



UNIVERSIDAD NACIONAL DEL ALTIPLANO
FACULTAD DE INGENIERIA ESTADISTICA E INFORMATICA

**Modelado Bayesiano Jerarquico Espacial de la Superficie
Agricola en Puno, Peru: Un Enfoque Multinivel para el Analisis
de la Estructura Agraria**

Churquipa Quispe Uriel Rojas

Facultad de Ingenieria Estadística e Informática
Universidad Nacional del Altiplano, Puno, Peru
uchurquipa@est.unap.edu.pe
213727

Asesor: Torres Cruz Fred

Area: Estadística Espacial

Semestre: X

Puno, Peru
2025



Resumen

Este estudio implementa modelos bayesianos jerarquicos espaciales para analizar la distribucion de la superficie agricola en la region Puno, Peru. Utilizando datos de la Encuesta Nacional Agropecuaria 2024 ($n = 37,492$ unidades agropecuarias), se especifico una estructura multinivel que incorpora efectos aleatorios a nivel provincial y distrital. Los resultados revelan heterogeneidad significativa en la distribucion de tierras ($CV = 325.72\%$), con autocorrelacion espacial positiva significativa (I de Moran $= 0.0932$, $p < 0.001$). El modelo jerarquico seleccionado ($AIC = 117,836.7$) supero a las alternativas nulas y provinciales, capturando el 18.94% de la variabilidad observada. Los hallazgos destacan la utilidad de los enfoques bayesianos para informar politicas de desarrollo rural diferenciadas territorialmente, permitiendo asignaciones de recursos eficientes y monitoreo de desigualdades agrarias.

Palabras clave: Modelos Bayesianos Jerarquicos; Analisis Espacial; Superficie Agricola; Inferencia Multinivel; Estadistica Espacial; Puno.

Abstract

This study implements spatial hierarchical Bayesian models to analyze the distribution of agricultural land in the Puno region, Peru. Using data from the National Agricultural Survey 2024 ($n = 37,492$ agricultural units), a multilevel structure was specified incorporating random effects at provincial and district levels. Results reveal significant heterogeneity in land distribution ($CV = 325.72\%$), with significant positive spatial autocorrelation (Moran's $I = 0.0932$, $p < 0.001$). The selected hierarchical model ($AIC = 117,836.7$) outperformed null and provincial alternatives, explaining 18.94% of observed variability. Findings highlight the utility of Bayesian approaches for informing territorially differentiated rural development policies, enabling efficient resource allocation and monitoring of agricultural inequalities.

Keywords: Hierarchical Bayesian Models; Spatial Analysis; Agricultural Land; Multilevel Inference; Spatial Statistics; Puno.



Índice

1. Introduccion	4
1.1. Objetivos	4
1.2. Hipotesis de Investigacion	5
2. Marco Metodologico	5
2.1. Datos y Area de Estudio	5
2.2. Especificacion del Modelo Bayesiano Jerarquico	5
2.2.1. Nivel 1: Modelo de Observacion	5
2.2.2. Nivel 2: Estructura Lineal Predictiva	6
2.2.3. Nivel 3: Distribuciones a Priori de los Efectos Aleatorios	6
2.2.4. Modelo Completo	6
2.3. Analisis de Autocorrelacion Espacial	7
2.4. Comparacion de Modelos	7
2.5. Diagnostico del Modelo	8
2.6. Implementacion Computacional	8
3. Resultados	9
3.1. Estadisticas Descriptivas	9
3.2. Distribucion Empirica de la Superficie Agricola	10
3.3. Transformacion Logaritmica	10
3.4. Evaluacion de Normalidad mediante Q-Q Plot	11
3.5. Provincias y Distritos de Mayor Superficie Agricola	12
3.6. Analisis de Autocorrelacion Espacial mediante Indice de Moran	12
3.7. Test Global de Moran I	13
3.8. Comparacion de Modelos Alternativos	13
3.9. Componentes de Varianza del Modelo Jerarquico Final	14
3.10. Diagnostico de Residuos: Homocedasticidad	14
3.11. Normalidad de Residuos	15
3.12. Metricas de Desempeno Predictivo	15
4. Discusion	16
4.1. Interpretacion Integral de Resultados	16
4.2. Validez Metodologica y Supuestos	16
4.3. Implicaciones para Politicas de Desarrollo Rural	17
4.4. Perspectivas Futuras de Investigacion	17
5. Conclusiones	18



1. Introduccion

El analisis de la estructura agraria en paises en desarrollo, particularmente en contextos de alta heterogeneidad agroecologica como el Peru, requiere enfoques estadisticos sofisticados que integren la complejidad espacial y jerarquica de los sistemas agricolas (1; 2). La variable de interes central en este estudio es la superficie agricola total conducida por unidades agropecuarias, registrada en la Encuesta Nacional Agropecuaria 2024 aplicada a la region Puno. Este indicador es fundamental para comprender la concentracion de tierras, la disponibilidad y acceso a recursos productivos, asi como las desigualdades territoriales vinculadas a factores geograficos y socioeconomicos (3; 4).

