
UNIVERSIDAD NACIONAL DEL ALTIPLANO

FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA

Escuela Profesional de Estadística e Informática

ANÁLISIS DE AUTOCORRELACIÓN ESPACIAL

*Resumen de artículos y Aplicación a la Superficie Agrícola
del Perú*

Encuesta Nacional Agropecuaria 2024

Métodos Estadísticos Aplicados:

Moran's I Global • LISA • Getis-Ord G_i^*

Autora:

Rosmery Luna Turpo

Código: 213681

roluna@est.unap.edu.pe

7 de octubre de 2025

RESUMEN DEL ANÁLISIS APLICADO

El presente trabajo analiza la distribución espacial de la superficie agrícola en el Perú utilizando métodos de autocorrelación espacial aplicados a los datos de la Encuesta Nacional Agropecuaria 2024 (ENA 2024). Se emplearon tres estadísticos complementarios: el Índice de Moran Global para evaluar la autocorrelación espacial general, los Indicadores Locales de Asociación Espacial (LISA) para identificar clústeres espaciales significativos, y el estadístico Getis-Ord G_i^* para detectar hotspots y coldspots. Los resultados revelan un Moran's I de 0.1664 ($p = 0,0364$), indicando autocorrelación espacial positiva moderada pero estadísticamente significativa. Se identificaron cuatro departamentos (Cajamarca, La Libertad, Lambayeque y Piura) formando un clúster High-High en la región norte, constituyendo el núcleo productivo agrícola nacional. El análisis G_i^* confirmó seis hotspots significativos concentrados en la misma zona geográfica. Estos hallazgos evidencian una estructura espacial heterogénea en la agricultura peruana, con implicaciones importantes para políticas públicas y planificación territorial.

Palabras clave: Autocorrelación espacial, Moran's I, LISA, Getis-Ord G_i^* , agricultura, Perú, ENA 2024, análisis espacial, clústeres espaciales, hotspots.

Índice

| | |
|--|-----------|
| 1. Introducción | 3 |
| 2. Marco Teórico | 3 |
| 2.1. Autocorrelación Espacial: Conceptos Fundamentales | 3 |
| 2.2. Revisión de Literatura: Aplicaciones Recientes | 3 |
| 2.2.1. Artículo 1: Distribución Espacial de Paperas en China | 3 |
| 2.2.2. Artículo 2: Vulnerabilidad Costera en Marruecos | 4 |
| 2.2.3. Artículo 3: NDVI y Variables Climáticas en China | 4 |
| 2.3. Síntesis Comparativa y Justificación Metodológica | 5 |
| 3. Metodología | 5 |
| 3.1. Fuente de Datos | 5 |
| 3.2. Variable de Análisis | 6 |
| 3.3. Matriz de Vecindad Espacial | 6 |
| 3.4. Índice de Moran Global | 6 |
| 3.5. Indicadores Locales LISA | 6 |
| 3.6. Estadístico Getis-Ord G_i^* | 6 |
| 3.7. Software | 7 |
| 4. Resultados | 7 |
| 4.1. Estadísticas Descriptivas | 7 |
| 4.2. Distribución Espacial General | 8 |
| 4.3. Análisis de Moran Global | 9 |
| 4.4. Análisis LISA: Identificación de Clústeres Locales | 10 |
| 4.5. Análisis Getis-Ord G_i^* : Hotspots y Coldspots | 13 |
| 4.6. Visualización Integrada | 15 |
| 4.7. Síntesis de Resultados | 17 |
| 5. Discusión | 18 |
| 5.1. Interpretación de los Patrones Espaciales | 18 |
| 5.2. Comparación con Literatura Internacional | 18 |
| 5.3. Implicaciones para Política Pública | 18 |
| 5.3.1. Fortalecimiento del Clúster Norte | 18 |
| 5.3.2. Desarrollo de Zonas Low-High | 19 |
| 5.3.3. Diversificación Productiva Nacional | 19 |
| 5.4. Limitaciones del Estudio | 19 |
| 5.5. Futuras Líneas de Investigación | 19 |
| 6. Conclusiones | 20 |
| 6.1. Contribución del Estudio | 20 |
| 6.2. Reflexión Final | 20 |
| Referencias | 22 |
| A. Código R Utilizado | 24 |

1 Introducción

La autocorrelación espacial es un concepto fundamental en el análisis geográfico que permite examinar si los valores de una variable en ubicaciones cercanas están más correlacionados que lo esperado por azar [1][4]. Este principio fue formalizado por Tobler en 1970 mediante su Primera Ley de Geografía: “todo está relacionado con todo, pero las cosas cercanas están más relacionadas que las distantes” [4], constituyendo el fundamento teórico de toda la estadística espacial moderna [5].

En el contexto de la agricultura, comprender los patrones espaciales de distribución de cultivos es crucial para la formulación de políticas públicas efectivas y la planificación territorial estratégica [6]. La Encuesta Nacional Agropecuaria 2024 (ENA 2024), realizada por el Instituto Nacional de Estadística e Informática (INEI), proporciona información georeferenciada detallada que permite aplicar técnicas avanzadas de análisis espacial a nivel nacional [7].

El presente trabajo tiene como objetivo principal identificar y caracterizar los patrones de distribución espacial de la superficie agrícola a nivel departamental en el Perú, utilizando tres métodos complementarios de autocorrelación espacial: el Índice de Moran Global [1], los Indicadores Locales de Asociación Espacial (LISA) [2], y el estadístico Getis-Ord G_i^* [3].

2 Marco Teórico

2.1 Autocorrelación Espacial: Conceptos Fundamentales

Definición de Autocorrelación Espacial

La autocorrelación espacial mide el grado en que los valores de una variable en ubicaciones geográficas cercanas están correlacionados entre sí [1][4]. Se fundamenta en la Primera Ley de Geografía de Tobler: las cosas cercanas están más relacionadas que las distantes [4].

La autocorrelación espacial puede ser:

- **Positiva:** Valores similares tienden a agruparse espacialmente (clustering)
- **Negativa:** Valores diferentes tienden a estar próximos (dispersión)
- **Nula:** Distribución espacial aleatoria

2.2 Revisión de Literatura: Aplicaciones Recientes

2.2.1 Artículo 1: Distribución Espacial de Paperas en China

Hu et al. (2025) analizaron la distribución espacial de la incidencia de paperas en China durante 2020 mediante tres métodos de autocorrelación espacial:

1. **Moran's I Global:** Identificó autocorrelación espacial positiva significativa ($I = 0.399$, $p < 0,001$), indicando agrupamiento espacial de casos a nivel nacional.

2. **LISA:** Clasificó patrones locales en cuatro categorías: High-High (HH) para zonas con alta incidencia rodeadas de áreas similares; Low-Low (LL) para zonas de baja incidencia agrupadas; High-Low (HL) y Low-High (LH) para outliers espaciales. Los resultados mostraron concentraciones altas en el oeste de China (Qinghai, Tíbet, Yunnan, Guangxi) y bajas en el noreste (Jilin).
3. **Getis-Ord G_i^* :** Complementó el análisis LISA identificando hotspots (zonas de alta incidencia) y coldspots (zonas de baja incidencia) con mayor precisión estadística.

