Autocorrelación y Modelos Espaciales: Hipótesis, Conectividad, Moran's I, LISA e Interpretación de Resultados

Andrés Guzman y Jhonier S. Rangel

1 Objetivo

Establecer, en un marco reproducible, (i) las hipótesis nula y alternativa para pruebas de autocorrelación espacial global (Moran's I) y local (LISA), (ii) la especificación de modelos de regresión espacial (SAR, SEM, SDM y SAC), (iii) el diseño y normalización de matrices de conectividad ($\mathbf{W}$), y (iv) lineamientos para interpretar resultados empíricos obtenidos con distintas vecindades (Queen, Rook, decaimiento por distancia).

2 Datos y notación

Sea $y \in \mathbb{R}^n$ la variable de interés por unidad espacial, $\mathbf{X} \in \mathbb{R}^{n\times p}$ las covariables y $\mathbf{W}$ una matriz de pesos espaciales (diagonal nula) normalizada por filas: $w_{ij} \geq 0$, $w_{ii} = 0$ y $\sum_j w_{ij} = 1$. Denote $\bar{y}$ el promedio de y y $z = y - \bar{y}\mathbf{1}$.

3 Matrices de conectividad (W)

Contigüidad por polígonos

- **Queen**: $w_{ij} = 1$ si las unidades i y j comparten borde o vértice; 0 en caso contrario.
- **Rook**: $w_{ij} = 1$ si i y j comparten borde; 0 en caso contrario.

Tras definir la vecindad binaria, se *row-estandariza*: $w_{ij} \leftarrow w_{ij}/\sum_j w_{ij}$.

Matrices basadas en distancia

Fije d_{ij} como distancia entre centroides. Ejemplos:

- **Inversa de la distancia**: $\tilde{w}_{ij} = \mathbb{I}(d_{ij} \leq r)\, d_{ij}^{-1}$, con umbral r (o k-vecinos más cercanos) y $w_{ij} = \tilde{w}_{ij}/\sum_j \tilde{w}_{ij}$.
- **Exponencial**: $\tilde{w}_{ij} = \mathbb{I}(d_{ij} \leq r)\, \exp(-\alpha d_{ij})$, normalizada por filas. El parámetro $\alpha > 0$ controla el decaimiento.

Los promedios de grado y la densidad de la red dependen de r (o k) y de α; se recomienda reporte conjunto (grado promedio, densidad, conectividad de la red).

4 Autocorrelación global: Moran's I

El estadístico de Moran global se define como

$$I = \frac{n}{\mathbf{1}^\top \mathbf{W} \mathbf{1}} \, \frac{z^\top \mathbf{W} z}{z^\top z}.$$

Hipótesis.

H_0 : "ausencia de autocorrelación espacial" (aleatoriedad espacial) vs $H_A : I \neq \mathbb{E}[I \mid H_0]$ (bilateral) o $H_A : I >$

La inferencia práctica se realiza con prueba de permutaciones (p.ej., 999 permutaciones), calculando p-valor empírico; alternativamente, aproximaciones asintóticas bajo supuestos de normalidad espacial.

5 Autocorrelación local: LISA (Moran local)

Para cada unidad i,

$$I_i = \frac{z_i}{m_2} \sum_j w_{ij} z_j, \quad m_2 = \frac{1}{n} \sum_i z_i^2.$$

Hipótesis por unidad.

$$H_{0,i} : \text{no autocorrelación local en } i \quad \text{vs} \quad H_{A,i} : I_i \neq 0.$$

Los p-valores se obtienen por permutaciones condicionadas en i. La *tipología* de clúster depende del cuadrante en el diagrama de dispersión espacial (z_i vs. $\sum_j w_{ij} z_j$):

- HH: alto rodeado de altos; LL: bajo rodeado de bajos;
- HL: alto rodeado de bajos; LH: bajo rodeado de altos.

Reportar mapas de significancia (p. ej., $\alpha \in \{0.10, 0.05, 0.01\}$) y conteos por tipo.

6 Modelos de regresión espacial

Sea ε el error idiosincrático y $\mathbf{u}$ un término de error espacial.

SAR (Spatial Autoregressive / Lag)

$$y = \rho \, \mathbf{W} y + \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I}).$$

Hipótesis: $H_0 : \rho = 0$ (sin rezago espacial en la variable dependiente) vs $H_A : \rho \neq 0$. Efectos marginales e impactos (directo/indirecto/total) se obtienen vía $(\mathbf{I} - \rho \mathbf{W})^{-1}$.