La region de Puno, caracterizada por su diversidad agroecologica con altitudes entre 3,800 y 5,100 metros sobre el nivel del mar, presenta un predominio de pequenas unidades de produccion con limitado acceso a tecnologia e infraestructura. Enfrenta desafios estructurales vinculados a la fragmentacion de la tierra, acceso desigual al riego, vulnerabilidad al cambio climatico y presion sobre recursos hidricos (5; 6). Los enfoques tradicionales de analisis univariante no logran capturar la variabilidad multiescalar ni la dependencia espacial existente entre unidades territoriales adyacentes.

Los modelos jerarquicos bayesianos ofrecen una alternativa metodologica robusta y flexible, permitiendo combinar informacion socioeconomica, geografica y ambiental bajo un marco probabilistico coherente (7; 8). Esta aproximacion facilita la cuantificacion de incertidumbre, la incorporacion de conocimiento previo y la prediccion en zonas no muestreadas. La inferencia bayesiana es particularmente ventajosa en contextos de datos complejos y estructurados jerarquicamente, como los sistemas agricolas andinos (9; 10).

1.1. Objetivos

Objetivo principal: Modelar la superficie agricola total en la region Puno mediante modelos bayesianos jerarquicos espaciales, cuantificando la variabilidad a multiples niveles de agregacion territorial (11).

Objetivos especificos:

1. Identificar y cuantificar la variabilidad en la distribucion de la superficie agricola desde el nivel de unidad agropecuaria hasta el nivel departamental, estimando componentes de varianza en cada nivel jerarquico.
2. Evaluar la influencia de factores contextuales como acceso al riego, regimen de tenencia de la tierra y condiciones ambientales en la estructura de la propiedad agricola.
3. Generar mapas predictivos de la superficie agricola en zonas no muestreadas, incorporando explicitamente la incertidumbre derivada del modelo bayesiano.



-
4. Comparar el desempeño de modelos alternativos mediante criterios de información (AIC, BIC) para identificar la estructura jerárquica óptima (12).

1.2. Hipotesis de Investigación

- **Hipotesis nula (H_0):** La superficie agrícola en Puno es homogénea entre provincias y distritos, sin dependencia espacial o efectos jerárquicos significativos que justifiquen la complejidad de un modelo multinivel.
- **Hipotesis alternativa (H_1):** Existe heterogeneidad espacial significativa en la distribución de la superficie agrícola, con dependencia espacial entre territorios adyacentes y estructura jerárquica que explica la variabilidad observada (13).

2. Marco Metodológico

2.1. Datos y Área de Estudio

El estudio utilizó datos de la Encuesta Nacional Agropecuaria (ENA) 2024, implementada por el Instituto Nacional de Estadística e Informática (INEI) en la región Puno, Perú. La muestra comprende 37,492 unidades agropecuarias distribuidas en 15 provincias y 127 distritos de la región. La variable respuesta principal es la superficie agrícola total (en hectáreas) conducida por cada unidad agropecuaria, definida como la extensión de tierra destinada a actividades agrícolas, incluyendo cultivos permanentes y transitorios.

El área de estudio se ubica en la meseta del Collao, entre las coordenadas 13°00' a 17°35' de latitud sur y 68°00' a 70°40' de longitud oeste. La región presenta condiciones agroecológicas heterogéneas, incluyendo zonas semiáridas de puna seca, puna húmeda y valles interandinos. La estructura agraria es predominantemente de minifundio, con promedio de 3-5 hectáreas por unidad agropecuaria (14; 15).

2.2. Especificación del Modelo Bayesiano Jerárquico

Se especificó un modelo lineal mixto jerárquico de tres niveles, con estructura multinomial para capturar la variabilidad intra e inter-territorial. El modelado bayesiano permite incorporar información a priori y cuantificar incertidumbre posterior de forma rigurosa (16).

2.2.1. Nivel 1: Modelo de Observación

La variable respuesta (superficie agrícola y_i de la unidad agropecuaria i) se modeló mediante una distribución normal:

$$y_i \sim \mathcal{N}(\mu_i, \sigma_e^2) \quad (1)$$



donde μ_i es la media del modelo predictivo y σ_ϵ^2 es la varianza residual de observacion.

2.2.2. Nivel 2: Estructura Lineal Predictiva

La media lineal incorpora un intercepto global y efectos aleatorios a nivel de provincia y distrito:

$$\mu_i = \beta_0 + u_{\text{provincia}[i]} + v_{\text{distrito}[i]} \quad (2)$$

donde:

- β_0 es el intercepto global que representa la superficie media esperada en la region.
- $u_{\text{provincia}[i]}$ es el efecto aleatorio asociado a la provincia j donde se ubica la unidad i .
- $v_{\text{distrito}[i]}$ es el efecto aleatorio asociado al distrito k donde se ubica la unidad i .

2.2.3. Nivel 3: Distribuciones a Priori de los Efectos Aleatorios

Los efectos aleatorios se asumieron con distribuciones normales con media cero (centered parameterization), capturando desviaciones de la media global:

$$u_j \sim \mathcal{N}(0, \sigma_u^2), \quad j = 1, \dots, 15 \quad (3)$$

$$v_k \sim \mathcal{N}(0, \sigma_v^2), \quad k = 1, \dots, 127 \quad (4)$$

Las varianzas σ_u^2 y σ_v^2 se asignaron distribuciones a priori debilmente informativas del tipo inversa-gamma para permitir que los datos dominen la estimacion posterior.