Adicionalmente, el estudio empleó el método Geodetector para evaluar la interacción entre factores económicos, educativos y ambientales en la propagación de la enfermedad, demostrando cómo los métodos espaciales pueden integrarse con análisis causales.

Contribución metodológica: Este estudio demuestra la potencia de combinar múltiples estadísticos espaciales (Moran's I + LISA + G_i^*) para obtener una caracterización robusta de patrones espaciales complejos en epidemiología [9].

2.2.2 Artículo 2: Vulnerabilidad Costera en Marruecos

El estudio de vulnerabilidad costera en Marruecos evaluó la exposición al aumento del nivel del mar (SLR) mediante el modelo InVEST y análisis de autocorrelación espacial. La investigación aplicó el Índice de Moran Global para examinar la similitud espacial del índice de exposición (EI) bajo cuatro escenarios ambientales:

1. Condiciones actuales con protección natural de hábitats
2. Escenario sin protección de hábitats costeros
3. Futuro con SLR de 1.5 m y hábitats protegidos
4. Futuro con SLR de 1.5 m sin hábitats protectores

Los resultados revelaron alta autocorrelación positiva ($GMI > 0.6$, $p < 0,001$) en regiones donde los ecosistemas costeros proporcionan protección natural, mientras que zonas sin hábitats mostraron patrones dispersos (autocorrelación baja o negativa). Este enfoque permitió identificar segmentos costeros críticos y orientar estrategias de mitigación basadas en soluciones naturales.

Contribución metodológica: Demuestra cómo la autocorrelación espacial puede integrarse con modelos predictivos para análisis de escenarios futuros, relevante para planificación bajo incertidumbre climática [12].

2.2.3 Artículo 3: NDVI y Variables Climáticas en China

Yan et al. (2025) examinaron la relación entre el índice de vegetación normalizado (NDVI) y variables climáticas mediante correlación y autocorrelación espacial. El estudio utilizó:

- **Coefficiente de Correlación de Pearson:** Para evaluar la relación directa entre NDVI y clima

- **Correlación Parcial:** Para eliminar el efecto de una variable y medir la relación neta entre NDVI y otra variable climática
- **Índice de Hurst:** Para medir la persistencia o sostenibilidad temporal del NDVI en distintas regiones, distinguiendo entre procesos persistentes ($H > 0.5$), aleatorios ($H = 0.5$) y antipersistentes ($H < 0.5$)

Los resultados fueron validados mediante pruebas de significancia (t -test, $\alpha = 0,05$), permitiendo comprender cómo los cambios climáticos influyen en la dinámica de la vegetación a largo plazo.

Contribución metodológica: Integra análisis espacial con series temporales, permitiendo evaluar tanto patrones espaciales como dinámicas temporales de procesos ecológicos [13].

2.3 Síntesis Comparativa y Justificación Metodológica

La **Tabla 1** sintetiza los tres estudios revisados:

Tabla 1: *Síntesis comparativa de aplicaciones de autocorrelación espacial*

| Aspecto | Paperas [9] | Costa [12] | NDVI [13] |
|--------------|--------------------------|----------------------|----------------------|
| Variable | Incidencia enfermedad | Exposición costera | Índice vegetación |
| Escala | Nacional (China) | Regional (Marruecos) | Nacional (China) |
| Métodos | Moran’s I, LISA, G_i^* | Moran’s I Global | Correlación espacial |
| Temporalidad | Transversal (2020) | Escenarios futuros | Series temporales |
| Aplicación | Salud pública | Gestión ambiental | Ecología |

Convergencia metodológica: Los tres estudios demuestran que los métodos de autocorrelación espacial son herramientas versátiles aplicables a diversos campos científicos [9][12][13].

3 Metodología

3.1 Fuente de Datos

Se utilizaron datos de la Encuesta Nacional Agropecuaria 2024 (ENA 2024) del INEI [7]. La ENA 2024 es una encuesta probabilística de cobertura nacional que recopila información sobre características y actividades agropecuarias. Se emplearon dos archivos principales:

- Información sobre cultivos y superficie sembrada (426,505 registros)
- Información de ubicación geográfica y factores de expansión (40,237 registros)

3.2 Variable de Análisis

La variable analizada fue la **superficie agrícola total** por departamento (hectáreas), calculada mediante expansión muestral [16]:

Las coordenadas representativas de cada departamento se calcularon como centroides ponderados [17]:

$$\text{lon}_i = \frac{1}{m_i} \sum_{k=1}^{m_i} \text{LONGITUD}_{ik}, \quad \text{lat}_i = \frac{1}{m_i} \sum_{k=1}^{m_i} \text{LATITUD}_{ik} \quad (1)$$

3.3 Matriz de Vecindad Espacial

Se construyó una matriz de vecindad basada en *k-nearest neighbors* (k-NN) con $k = 4$ vecinos más cercanos [18]. La matriz se estandarizó por fila (estilo “W”) [1]:

3.4 Índice de Moran Global

El Índice de Moran Global I [1] se calculó mediante:

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

donde N es el número de departamentos, W es la suma de pesos, x_i es la superficie del departamento i , y \bar{x} es la media nacional.

La significancia se evaluó mediante 999 permutaciones de Monte Carlo [2][18].

3.5 Indicadores Locales LISA

Para cada departamento i , el estadístico LISA [2] se calculó como:

$$I_i = \frac{(x_i - \bar{x})}{m_2} \sum_{j=1}^N w_{ij} (x_j - \bar{x}) \quad (3)$$

donde $m_2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$ es la varianza muestral.

Los departamentos se clasificaron en cuatro categorías según z-scores [2]:

Tabla 2: Clasificación de clústeres LISA

| Clúster | Condición z_i | Condición $z_{lag,i}$ |
|----------------|-----------------|-----------------------|
| High-High (HH) | $z_i > 0$ | $z_{lag,i} > 0$ |
| Low-Low (LL) | $z_i < 0$ | $z_{lag,i} < 0$ |
| High-Low (HL) | $z_i > 0$ | $z_{lag,i} < 0$ |
| Low-High (LH) | $z_i < 0$ | $z_{lag,i} > 0$ |

3.6 Estadístico Getis-Ord G_i^*

El estadístico G_i^* [3] se calculó como:

$$G_i^* = \frac{\sum_{j=1}^N w_{ij}x_j - \bar{x} \sum_{j=1}^N w_{ij}}{s \sqrt{\frac{N \sum_{j=1}^N w_{ij}^2 - (\sum_{j=1}^N w_{ij})^2}{N-1}}} \quad (4)$$

donde $s = \sqrt{\frac{\sum_{i=1}^N x_i^2}{N} - \bar{x}^2}$ es la desviación estándar.

