and Escobar, Y. C. (2014). Índice de segregación espacial y socioeconómico (ises) en las comunas de Santiago de Cali. *Cuadernos de Vivienda y Urbanismo*, 7(13).

Chakraborty, J. and Armstrong, M. P. (1997). Exploring the use of buffer analysis for the identification of impacted areas in environmental equity assessment. *Cartography and Geographic Information Systems*, 24(3):145–157.

Chapman, C. A. and Onderdonk, D. A. (1998). Forests without primates: primate/plant codependency. *American Journal of Primatology*, 45(1):127–141.

CIAT (2015a). Plan integral de adaptación y mitigación al cambio climático.

CIAT (2015b). Identificación de Zonas y Formulación de Propuestas para el Tratamiento de Islas de Calor.

Comber, A., Brunsdon, C., and Green, E. (2008). Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious groups. *Landscape and Urban Planning*, 86(1):103–114.

Comber, A. J., Brunsdon, C., and Radburn, R. (2011). A spatial analysis of variations in health access: linking geography, socio-economic status and access perceptions. *International Journal of Health Geographics*, 10:44.

Cowett, F. D. (2014). Methodology for Spatial Analysis of Municipal Street Tree Benefits. *Arboriculture & Urban Forestry*, 40(2).

Cutter, S. L., Holm, D., and Clark, L. (1996). The role of geographic scale in monitoring environmental justice. *Risk Analysis*, 16(4):517–526.

DANE (2005). R+::CEPAL/CELADE - R+SP WebServer.

Durán Rivera, B. and Alzate Guarín, F. (2009). Intercepción de partículas suspendidas totales (PST) por cinco especies de árboles urbanos en el Valle de Aburrá.

Escobar, G. (2015). Cali en cifras 2015. *Alcaldía de Santiago de Cali, Departamento Administrativo de Planeación*, 26.

Ferro Medina, G. (2010). *Árboles ciudadanos, en la memoria y en el paisaje cultural de Bogotá*. Bogotá: Instituto Distrital de Patrimonio Cultural.

Fotheringham, A. S., Charlton, M. E., and Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and planning A*, 30(11):1905–1927.

Garzón, B., Brañes, N., Abella, M. L., and Auad, A. (2004). Vegetación urbana y hábitat popular: el caso de san miguel de tucumán. *Revista invi*, 18(49).

- Gibbons, S. and Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, 52(2):172–191.
- Gómez-Baggethun, E. and Barton, D. N. (2013). Classifying and valuing ecosystem services for urban planning. *Ecological Economics*, 86:235–245.
- Heynen, N., Perkins, H. A., and Roy, P. (2006). The political ecology of uneven urban green space the impact of political economy on race and ethnicity in producing environmental inequality in Milwaukee. *Urban Affairs Review*, 42(1):3–25.
- Kabisch, N. and Haase, D. (2014). Green justice or just green? Provision of urban green spaces in Berlin, Germany. *Landscape and Urban Planning*, 122:129–139.
- Killicoat, P., Puzio, E., and Stringer, R. (2002). The economic value of trees in urban areas: estimating the benefits of Adelaide’s street trees. *TREENET*, page 90.
- Kissling, W. D. and Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, 17(1):59–71.
- Konijnendijk, C., Gauthier, M., and Van Veenhuizen, R. (2005). Árboles y Ciudades Creciendo Juntos. *Revista Agricultura Urbana*, 13:1–7.
- Landry, S. M. and Chakraborty, J. (2009). Street trees and equity: evaluating the spatial distribution of an urban amenity. *Environment and Planning A*, 41(11):2651–2670.
- Laredo, D. R. and Mirtha, D. (2011). La gestión del verde urbano como un criterio de mitigación y adaptación al cambio climático. *Revista de la Asociación Argentina de Ecología de Paisajes*, 2(2):123–130.
- Leff, E. (2012). Pensamiento ambiental latinoamericano: patrimonio de un saber para la sustentabilidad. *Environmental Ethics*, 34(Supplement):97–112.
- LeSage, J. P. and Pace, R. K. (2014). The Biggest Myth in Spatial Econometrics. *Econometrics*, 2(4):217–249.

- León Calle, S. and others (2011). *Árboles, simbolismo, cultura, memoria e identidad. Representaciones en el paisaje arbóreo de Gualaquiza*. PhD thesis, Universidad Andina Simón Bolívar, Sede Ecuador.
- Low, S. (2013). Public space and diversity: Distributive, procedural and interactional justice for parks. *The Ashgate Research companion to planning and culture*. Surrey, UK: Ashgate Publishing, pages 295–310.
- Martinez Alier, J., Anguelovski, I., Bond, P., Del Bene, D., Demaria, F., Gerber, J.-F., Greyl, L., Haas, W., Healy, H., Marín-Burgos, V., and others (2014). Between activism and science. *Journal of Political Ecology*, 21:19–60.
- McPherson, E. G., Nowak, D., Heisler, G., Grimmond, S., Souch, C., Grant, R., and Rounsevell, R. (1997). Quantifying urban forest structure, function, and value: the Chicago Urban Forest Climate Project. *Urban ecosystems*, 1(1):49–61.
- McPherson, E. G., Xiao, Q., and Aguaron, E. (2013). A new approach to quantify and map carbon stored, sequestered and emissions avoided by urban forests. *Landscape and Urban Planning*, 120:70–84.
- Mora, J. J. and Arcila, A. M. (2014). Brechas salariales por etnia y ubicación geográfica en Santiago de Cali. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 18:34–53.
- Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23.
- Nelson, E., Mendoza, G., Regetz, J., Polasky, S., Tallis, H., Cameron, D., Chan, K., Daily, G. C., Goldstein, J., Kareiva, P. M., and others (2009). Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. *Frontiers in Ecology and the Environment*, 7(1):4–11.
- Nesbitt, L. and Meitner, M. J. (2016). Exploring Relationships between Socioeconomic Background and Urban Greenery in Portland, OR. *Forests*, 7(8):162.