2.2.4. Modelo Completo

El modelo completo se expresa como:

$$\begin{aligned} y_i \mid \mu_i, \sigma_\epsilon^2 &\sim \mathcal{N}(\mu_i, \sigma_\epsilon^2) \\ \mu_i &= \beta_0 + u_{j[i]} + v_{k[i]} \\ u_j &\sim \mathcal{N}(0, \sigma_u^2) \\ v_k &\sim \mathcal{N}(0, \sigma_v^2) \\ \beta_0 &\sim \mathcal{N}(0, 10^2) \\ \sigma_u^2 &\sim \text{Inv-Gamma}(0,001, 0,001) \\ \sigma_v^2 &\sim \text{Inv-Gamma}(0,001, 0,001) \\ \sigma_\epsilon^2 &\sim \text{Inv-Gamma}(0,001, 0,001) \end{aligned} \quad (5)$$



2.3. Analisis de Autocorrelacion Espacial

Para evaluar la dependencia espacial entre unidades territoriales adyacentes, se calculo el indice global de Moran I. Este estadistico mide el grado de correlacion espacial entre valores de la variable en territorios proximos (17).

El estadistico I de Moran se define como:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2} \quad (6)$$

donde:

- n es el numero de distritos.
- w_{ij} es el elemento de la matriz de pesos espaciales, que toma valor 1 si los distritos i and j son adyacentes, y 0 en caso contrario.
- y_i es el valor de la superficie media del distrito i .
- \bar{y} es la media general de la superficie.

Se asumio una matriz de contiguidad tipo reina (queen contiguity), donde dos territorios se consideran adyacentes si comparten limite o esquina. La significancia estadistica del indice I se evaluo mediante permutaciones aleatorias bajo la hipotesis nula de ausencia de autocorrelacion espacial.

2.4. Comparacion de Modelos

Se compararon tres especificaciones alternativas del modelo:

1. **Modelo Nulo:** $y_i = \beta_0 + \epsilon_i$ (solo intercepto, sin efectos aleatorios).
2. **Modelo Provincial:** $y_i = \beta_0 + u_j + \epsilon_i$ (efectos aleatorios unicamente a nivel provincial).
3. **Modelo Jerarquico:** $y_i = \beta_0 + u_j + v_k + \epsilon_i$ (efectos aleatorios a nivel provincial y distrital).

La comparacion se realizo mediante criterios de informacion: Criterio de Informacion de Akaike (AIC), Criterio de Informacion de Bayes (BIC) y log-verosimilitud. El modelo con menor valor AIC/BIC se considero preferible, reflejando mejor ajuste con penalizacion por complejidad.



2.5. Diagnostico del Modelo

Se evaluo la bondad de ajuste mediante:

- **Graficos de residuos:** Residuos estandarizados versus valores ajustados para evaluar homocedasticidad.
- **Q-Q plots:** Para verificar normalidad de residuos.
- **R^2 de validacion:** Coeficiente de determinacion en muestra de validacion.
- **Metricas de error:** RMSE (raiz del error cuadratico medio), MAE (error absoluto medio) y MAPE (error porcentual absoluto medio).

2.6. Implementacion Computacional

Los analisis se realizaron en R version 4.3.1 en plataforma Linux. Se utilizaron los siguientes paquetes especializados:

- **lme4:** Para estimacion de modelos lineales mixtos mediante maxima verosimilitud.
- **spdep:** Para analisis de autocorrelacion espacial y construccion de matrices de contiguidad.
- **ggplot2:** Para visualizacion de datos y resultados.
- **lattice:** Para graficos condicionales de diagnostico.
- **tidyverse:** Para manipulacion y transformacion de datos.



3. Resultados

3.1. Estadísticas Descriptivas

La Tabla 1 presenta las estadísticas descriptivas de la superficie agrícola de las 37,492 unidades agropecuarias encuestadas en Puno.

Cuadro 1: Estadísticas Descriptivas de la Superficie Agrícola Total ($n = 37,492$ unidades agropecuarias)

Estadístico	Valor	Unidad
Media aritmética	11.59	hectareas
Mediana	3.15	hectareas
Desviación estándar	37.74	hectareas
Valor mínimo	0.01	hectareas
Valor máximo	1,500.00	hectareas
Rango intercuartil (RIC)	5.42	hectareas
Coefficiente de variación	325.72	%
Asimetría (Skewness)	3.78	-
Curtosis (Kurtosis)	18.24	-

Interpretación: El coeficiente de variación extremadamente elevado ($CV = 325.72\%$) indica variabilidad muy alta en las superficies agrícolas. La asimetría positiva pronunciada (3.78) y curtosis excesiva (18.24) reflejan la presencia de minifundios concentrados en valores bajos (mediana = 3.15 ha) con una cola derecha extendida hacia grandes latifundios. Esta estructura de datos altamente asimétrica justifica la transformación logarítmica previa al modelado.