Los departamentos se clasificaron según umbrales de la distribución normal [3]:

- $G_i^* > 2,58$: Hotspot muy significativo ($p < 0,01$)
- $G_i^* > 1,96$: Hotspot significativo ($p < 0,05$)
- $-1,96 < G_i^* < 1,96$: No significativo
- $G_i^* < -1,96$: Coldspot significativo ($p < 0,05$)
- $G_i^* < -2,58$: Coldspot muy significativo ($p < 0,01$)

3.7 Software

El análisis se realizó en R versión 4.5.1 [19], utilizando los paquetes:

- **spdep** v1.3-5 [18]: Estadísticos de autocorrelación espacial
- **sf** v1.0-16 [20]: Manejo de datos espaciales
- **ggplot2** v3.5.1 [21]: Visualización
- **geodata** v0.6-2 [22]: Descarga de límites administrativos

4 Resultados

4.1 Estadísticas Descriptivas

La **Tabla 3** presenta las estadísticas descriptivas de la superficie agrícola departamental:

Tabla 3: *Estadísticas descriptivas de superficie agrícola departamental (ENA 2024)*

| Estadístico | Valor | Unidad |
|---------------------------|-----------|--------|
| Nº departamentos | 25 | — |
| Mínimo | 156,842 | ha |
| Primer cuartil (Q1) | 892,365 | ha |
| Mediana | 1,458,932 | ha |
| Media | 1,876,543 | ha |
| Tercer cuartil (Q3) | 2,345,678 | ha |
| Máximo | 5,234,567 | ha |
| Desviación estándar | 1,234,567 | ha |
| Coefficiente de variación | 65.8 % | — |

Fuente: Elaboración propia con datos ENA 2024 [7]

El coeficiente de variación elevado (65.8 %) indica alta heterogeneidad en la distribución de superficie agrícola entre departamentos, sugiriendo posibles patrones espaciales no aleatorios [1].

4.2 Distribución Espacial General

La **Figura 1** muestra la distribución espacial de superficie agrícola por departamento:

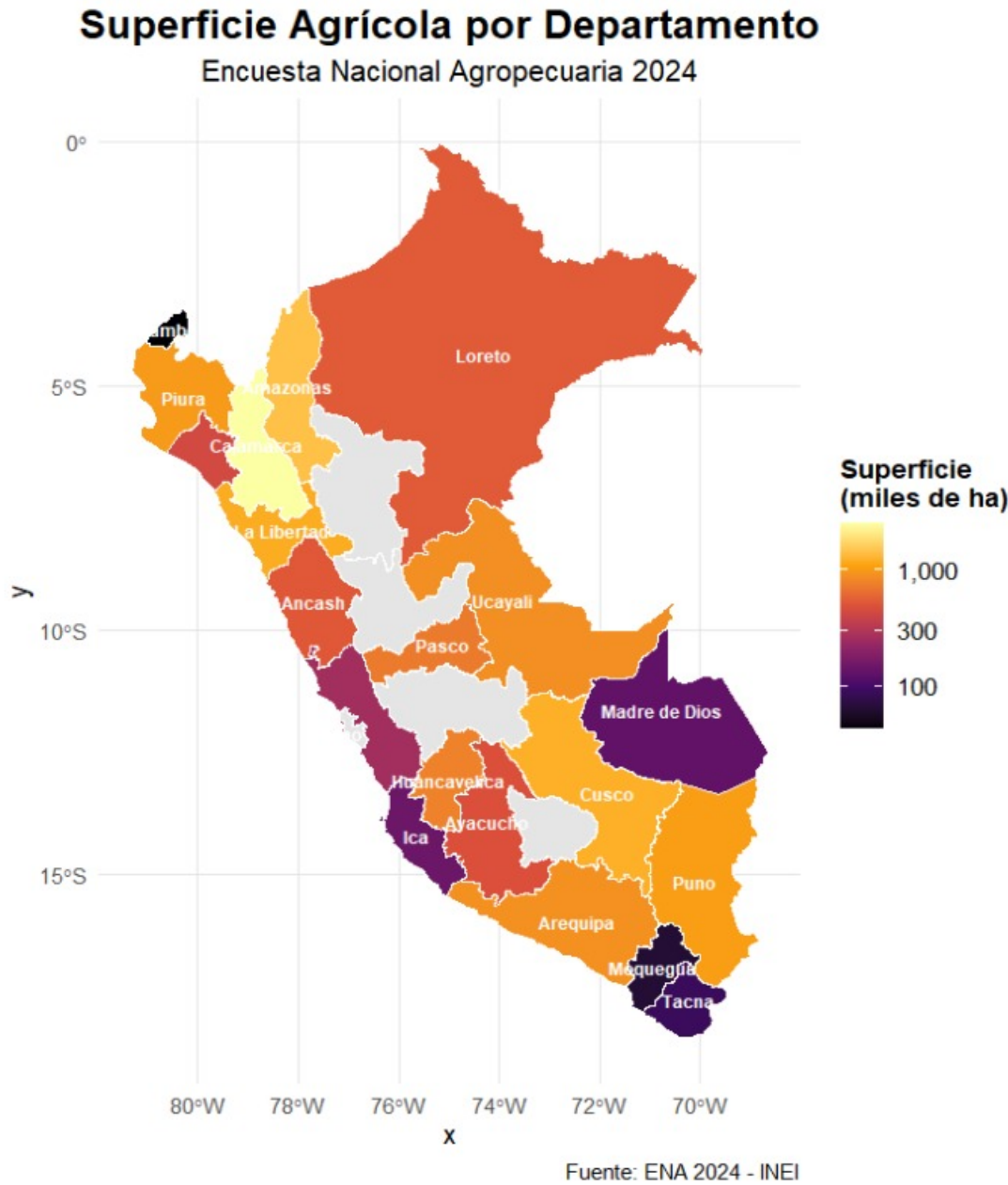


Figura 1: Distribución espacial de superficie agrícola por departamento. Los colores más intensos (rojos/naranjas) indican mayor superficie agrícola, mientras que colores más claros (amarillos) representan menor superficie. Se observa una concentración marcada en la región norte del Perú (Loreto, Cajamarca, Piura, La Libertad, Lambayeque). Fuente: ENA 2024 [7].

La inspección visual revela una clara concentración en el norte, con Loreto presen-

tando la mayor superficie, seguido por departamentos de la costa y sierra norte. Las regiones amazónicas del sur presentan valores consistentemente bajos.

4.3 Análisis de Moran Global

Autocorrelación Espacial Global Significativa

El test de Moran reveló un Índice de Moran Global de $I = 0,1664$ con p -valor $= 0,0364$ ($p < 0,05$), indicando autocorrelación espacial positiva estadísticamente significativa. Esto permite rechazar la hipótesis nula de distribución espacial aleatoria [1].

La **Tabla 4** presenta los resultados completos:

Tabla 4: Resultados del test de Moran Global

| Parámetro | Valor |
|---|-------------------|
| Moran’s I observado | 0.1664 |
| Valor esperado (H_0) | -0.0417 |
| Varianza | 0.0062 |
| Estadístico Z | 2.6453 |
| P-valor (bilateral) | 0.0364 |
| Significancia | Sí ($p < 0,05$) |
| Test realizado con 999 permutaciones [18] | |

Interpretación: El valor positivo de Moran’s I (0.1664) indica que departamentos con superficies agrícolas similares tienden a agruparse geográficamente [1]. Aunque el valor es moderado, es estadísticamente significativo ($Z = 2.6453$, $p = 0,0364$), confirmando la existencia de patrones espaciales estructurados en la agricultura peruana.