- Nolazco, S. (2012). Diversidad de aves silvestres y correlaciones con la cobertura vegetal en parques y jardines de la ciudad de Lima. *Boletín Informativo UNOP*, 7(1):4–16.
- Nowak, D. J. and Crane, D. E. (2000). The Urban Forest Effects (UFORE) Model: quantifying urban forest structure and functions.
- Nowak, D. J. and Crane, D. E. (2002). Carbon storage and sequestration by urban trees in the USA. *Environmental pollution*, 116(3):381–389.
- Nowak, D. J. and Greenfield, E. J. (2012). Tree and impervious cover change in US cities. *Urban Forestry & Urban Greening*, 11(1):21–30.
- Nowak, D. J., Stein, S. M., Randler, P. B., Greenfield, E. J., Comas, S. J., Carr, M. A., and Alig, R. J. (2010). Sustaining America's urban trees and forests: a Forests on the Edge report.
- Osorio, J. and Molina, L. (2009). A vuelo de pájaro Las ciudades como refugio para las aves. *Revista nodo*, 4(7).
- Pacheco, H. V. (2013). Persistencia de la segregación residencial y composición del capital humano por barrios en la ciudad de cali. *Ensayos sobre Política Económica*, 31(70):121 – 155.
- Pebesma, E. and Bivand, R. (2018). *sp: Classes and Methods for Spatial Data*. R package version 1.2-6.
- Phelps, J. B. (2012). *The association between tree canopy cover and socio-demographics in Lubbock, Texas*. PhD thesis, Texas Tech University.
- Páez, A. and Scott, D. M. (2005). Spatial statistics for urban analysis: a review of techniques with examples. *GeoJournal*, 61(1):53–67.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Restrepo, O., Iván, H., Moreno, H., Hoyos, E., Helena, C., and others (2015). Incidence of progressive deterioration of urban trees in the Aburrá valley, Colombia. *Colombia Forestal*, 18(2):225–240.
- Ripoll, M. V., Kurbán, A., Papparelli, A., Cúnsulo, M., and Roca, G. (2010). Condiciones térmicas de un espacio verde urbano en clima árido. *Avances en Energías Renovables y Medio Ambiente*, 14:11–09.
- Schlosberg, D. (2013). Theorising environmental justice: the expanding sphere of a discourse. *Environmental Politics*, 22(1):37–55.
- Schwarz, K., Frakias, M., Boone, C. G., Zhou, W., McHale, M., Grove, J. M., O’Neil-Dunne, J., McFadden, J. P., Buckley, G. L., Childers, D., Ogden, L., Pincetl, S., Pataki, D., Whitmer, A., and Cadenasso, M. L. (2015). Trees Grow on Money: Urban Tree Canopy Cover and Environmental Justice. *PLoS ONE*, 10(4).
- Shanahan, D. F., Lin, B. B., Gaston, K. J., Bush, R., and Fuller, R. A. (2014). Socio-economic inequalities in access to nature on public and private lands: a case study from Brisbane, Australia. *Landscape and Urban Planning*, 130:14–23.
- Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.
- Talen, E. and Anselin, L. (1998). Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds. *Environment and Planning a*, 30(4):595–613.
- Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic geography*, 46(sup1):234–240.
- Tratalos, J., Fuller, R. A., Warren, P. H., Davies, R. G., and Gaston, K. J. (2007). Urban form, biodiversity potential and ecosystem services. *Landscape and urban planning*, 83(4):308–317.

Troy, A. R., Grove, J. M., O'Neil-Dunne, J. P., Pickett, S. T., and Cadenasso, M. L. (2007). Predicting opportunities for greening and patterns of vegetation on private urban lands. *Environmental management*, 40(3):394–412.

Vásquez, . (1990). Historia del desarrollo económico y urbano en Cali". *Boletín socioeconómico*, 20:1–28.

Vásquez Fuentes, A. and Romero Aravena, H. (2008). Vegetación urbana y desigualdades socio-económicas en la comuna de Peñalolen, Santiago de Chile. Una perspectiva de justicia ambiental.

Watkins, S. L., Mincey, S. K., Vogt, J., and Sweeney, S. P. (2016). Is Planting Equitable? An Examination of the Spatial Distribution of Nonprofit Urban Tree-Planting Programs by Canopy Cover, Income, Race, and Ethnicity. *Environment and Behavior*, page 0013916516636423.

Wikipedia (2017). Indicadores de asociación espacial — wikipedia, la enciclopedia libre. [Internet; descargado 29-septiembre-2017 ].

Wolch, J. R., Byrne, J., and Newell, J. P. (2014). Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough'. *Landscape and Urban Planning*, 125:234–244.

Xie, Y. (2018). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.20.

Zhou, X. and Kim, J. (2013). Social disparities in tree canopy and park accessibility: A case study of six cities in Illinois using GIS and remote sensing. *Urban forestry & urban greening*, 12(1):88–97.