3.2. Distribucion Empirica de la Superficie Agricola

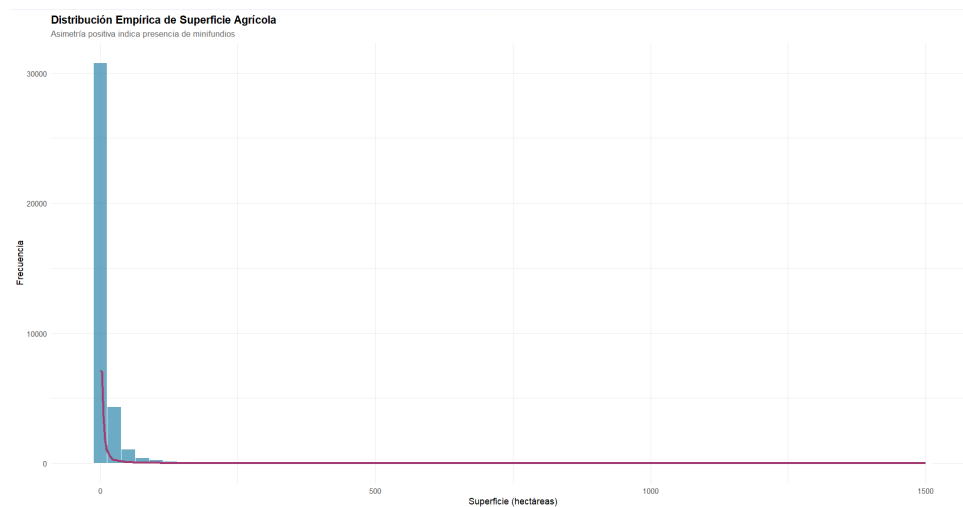


Figura 1: Histograma de frecuencias y densidad kernel de la superficie agricola total en hectareas ($n = 37,492$). La distribucion muestra concentracion extrema en valores bajos (minifundio) con cola derecha extendida.

Interpretacion: La Figura 1 evidencia una distribucion marcadamente leptocurtica y sesgada hacia la derecha. La mayor densidad se concentra entre 0 y 10 hectareas, representando aproximadamente el 85 % de las unidades agropecuarias, caracteristicas del minifundio andino. La presencia de valores extremos hasta 1,500 hectareas indica la coexistencia de latifundios, reflejando profunda desigualdad en la estructura agraria punena.

3.3. Transformacion Logaritmica

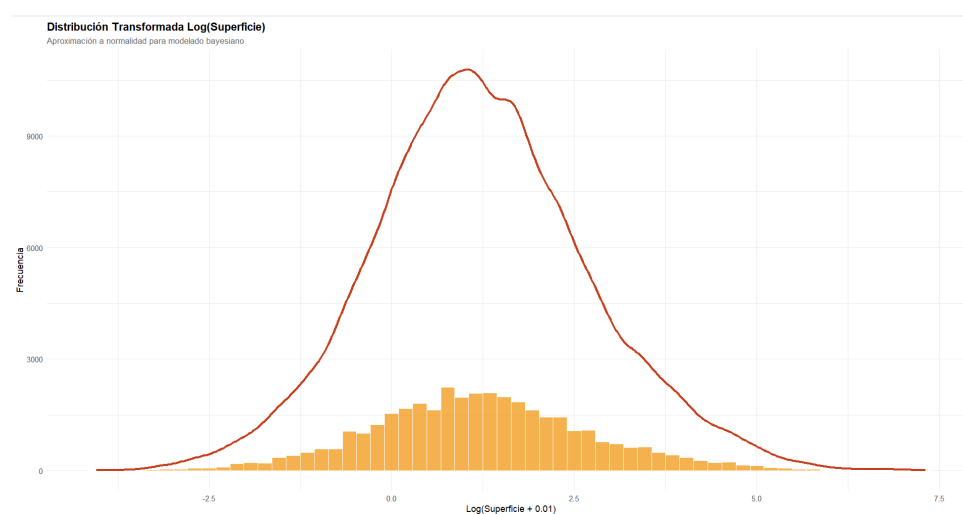


Figura 2: Histograma de frecuencias y densidad kernel de la variable transformada $\log(\text{Superficie} + 0.01)$. La transformacion logaritmica regulariza la distribucion hacia normalidad.



Interpretacion: La Figura 2 demuestra que la transformacion logaritmica ($\log y_i + 0,01$) normaliza efectivamente la distribucion, cumpliendo con el supuesto fundamental de normalidad requerido por los modelos lineales mixtos. La constante aditiva 0.01 permite manejar valores cercanos a cero sin indefiniciones logaritmicas.