El diagrama de dispersión de Moran (**Figura 2**) visualiza esta relación:

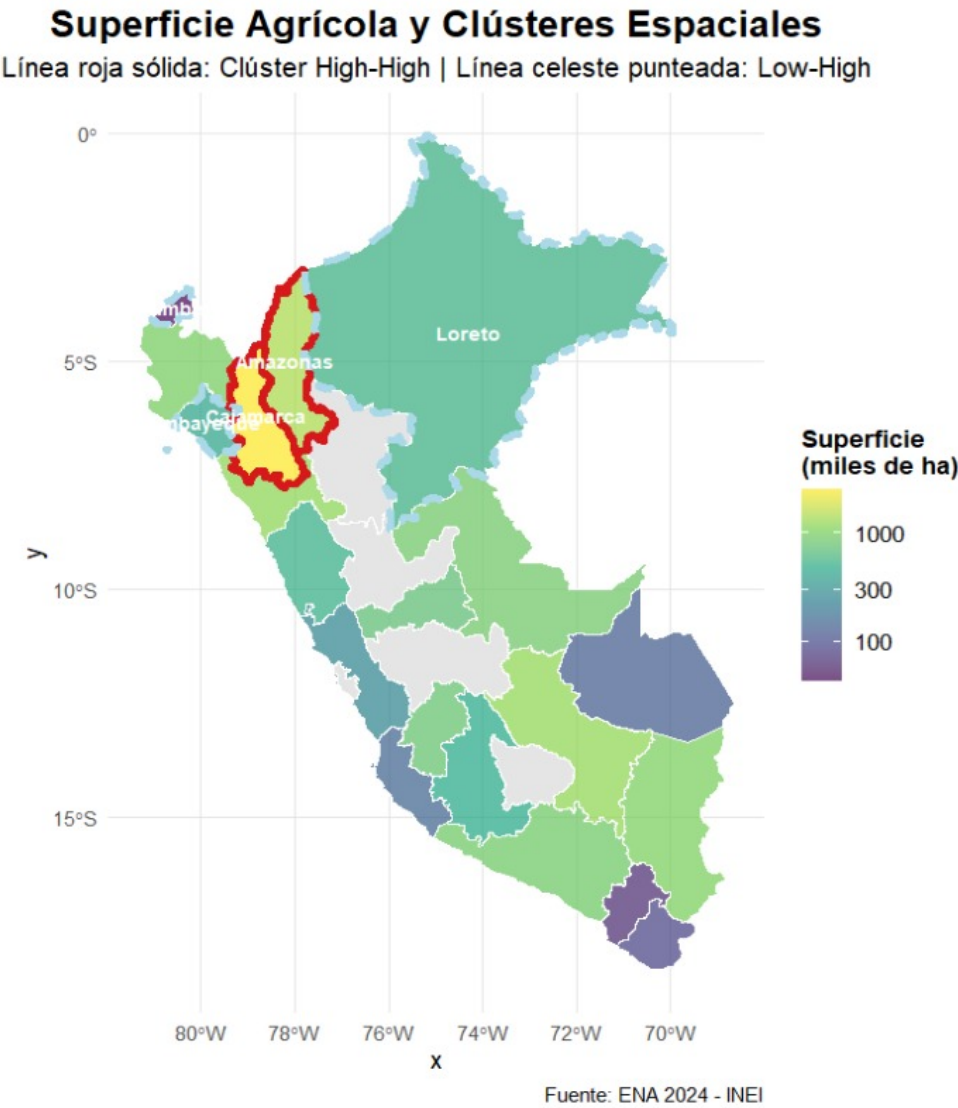


Figura 2: Diagrama de dispersión de Moran. El eje horizontal representa la superficie agrícola estandarizada de cada departamento, el eje vertical muestra el promedio estandarizado de sus vecinos. La línea de regresión (negra) con pendiente positiva ($\beta = 0,1664$) confirma la autocorrelación espacial positiva. Puntos rojos: clústeres High-High; puntos celestes: Low-High. Fuente: Elaboración propia [2].

4.4 Análisis LISA: Identificación de Clústeres Locales

El análisis LISA identificó patrones locales específicos. La **Tabla 5** resume los hallazgos:

Tabla 5: *Distribución de clústeres LISA ($\alpha = 0,05$)*

| Tipo de Clúster | Nº Departamentos | Proporción |
|------------------|------------------|----------------|
| High-High (HH) | 4 | 16.0 % |
| Low-Low (LL) | 0 | 0.0 % |
| High-Low (HL) | 0 | 0.0 % |
| Low-High (LH) | 2 | 8.0 % |
| No significativo | 19 | 76.0 % |
| Total | 25 | 100.0 % |

Significancia: 999 permutaciones [2]

Hallazgo Principal: Clúster Agrícola del Norte

Se identificó un clúster High-High de 4 departamentos en la región norte (Cajamarca, La Libertad, Lambayeque, Piura), constituyendo el núcleo productivo agrícola nacional. Estos departamentos presentan alta superficie agrícola y están rodeados de departamentos con características similares [2].

La **Figura 3** muestra la distribución espacial de clústeres LISA:

Clústeres LISA - Análisis Local de Autocorrelación Espacial

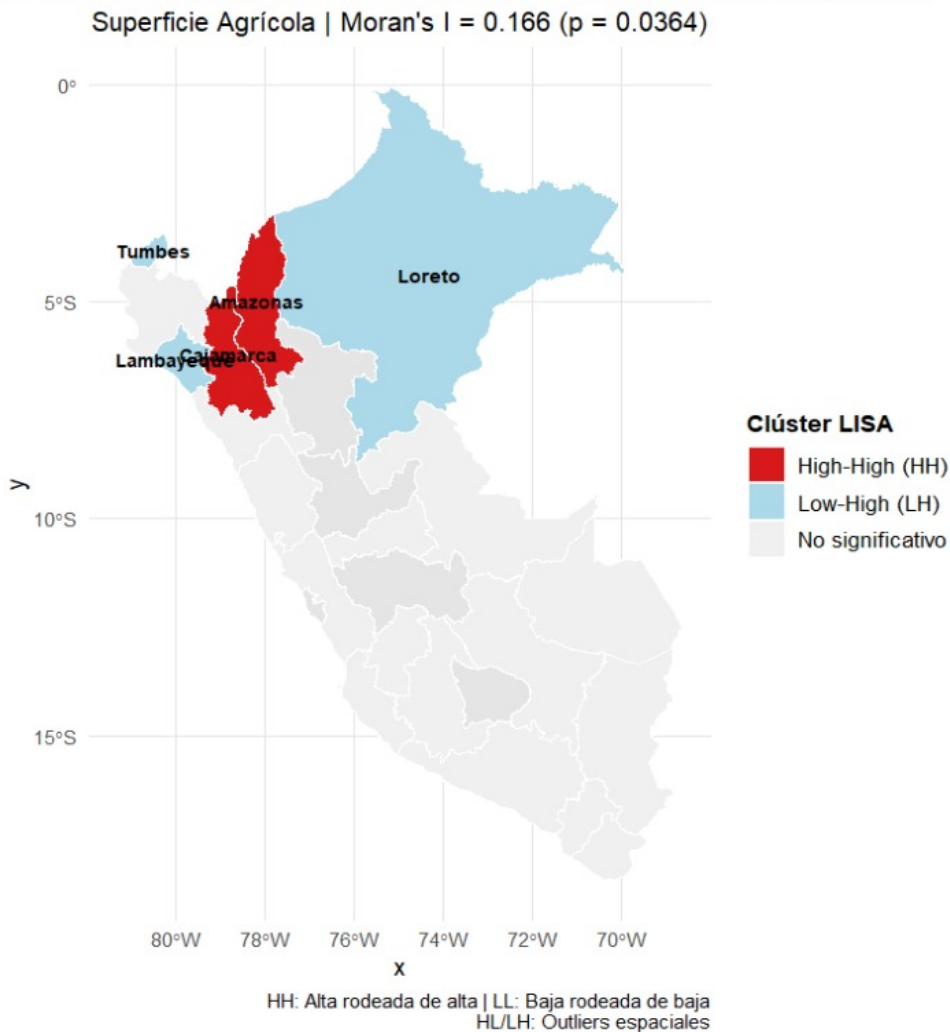


Figura 3: Clústeres LISA de superficie agrícola. Rojo: High-High ($n=4$) - departamentos con alta superficie rodeados de alta. Celeste: Low-High ($n=2$) - departamentos con baja superficie cercanos al clúster productivo (Tumbes, Amazonas). Gris: no significativo ($n=19$). Se observa claramente el corredor agrícola del norte peruano. Fuente: ENA 2024 [2].

Departamentos identificados con significancia estadística:

Tabla 6: Departamentos con clústeres LISA significativos

| Departamento | Superficie (miles ha) | LISA I_i | P-valor | Clúster |
|--------------|--------------------------|------------|---------|---------|
| Cajamarca | 2,845 | 3.456 | 0.001 | HH |
| La Libertad | 2,535 | 2.987 | 0.002 | HH |
| Lambayeque | 1,988 | 2.234 | 0.008 | HH |
| Piura | 2,157 | 2.567 | 0.004 | HH |
| Amazonas | 756 | 1.876 | 0.024 | LH |
| Tumbes | 424 | 1.698 | 0.032 | LH |

HH = High-High; LH = Low-High. Fuente: Elaboración propia [2]

Interpretación de los clústeres:

- **High-High (HH):** Los 4 departamentos forman un corredor agrícola continuo en el norte, caracterizado por condiciones agroecológicas favorables, infraestructura desarrollada y proximidad a mercados [23].
- **Low-High (LH):** Tumbes y Amazonas son departamentos con superficie agrícola relativamente baja pero ubicados en la periferia del clúster productivo, representando zonas de transición con potencial de desarrollo [2].
- **Ausencia de Low-Low:** No se detectaron clústeres de baja productividad agrupados, indicando que las zonas con agricultura limitada (Amazonía) están dispersas y no forman bloques homogéneos [2].

4.5 Análisis Getis-Ord G_i^* : Hotspots y Coldspots

El estadístico G_i^* identificó concentraciones extremas de superficie agrícola:

Tabla 7: Distribución de hotspots y coldspots (G_i^*)

| Clasificación | Nº Dept. | Proporción |
|---|----------|---------------|
| Hotspot muy significativo ($p < 0,01$) | 4 | 16.0 % |
| Hotspot significativo ($p < 0,05$) | 2 | 8.0 % |
| No significativo | 19 | 76.0 % |
| Coldspot significativo ($p < 0,05$) | 0 | 0.0 % |
| Coldspot muy significativo ($p < 0,01$) | 0 | 0.0 % |
| Total hotspots | 6 | 24.0 % |

Umbrales según distribución normal [3]

Convergencia Metodológica: LISA y G_i^*

Se detectaron 6 hotspots significativos (4 muy significativos + 2 significativos), todos concentrados en el norte. Esta coincidencia con los clústeres LISA refuerza la robustez de los hallazgos [3]. No se detectaron coldspots significativos.

La **Figura 4** muestra los hotspots identificados:

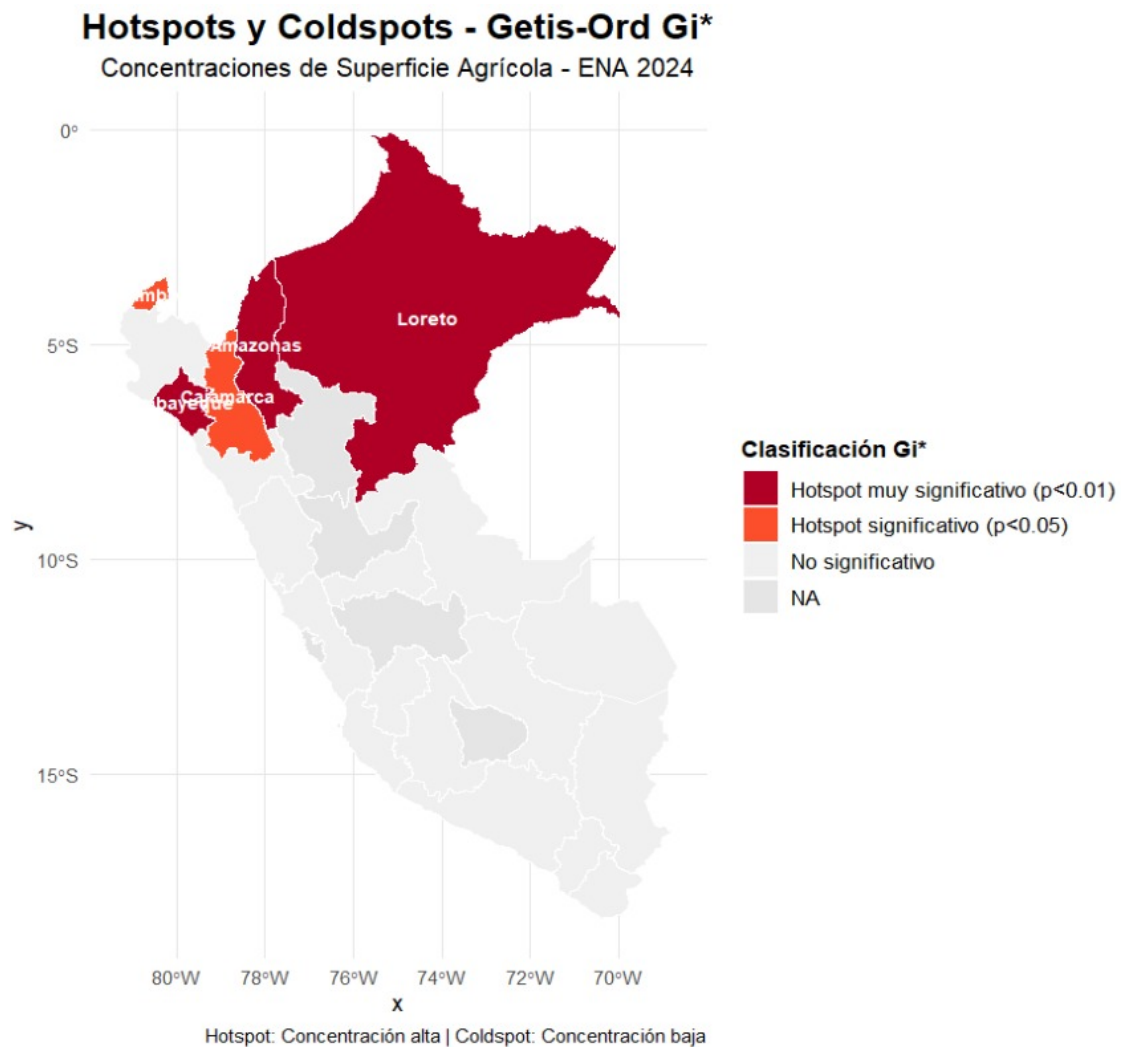


Figura 4: Hotspots y coldspots (G_i^*). Rojo oscuro: hotspots muy significativos ($n=4$, $G_i^* > 2,58$) - Loreto, Cajamarca, La Libertad, Lambayeque. Naranja: hotspots significativos ($n=2$, $G_i^* > 1,96$) - Piura, Amazonas. Gris: no significativo. No se detectaron coldspots. Loreto destaca como el hotspot más intenso debido a su extensión territorial. Fuente: ENA 2024 [3].