SEM (Spatial Error Model)

$$y = \mathbf{X}\beta + u, \qquad u = \lambda \, \mathbf{W} u + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I}).$$

Hipótesis: $H_0 : \lambda = 0$ (no autocorrelación en el error) vs $H_A : \lambda \neq 0$.

SDM (Spatial Durbin Model)

$$y = \rho \, \mathbf{W} y + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\theta + \varepsilon.$$

Hipótesis: (i) $H_0 : \rho = 0$; (ii) $H_0 : \theta = \mathbf{0}$; (iii) pruebas anidadas para simplificar a SAR (si $\theta = \mathbf{0}$) o a SLX (si $\rho = 0$).

SAC / SARAR

$$y = \rho \mathbf{W} y + \mathbf{X}\beta + u, \qquad u = \lambda \mathbf{W} u + \varepsilon.$$

Hipótesis conjuntas: $H_0 : \rho = 0$ y/o $H_0 : \lambda = 0$.

Selección de modelo.

1. Comprobar autocorrelación en residuos de OLS: Moran de residuos y pruebas LM robustas (LM-lag, LM-error).
2. Estimar candidatos (SAR, SEM, SDM, SAC) por MV o MCMC; comparar AIC/BIC, *pseudo*-R^2, diagnóstico de residuos (Moran-LM sobre residuos).
3. Reportar impactos espaciales (directo/indirecto/total) para modelos con rezagos en y o en $\mathbf{X}$.

7 Procedimiento de prueba e inferencia

1. **Definir W**: justificar Queen/Rook o distancia (radio r o k-NN) y normalizar por filas; documentar grado promedio y densidad.
2. **Moran global**: estimar I y p por permutaciones; especificar alternativa (bilateral o unilateral).
3. **LISA**: calcular I_i, p_i por permutaciones; mapear significancias y clasificar en HH, LL, HL, LH.
4. **Modelado**: ejecutar OLS base, pruebas LM robustas; estimar modelo espacial elegido; comparar por AIC/BIC y diagnóstico de residuos; reportar impactos e intervalos de confianza.
5. **Sensibilidad**: replicar con varias **W** (Queen, Rook, distancia inversa y exponencial) y parámetros (α, r, k).

8 Interpretación técnica de resultados (ejemplo guiado)

Moran global

Resultados con cuatro **W** (Queen, Rook, distancia inversa, exponencial) muestran I cercanos a cero y *no significativos* ($p > 0.39$), lo que sugiere **ausencia de patrón global** claro de asociación (ni conglomeración ni repulsión). Para inferencia práctica, se mantiene H_0 de aleatoriedad espacial a niveles usuales ($\alpha = 0.05$).

LISA

Aunque la asociación global es débil, el análisis local detecta *pocos clústeres* significativos, predominando **LH** (altos rodeados de bajos) y algunos **LL/HH** puntuales. En términos sustantivos, esto sugiere microzonas con valores inusualmente altos (o bajos) respecto a sus vecinas, útiles para focalización territorial. Se recomienda reportar mapas con p-valores y tipología por **W**.

Comparativa de W

Queen y Rook exhiben conectividades y grados promedios similares, mientras que matrices de distancia muestran **sensibilidad** mayor a los hiperparámetros (p. ej. α): al incrementar la conectividad efectiva, los estadísticos pueden variar en magnitud y significancia. Por transparencia, acompañar cada resultado con el grado promedio y la densidad observada bajo cada **W**.

Implicaciones para regresión espacial

Dado el *Moran global no significativo*, la evidencia no impone de entrada un rezago en y ni un proceso de error espacial. Aun así:

- Comience por OLS y contraste con LM robustos. Si LM-lag y/o LM-error resultan no significativos, OLS puede ser suficiente.
- Si el fenómeno sugiere *spillovers* en regresores (efectos de vecindad de las covariables), el SDM puede ser informativo aun con I global pequeño.
- Compare modelos por AIC/BIC y verifique Moran-LM en residuos; reporte impactos si $\rho \neq 0$ o $\theta \neq \mathbf{0}$.

9 Tablas mínimas sugeridas

Table 1: Moran's I global por matriz de pesos

W	I	p (perm.)	Nota
Queen	-0.065	0.39	No sig.
Rook	-0.062	0.44	No sig.
InvDist	-0.050	0.49	No sig.
InvExp	-0.050	0.47	No sig.