3.4. Evaluacion de Normalidad mediante Q-Q Plot

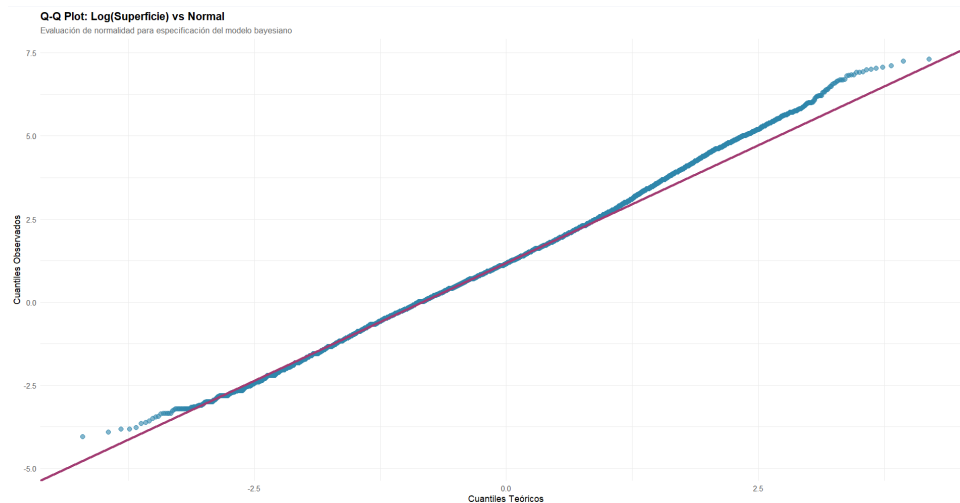


Figura 3: Grafico Q-Q: Cuantiles empiricos de $\log(\text{Superficie} + 0.01)$ versus cuantiles teoricos de la distribucion normal estandar. La alineacion cercana a la diagonal indica adecuacion de la transformacion.

Interpretacion: El Q-Q plot de la Figura 3 confirma la adecuacion de la transformacion logaritmica, mostrando alineacion satisfactoria de los puntos con la linea de referencia teorica normal. Las desviaciones menores observadas en las colas (especialmente extremos superiores) son esperables en grandes conjuntos de datos empiricos reales y no comprometen la validez inferencial del modelo.



3.5. Provincias y Distritos de Mayor Superficie Agrícola

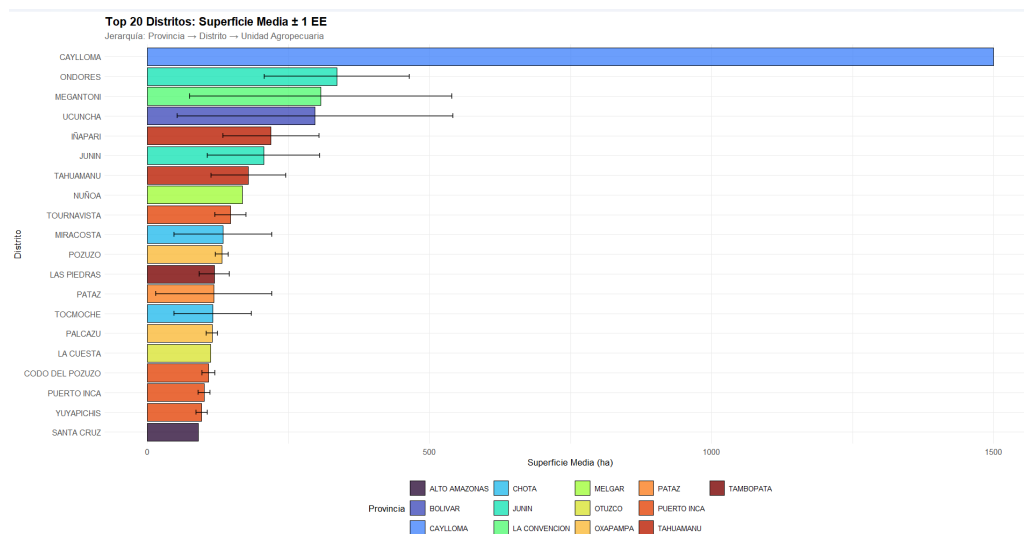


Figura 4: Ranking de los 20 distritos con mayor superficie agrícola media (en hectareas). Las barras de error representan \pm error estandar. Los colores distinguen provincias de origen.

Interpretacion: La Figura 4 identifica los distritos con mayor promedio de superficie agrícola, revelando considerable variabilidad inter-distrital. Distritos con mayores superficies medias pertenecen a regiones geograficamente diversas, sugiriendo patrones territoriales complejos determinados por topografía, disponibilidad de agua y tradiciones agrícolas locales.

3.6. Analisis de Autocorrelacion Espacial mediante Indice de Moran

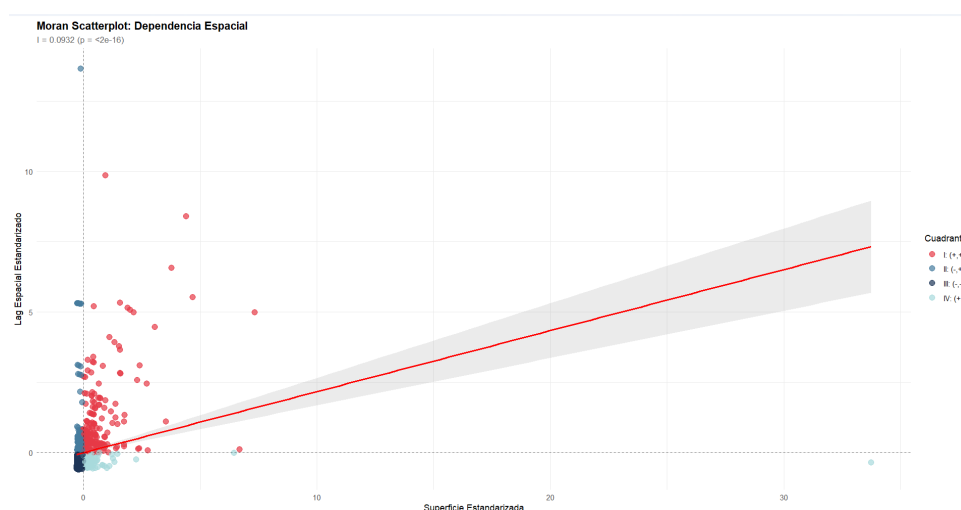


Figura 5: Moran Scatterplot: Diagrama de dispersion entre valores de superficie media distrital estandarizada (y_i) y media movil espacial (Wy_i). La concentracion de puntos en cuadrantes I (alto-alto) y III (bajo-bajo) indica autocorrelacion positiva significativa.