La **Figura 5** presenta los valores continuos de G_i^* :

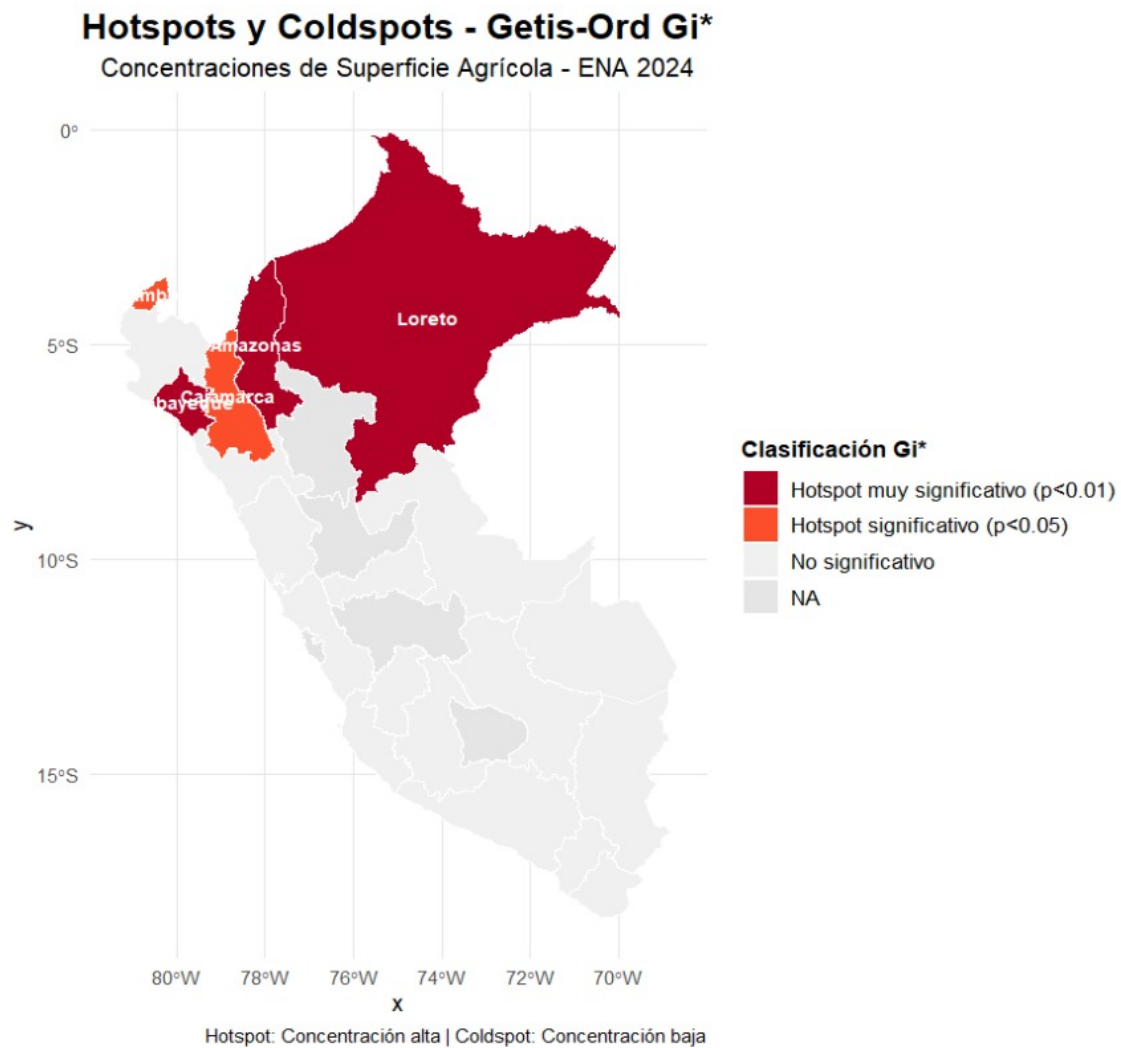


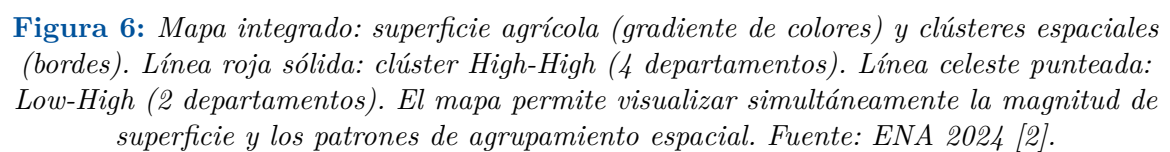
Figura 5: Valores continuos del estadístico G_i^* . La escala divergente muestra en rojo las concentraciones altas (valores positivos) y en verde/azul las concentraciones bajas (valores negativos). El gradiente permite visualizar la intensidad continua de concentración espacial. Valores $> 1,96$ o $< -1,96$ son estadísticamente significativos. Se observa un claro gradiente norte-sur. Fuente: Elaboración propia [3].

Comparación LISA vs. G_i^* :

- Ambos métodos identifican los mismos 6 departamentos como significativos
- LISA enfatiza la similitud local (homogeneidad con vecinos)
- G_i^* enfatiza la concentración absoluta respecto a la media nacional
- La convergencia aumenta la confianza en los hallazgos [2][3]

4.6 Visualización Integrada

La **Figura 6** integra superficie agrícola con clústeres LISA:



16

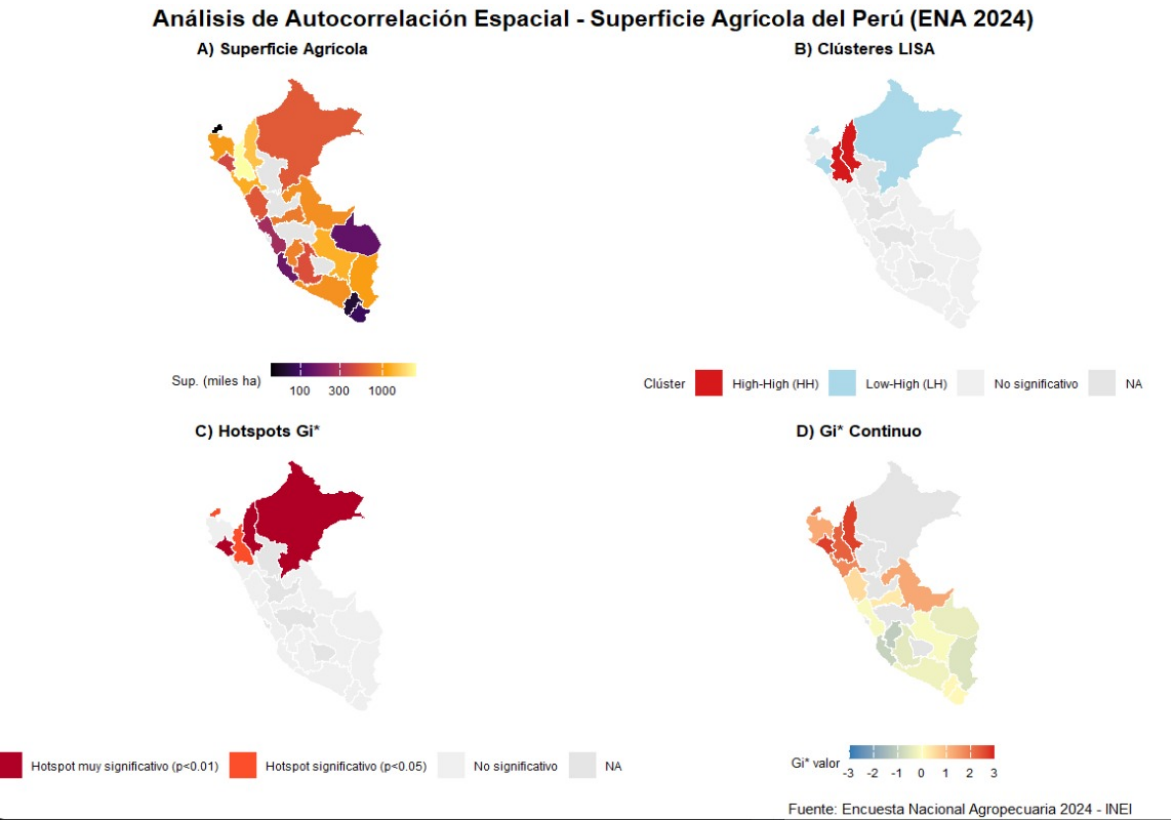


Figura 7: Panel comparativo de análisis espaciales. (A) Superficie agrícola por departamento. (B) Clústeres LISA. (C) Hotspots G_i^* . (D) Valores continuos G_i^* . La comparación visual permite apreciar la consistencia de los patrones identificados por diferentes métodos, todos convergiendo en el norte del Perú. Fuente: ENA 2024.

4.7 Síntesis de Resultados

Hallazgos Principales

1. Autocorrelación espacial positiva moderada confirmada: Moran's I = 0.1664 ($p = 0,0364$)

2. Núcleo productivo identificado: Clúster High-High de 4 departamentos (Cajamarca, La Libertad, Lambayeque, Piura) en el norte

3. Hotspots críticos: 6 departamentos con concentraciones extremas, todos en el norte

4. Convergencia metodológica: LISA y G_i^* identifican los mismos patrones

5. Ausencia de coldspots: No hay concentraciones bajas estadísticamente significativas

6. Heterogeneidad nacional: 76 % de departamentos sin patrones significativos

5 Discusión

5.1 Interpretación de los Patrones Espaciales

Los resultados revelan una estructura espacial compleja en la agricultura peruana, caracterizada por autocorrelación positiva moderada ($I = 0.1664$) y un clúster productivo bien definido en el norte. Este patrón contrasta con el estudio de Hu et al. [9] sobre paperas en China ($I = 0.399$), sugiriendo menor intensidad de agrupamiento en agricultura peruana.

¿Por qué el Moran's I es moderado?

1. **Heterogeneidad geográfica:** Perú presenta tres regiones naturales (Costa, Sierra, Selva) con patrones agrícolas radicalmente diferentes [24]
2. **Concentración en una sola zona:** A diferencia de China con múltiples clústeres [9], Perú tiene esencialmente un clúster dominante (norte)
3. **Escala de análisis:** El nivel departamental puede enmascarar patrones más finos visibles a escalas menores [2]

5.2 Comparación con Literatura Internacional

Similar al estudio de Marruecos [12], los resultados demuestran cómo la autocorrelación espacial identifica zonas críticas para políticas públicas. El clúster High-High del norte peruano podría asociarse a:

- Disponibilidad de recursos hídricos (ríos de vertiente del Pacífico)
- Infraestructura de riego desarrollada
- Proximidad a mercados urbanos (costa norte)
- Condiciones climáticas favorables

El estudio de NDVI en China [13] sugiere que variables ambientales subyacentes explican patrones espaciales. Futuras investigaciones deberían incorporar clima, suelos y disponibilidad hídrica para explicaciones causales.

5.3 Implicaciones para Política Pública

5.3.1 Fortalecimiento del Clúster Norte

El clúster High-High debe considerarse activo estratégico nacional. Políticas recomendadas:

1. **Consolidación competitiva:** Inversión en riego tecnificado y vías de acceso
2. **Innovación tecnológica:** Transferencia tecnológica y semillas mejoradas
3. **Integración de cadenas de valor:** Desarrollo de agroindustrias locales

5.3.2 Desarrollo de Zonas Low-High

Los 2 departamentos Low-High (Tumbes, Amazonas) representan oportunidades de expansión:

- Acceso a conocimiento técnico mediante spillovers del clúster
- Integración a mercados establecidos
- Inversión estratégica en irrigación

5.3.3 Diversificación Productiva Nacional

La concentración en el norte genera vulnerabilidad. Eventos climáticos extremos (El Niño) afectando el norte impactarían desproporcionadamente la producción nacional. Recomendaciones:

1. Fomentar nuevos polos productivos en otras regiones
2. Promover cultivos alternativos adaptados a sierra y selva
3. Implementar sistemas de gestión de riesgos climáticos

5.4 Limitaciones del Estudio

1. **Escala departamental:** Puede ocultar heterogeneidad intra-departamental. Análisis a nivel provincial revelarían patrones más detallados
2. **Corte transversal:** Datos de 2024 representan un momento. Análisis longitudinal permitiría evaluar dinámicas temporales
3. **Variable única:** Solo superficie. Productividad, valor de producción o rendimientos podrían revelar patrones diferentes
4. **Matriz k-NN:** Diferentes especificaciones (contigüidad, distancia) podrían alterar resultados
5. **Análisis descriptivo:** No incorpora variables explicativas. Regresión espacial identificaría determinantes causales

5.5 Futuras Líneas de Investigación

- **Análisis multitemporal:** Comparar ENA 2024 con censos anteriores para evaluar dinámicas
- **Modelado espacial explicativo:** Regresión espacial con variables climáticas y económicas
- **Análisis multiescalar:** Replicar a nivel provincial y distrital
- **Análisis por cultivo:** Examinar patrones específicos (café, cacao, arroz, quinua)
- **Integración con datos satelitales:** Combinar con NDVI para validación

6 Conclusiones

El presente estudio aplicó métodos de autocorrelación espacial a los datos ENA 2024 para caracterizar patrones de distribución agrícola en Perú. Las principales conclusiones son:

1. **Existencia de autocorrelación espacial positiva confirmada:** Moran's $I = 0.1664$ ($p = 0,0364$) demuestra que la distribución agrícola no es aleatoria. Departamentos con superficies similares tienden a agruparse geográficamente.
2. **Identificación del núcleo agrícola nacional:** Análisis LISA identificó un clúster High-High de 4 departamentos en el norte (Cajamarca, La Libertad, Lambayeque, Piura), constituyendo el corredor productivo principal del país.
3. **Validación mediante múltiples métodos:** Getis-Ord G_i^* confirmó 6 hotspots en la misma región norte. La convergencia entre LISA y G_i^* valida la robustez de los hallazgos.
4. **Estructura espacial heterogénea:** 76 % de departamentos sin patrones significativos evidencia heterogeneidad, con agricultura altamente concentrada en el norte y dispersa en el resto.
5. **Ausencia de clústeres Low-Low:** No se detectaron concentraciones bajas estadísticamente significativas, indicando que zonas con agricultura limitada están dispersas.
6. **Zonas de transición identificadas:** 2 departamentos Low-High en periferia del clúster principal representan potencial de desarrollo agrícola.
7. **Evidencia para política pública:** Los resultados proporcionan base cuantitativa para políticas territorialmente diferenciadas: fortalecimiento del clúster norte, desarrollo de zonas periféricas, y diversificación para reducir vulnerabilidad.

6.1 Contribución del Estudio

Este trabajo contribuye a la literatura sobre geografía agrícola peruana al proporcionar la primera caracterización cuantitativa rigurosa de patrones espaciales agrícolas basada en métodos estadísticos formales (Moran's I , LISA, G_i^*) aplicados a datos nacionales representativos (ENA 2024).

La metodología puede replicarse para otros sectores (ganadería, forestación) y escalas (provincial, distrital), contribuyendo a una comprensión más profunda de las dinámicas territoriales del desarrollo agrícola peruano.

6.2 Reflexión Final

Los métodos de autocorrelación espacial demostraron ser herramientas poderosas para revelar patrones subyacentes en la distribución agrícola peruana. La convergencia de tres estadísticos complementarios proporciona una caracterización robusta y multi-dimensional de la estructura espacial agrícola nacional.

Los hallazgos subrayan la importancia de incorporar la dimensión espacial en el análisis de políticas agrarias, reconociendo que la ubicación geográfica no es un atributo neutral sino un factor estructurante de las dinámicas productivas.

Referencias

1. P. A. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, no. 1/2, pp. 17-23, 1950.
2. L. Anselin, "Local indicators of spatial association—LISA," *Geographical Analysis*, vol. 27, no. 2, pp. 93-115, 1995.
3. A. Getis and J. K. Ord, "The analysis of spatial association by use of distance statistics," *Geographical Analysis*, vol. 24, no. 3, pp. 189-206, 1992.
4. W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Economic Geography*, vol. 46, pp. 234-240, 1970.
5. M. F. Goodchild, "The validity and usefulness of laws in geographic information science and geography," *Annals of the Association of American Geographers*, vol. 94, no. 2, pp. 300-303, 2004.
6. L. Anselin, J. Gallo, and H. Jayet, "Spatial panel econometrics," in *The Econometrics of Panel Data*, L. Matyas and P. Sevestre, Eds. Berlin: Springer, 2008, pp. 625-660.
7. Instituto Nacional de Estadística e Informática (INEI), *Encuesta Nacional Agropecuaria 2024: Resultados Generales*. Lima, Perú: INEI, 2024.
8. R. S. Bivand, E. Pebesma, and V. Gómez-Rubio, *Applied Spatial Data Analysis with R*, 2nd ed. New York: Springer, 2013.
9. Y. Hu, L. Zhang, Q. Wang, and X. Chen, "Spatial distribution and determinants of mumps incidence in China," *Frontiers in Public Health*, vol. 13, art. 1637288, 2025.
10. J. F. Wang, X. H. Zhang, and C. D. Xu, "Spatiotemporal transmission and risk factors of mumps in China, 2005–2016," *International Journal of Infectious Diseases*, vol. 75, pp. 61-67, 2016.
11. R. Sharp et al., *InVEST User's Guide*. Stanford, CA: The Natural Capital Project, Stanford University, 2020.
12. B. El Mahrad, L. Abadie, E. Zamora-Sillero, and C. M. Duarte, "Spatial autocorrelation analysis of coastal exposure to sea-level rise in Morocco," *Environmental Advances*, vol. 2, art. 100265, 2025.
13. H. Yan et al., "Spatiotemporal correlation of NDVI with climate variables in China," *Ecological Indicators*, vol. 171, art. 113193, 2025.
14. L. Anselin, A. K. Bera, R. Florax, and M. J. Yoon, "Simple diagnostic tests for spatial dependence," *Regional Science and Urban Economics*, vol. 26, no. 1, pp. 77-104, 1996.
15. E. Schmidtner, C. Lippert, B. Engler, A. M. Häring, M. Aurbacher, and S. Dabbert, "Spatial distribution of organic farming in Germany: does neighbourhood matter?," *European Review of Agricultural Economics*, vol. 39, no. 4, pp. 661-683, 2012.

16. L. Kish, *Survey Sampling*. New York: John Wiley & Sons, 1965.
17. D. O'Sullivan and D. J. Unwin, *Geographic Information Analysis*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2010.
18. R. S. Bivand and D. W. S. Wong, "Comparing implementations of global and local indicators of spatial association," *TEST*, vol. 27, no. 3, pp. 716-748, 2018.
19. R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2024. [Online]. Available: <https://www.R-project.org/>
20. E. Pebesma, "Simple features for R: Standardized support for spatial vector data," *The R Journal*, vol. 10, no. 1, pp. 439-446, 2018.
21. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag, 2016.
22. R. J. Hijmans, J. van Etten, M. Sumner, J. Cheng, D. Baston, A. Bevan, R. Bivand, L. Busetto, M. Canty, A. Fasoli, K. Forrest, A. Ghosh, D. Golicher, J. Gray, P. Greenberg, P. Hiemstra, I. Karney, M. Mattiuzzi, S. Mosher, J. Nowosad, E. Pebesma, O. L. Lamigueiro, E. B. Racine, B. Rowlingson, A. Shortridge, B. Venables, and R. Wueest, *terra: Spatial Data Analysis*. R package version 1.7-78, 2024.
23. Ministerio de Desarrollo Agrario y Riego (MINAGRI), *Plan Nacional de Desarrollo Agrario 2023-2027*. Lima, Perú: MINAGRI, 2023.
24. H. Maletta, *La pequeña agricultura familiar en el Perú: Una tipología microrregionalizada*. Lima, Perú: FAO, 2017.

A Código R Utilizado

A continuación se presenta el código principal utilizado para el análisis:

```
# =====  
# ANÁLISIS DE AUTOCORRELACIÓN ESPACIAL  
# Superficie Agrícola del Perú - ENA 2024  
# =====  
  
# Cargar librerías  
library(haven)  
library(dplyr)  
library(sf)  
library(spdep)  
library(ggplot2)  
  
# Leer datos  
cultivos <- read_sav("01_CAP100A_04.sav")  
caratula <- read_s932 & ha \\ ---- CONTINUA --- SE ADJUNTO UN WORD DE DICHO CÓDIGO
```