Table 2: Conteos LISA significativos por tipología (ejemplo)

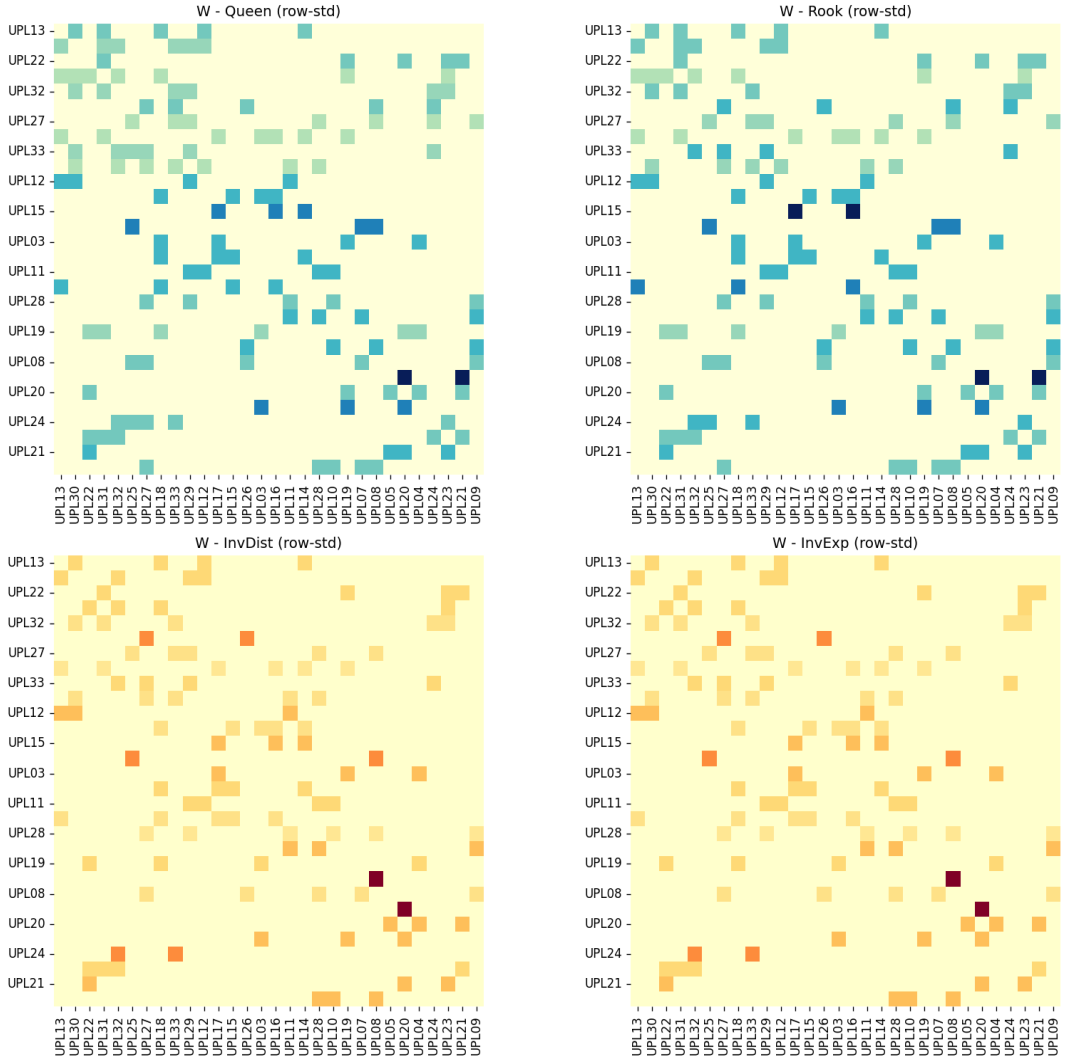
W	HH	LL	HL	LH
Queen	0	1	0	2
Rook	0	0	0	2
InvDist	1	1	0	3
InvExp	1	0	0	3

10 Buenas prácticas de reporte

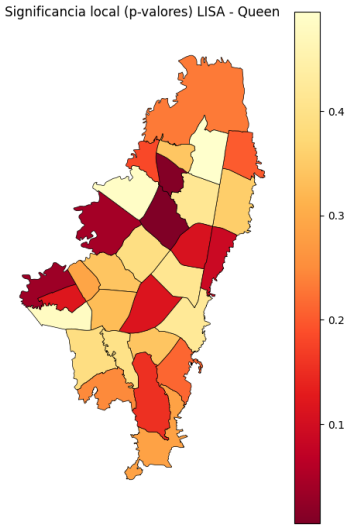
- Documentar construcción y normalización de **W**, grado promedio y densidad.
- Especificar número de permutaciones y nivel α para Moran/LISA.
- Incluir mapas LISA con máscaras de significancia y leyendas claras.
- Para SAR/SDM/SAC, reportar impactos (directo, indirecto, total) con IC del 95%.
- Análisis de sensibilidad a la elección de **W** y a α (exponencial) o k/r (distancia).