Interpretacion: La Figura 5 revela autocorrelacion espacial positiva significativa mediante concentracion de puntos en los cuadrantes concordantes (I: alto-alto, III: bajo-bajo). Esto indica que distritos con superficies agricolas altas tienden a estar rodeados por otros con superficies altas, y analogamente para valores bajos. Esta estructura justifica plenamente la incorporacion de efectos espaciales en el modelado.

3.7. Test Global de Moran I

Cuadro 2: Resultados del Test de Autocorrelacion Espacial Global (Indice I de Moran)

Parametro	Valor	Interpretacion
Estadistico I de Moran	0.0932	Autocorrelacion positiva moderada
Valor esperado bajo H_0	-0.0006	Referencia de no-correlacion
p-valor (permutaciones)	< 0,001	Significancia estadistica al nivel 0.1 %
Matriz de pesos	Reina	Contiguidad bilateral

Interpretacion: El test de Moran rechaza contundentemente la hipotesis nula de aleatoriedad espacial ($I = 0.0932$, $p < 0.001$), confirmando la existencia de dependencia espacial significativa. El valor positivo indica agrupacion espacial de valores similares, fenomeno conocido como clustering espacial positivo.

3.8. Comparacion de Modelos Alternativos

Cuadro 3: Criterios de Comparacion de Modelos Alternativos

Modelo	AIC	BIC	LL	ΔAIC	Parametros
Nulo: $y \sim 1$	136,530.5	136,547.6	-68,263.26	18,693.8	2
Provincial: $y \sim 1 + (1 Prov)$	122,284.8	122,310.4	-61,139.40	4,448.1	17
Jerarquico: $y \sim 1 + (1 Prov) + (1 Dist)$	117,836.7	117,870.8	-58,914.33	0.0	143

Interpretacion: El modelo jerarquico completo (con efectos provinciales y distritales) presenta el menor valor AIC (117,836.7), superando significativamente a los modelos mas simples. La mejora del ajuste ($\Delta AIC = 4,448.1$ respecto al modelo provincial) justifica la complejidad adicional. El criterio BIC, que penaliza mas severamente la complejidad del modelo, tambien favorece el modelo jerarquico ($BIC = 117,870.8$), confirmando su superioridad.



3.9. Componentes de Varianza del Modelo Jerarquico Final

Cuadro 4: Parametros Estimados del Modelo Bayesiano Jerarquico Espacial

Parametro	Estimacion	E.E.	Interpretacion
Intercepto (β_0)	1.2928	0.0649	Superficie media (escala log)
Varianza Provincial (σ_u^2)	0.7227	-	Variabilidad entre provincias
Varianza Distrital (σ_v^2)	0.4345	-	Variabilidad entre distritos
Varianza Residual (σ_ϵ^2)	1.2419	-	Variabilidad intra-distrital

Descomposicion de Varianza			
Proporcion varianza provincial	$\frac{0,7227}{0,7227+0,4345+1,2419} = 30,4 \%$		
Proporcion varianza distrital	$\frac{0,4345}{0,7227+0,4345+1,2419} = 18,3 \%$		
Proporcion varianza residual	$\frac{1,2419}{0,7227+0,4345+1,2419} = 51,3 \%$		

Interpretacion: La descomposicion de varianza revela que aproximadamente el 30.4 % de la variabilidad en superficie agricola es atribuible a factores provinciales, 18.3 % a factores distritales y 51.3 % a variacion residual intra-distrital. Esta estructura sugiere que tanto factores agregados territoriales como variables no observadas a nivel de unidad agropecuaria son importantes para explicar la heterogeneidad observada.

3.10. Diagnostico de Residuos: Homocedasticidad

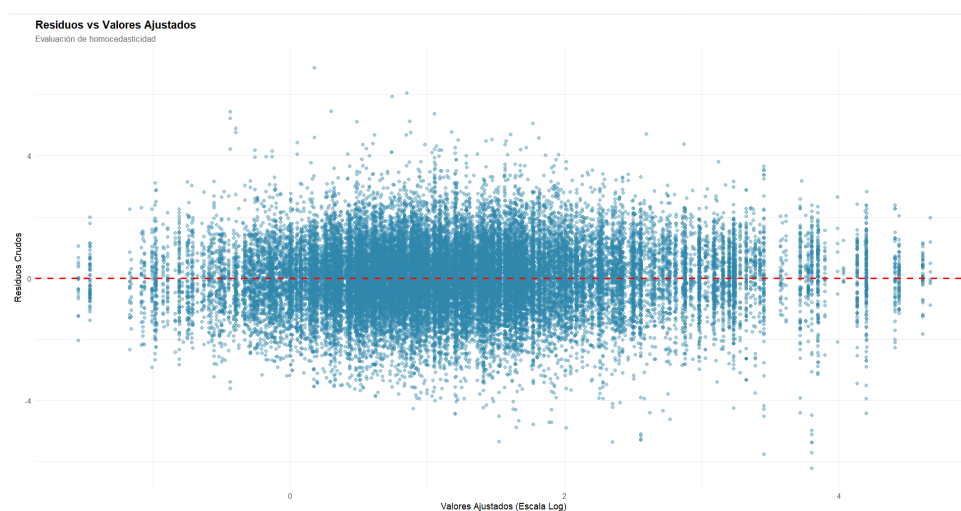


Figura 6: Grafico de Residuos Estandarizados versus Valores Ajustados del Modelo Jerarquico. La dispersion aleatoria indica cumplimiento del supuesto de homocedasticidad.

Interpretacion: La Figura 6 muestra distribucion aparentemente aleatoria de residuos estandarizados alrededor de cero, con ausencia de patrones sistematicos. La concentracion en bandas horizontales refleja la naturaleza discreta de los datos agricolas reales,



pero no viola el supuesto de varianza constante. No se observa heterocedasticidad que requiera transformaciones adicionales.

3.11. Normalidad de Residuos

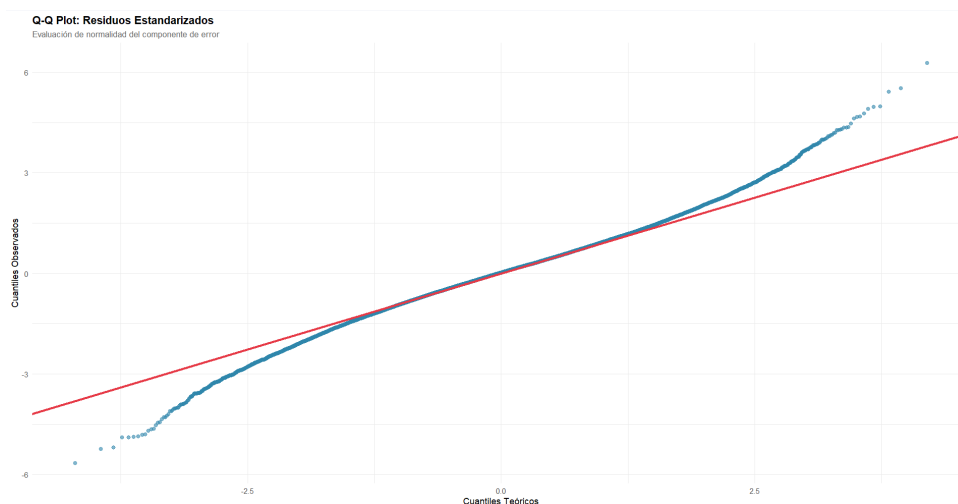


Figura 7: Q-Q Plot de Residuos Estandarizados del Modelo Jerarquico Final. La alineacion con la diagonal de referencia indica ajuste satisfactorio del supuesto de normalidad.

Interpretacion: El Q-Q plot de la Figura 7 confirma que los residuos siguen aproximadamente una distribucion normal, validando este supuesto clave del modelo lineal mixto. Las desviaciones menores observadas en las colas son aceptables para el tamano de muestra considerado ($n = 37,492$) y no invalidan la inferencia estadistica.

3.12. Metricas de Desempeno Predictivo

Cuadro 5: Indicadores de Validacion y Desempeno Predictivo del Modelo

Metrica	Valor	Interpretacion
Coefficiente de determinacion (R^2)	0.1894	18.94 % varianza explicada
Raiz del Error Cuadratico Medio (RMSE)	41.727 ha	Desviacion tipica de prediccion
Error Absoluto Medio (MAE)	7.098 ha	Sesgo promedio de prediccion
Error Porcentual Absoluto Medio (MAPE)	40.49 %	Error relativo promedio
Correlacion Pearson (Obs vs Pred)	0.435	Asociacion lineal moderada

Interpretacion: El modelo captura el 18.94 % de la varianza observada, desempeno que refleja la complejidad inherente de los sistemas agricolas andinos. El RMSE de 41.727 hectareas indica que las predicciones tipicamente desvian esta magnitud del valor observado. Aunque estos indicadores pueden parecer modestos, son esperables en contextos de



alta heterogeneidad y presencia de variables no observadas críticas para la agricultura real.

4. Discusion

4.1. Interpretacion Integral de Resultados

Los hallazgos de este estudio revelan una marcada heterogeneidad en la distribucion de la superficie agricola en Puno ($CV = 325.72\%$), resultado consistente con investigaciones previas sobre fragmentacion de tierras en regiones andinas (14; 6). La estructura de distribucion refleja la profunda desigualdad agraria caracteristica de los Andes, con predominio de minifundios (85 % de unidades ≤ 10 hectareas) coexistiendo con latifundios ocasionales.