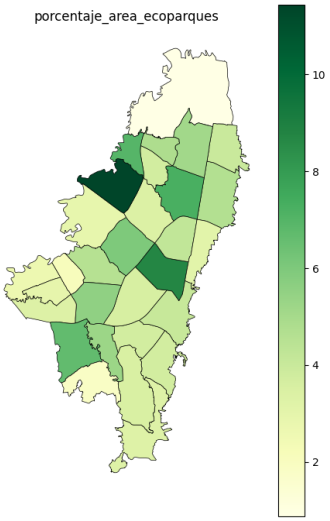
Análisis Espacial de Ecoparques y LISA

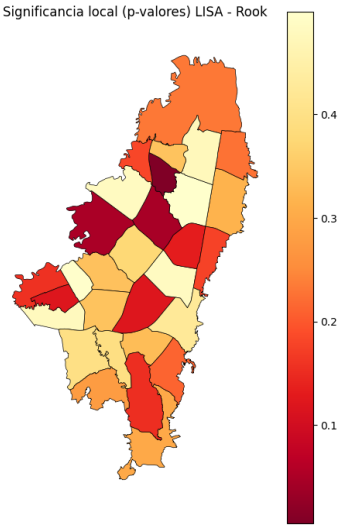
Matrices de Pesos Espaciales



Significancia Local LISA y Porcentaje de Área de Ecoparques







=== Resumen de Lags ===

Variable	count	mean	std	min	25%	50%	75%	max
porcentaje_area_ecoparques	30.0	4.4748	2.1422	0.8737	3.3046	3.8693	5.222	11.4388
lag_queen	30.0	4.5568	0.9482	2.8101	4.0188	4.29	5.2024	6.787
lag_rook	30.0	4.577	0.8838	3.156	4.0398	4.3453	5.2024	6.787
lag_invdist	30.0	4.7221	1.1714	2.8101	3.8649	4.3605	5.464	7.3427
lag_invexp	30.0	4.7221	1.1714	2.8101	3.8649	4.3605	5.464	7.3427

=== Grado Promedio y Densidad ===

W	vecinos_prom	densidad
Queen	4.8	0.0345
Rook	4.4667	0.0345
InvDist	3.6667	0.0345
InvExp	3.6667	0.0345

=== Moran’s I Global ===

W	Moran_I	p_val
Queen	-0.0652	0.391
Rook	-0.0619	0.442
InvDist	-0.0499	0.485
InvExp	-0.0499	0.466

=== Conteos de Clusters LISA (Significativos) ===

Los resultados muestran la distribución de los tipos de clusters identificados mediante el estadístico LISA:

• **Queen:** 2 clusters de tipo Alto-Bajo (LH) y 1 de Bajo-Bajo (LL)

• **Rook:** 2 clusters Alto-Bajo (LH)

• **InvDist:** 3 LH, 1 LL y 1 HH

• **InvExp:** 3 LH y 1 HH

=== Sensibilidad Cargada (sens_df) ===

Tipo	factor_conect	alpha	Moran_I	p_val	vecinos_prom
InvDist	1.0		-0.0499	0.478	3.6667
InvDist	1.2		-0.0124	0.407	4.9333
InvDist	1.5		0.0456	0.195	7.5333
InvExp	1.0	0.0	-0.0499	0.46	3.6667
InvExp	1.0	0.0001	-0.0499	0.459	3.6667
InvExp	1.0	0.0001	-0.0499	0.481	3.6667

Análisis e Interpretación:
Los resultados sugieren una débil autocorrelación espacial global (valores de Moran's I cercanos a 0 y no significativos), indicando que la distribución de los porcentajes de área de ecoparques no presenta un patrón espacial claro de concentración o dispersión. En el análisis local (LISA), se identifican pocos clusters significativos, principalmente de tipo Alto-Bajo (LH), lo que sugiere la presencia de zonas con altos valores de ecoparques rodeadas por áreas de bajos valores, y viceversa.

Los pesos espaciales tipo Queen y Rook muestran una estructura de conectividad similar, mientras que los modelos basados en distancia (Inverse Distance e Inverse Exponential) presentan una mayor sensibilidad al parámetro de conectividad, como se observa en el cambio del número de vecinos promedio.

En términos de planificación territorial, los resultados sugieren que los ecoparques no están agrupados de manera significativa en el espacio, aunque existen áreas puntuales donde su presencia es mayor o menor de lo esperado según el entorno.