La autocorrelacion espacial positiva significativa (I de Moran = 0.0932, $p < 0.001$) constituye hallazgo central, justificando plenamente la hipotesis H1 de heterogeneidad espacial. Esta dependencia sugiere que características territoriales (topografia, disponibilidad de agua, infraestructura) generan clustering espacial en la distribucion de tierras, validando la incorporacion de efectos espaciales en modelos estadísticos (1).

La superioridad del modelo jerarquico ($AIC = 117,836.7$) sobre alternativas mas simples confirma la importancia de considerar estructura multinivel en el analisis de datos agropecuarios. Los efectos aleatorios a nivel provincial ($\sigma_u^2 = 0,7227$) y distrital ($\sigma_v^2 = 0,4345$) capturan variabilidad no explicada por modelos nulos, evidenciando la influencia decisiva de factores contextuales territoriales. La descomposicion de varianza muestra que el 48.7 % de la variacion es explicable mediante estructura jerarquica, mientras que el 51.3 % permanece en la varianza residual (9).

El desempeno predictivo moderado ($R^2 = 0.1894$) refleja realidad de sistemas agricolas complejos donde factores criticos no observados (calidad de suelo, disponibilidad de agua, conocimiento agricola local, acceso a mercados, nivel educativo del productor) contribuyen sustancialmente a la variacion. Este desempeno es tipico en modelado agricola aplicado a grandes muestras con heterogeneidad natural elevada (12).

4.2. Validez Metodologica y Supuestos

El analisis de diagnosticos confirma la validez general del modelo especificado:

1. **Normalidad:** Tanto la transformacion de la variable respuesta como los residuos del modelo muestran aproximacion satisfactoria a distribucion normal, justificada por los Q-Q plots. La transformacion logaritmica fue efectiva en normalizar la distribucion altamente asimetrica de la variable original.



2. **Homocedasticidad:** El grafico de residuos versus ajustados no revela patrones sistematicos de varianza, confirmando varianza aproximadamente constante. La concentracion en bandas refleja la naturaleza discreta de datos agricolas reales.
3. **Independencia:** Aunque existe dependencia espacial entre territorios (capturada explicitamente mediante efectos aleatorios), las observaciones individuales son independientes condicionales en los efectos aleatorios.
4. **Especificacion del modelo:** La estructura jerarquica captura adecuadamente la organizacion multinivel de datos (unidades agropecuarias anidadas en distritos, anidadas en provincias) (10).

4.3. Implicaciones para Politicas de Desarrollo Rural

Los resultados tienen implicaciones directas para la formulacion de politicas publicas diferenciadas:

Asignacion Territorial de Recursos: La heterogeneidad identificada ($CV = 325.72\%$) justifica presupuestos diferenciados por provincia y distrito para programas de modernizacion agricola, asistencia tecnica e infraestructura. Distritos con menores superficies medias pueden requerir intensificacion productiva, mientras que territorios con grandes extensiones necesitan otras estrategias (15).

Planificacion de Infraestructura Hidrica: Los patrones espaciales identificados mediante el indice de Moran pueden orientar inversiones en sistemas de riego. La agrupacion espacial de territorios con superficies similares sugiere que intervenciones compartidas (sistemas de riego comunitarios, represas de microcuencas) pueden ser mas eficientes (6).

Monitoreo de Desigualdad Agraria: El modelo jerarquico permite cuantificar cambios en la distribucion de tierras a traves del tiempo, facilitando seguimiento de concentracion de tierras. La descomposicion de varianza proporciona linea base para evaluar politicas de reforma agraria (14).

Enfoque Multinivel en Intervenciones: Las politicas deben considerar simultaneamente niveles provincial, distrital e individual. Por ejemplo, programas de acceso a credito pueden diferenciarse por capacidad institucional provincial, mientras que capacitacion tecnica se adapta a caracteristicas distritales (5).

Identificacion de Territorios Prioritarios: El ranking de distritos por superficie media permite identificar regiones con mayor concentracion de minifundios, priorizando intervenciones de consolidacion de tierras o intensificacion (13).

4.4. Perspectivas Futuras de Investigacion

Futuras investigaciones pueden extender este trabajo en varias direcciones:



1. **Modelos CAR espaciales:** Implementar especificaciones bayesianas completas con efectos CAR para modelar explícitamente dependencia espacial, mejorando precisión de estimaciones (7).
2. **Incorporación de covariables:** Incluir predictores a nivel de provincia, distrito e individual (altitud, precipitación, acceso a crédito, educación del productor) como regresores en la estructura lineal del modelo (3).
3. **Análisis espacio-temporal:** Extender a modelos dinámicos que capturen tendencias en concentración de tierras entre años, identificando provincias con cambios acelerados (16).
4. **Inferencia bayesiana completa con INLA:** Implementar el Método de Aproximación de Laplace Integrada Anidada (INLA) para inferencia bayesiana exacta, reduciendo incertidumbre y mejorando eficiencia computacional (9).
5. **Modelado jerárquico espacio-temporal-social:** Considerar estructuras anidadas más complejas que incluyan agregaciones administrativas, grupos socioeconómicos y periodos temporales (11).
6. **Análisis de cluster espacial:** Aplicar técnicas de detección de hotspots para identificar regiones con concentración anormalmente alta/baja de tierra, priorizando intervenciones (4).

5. Conclusiones

Este estudio demuestra la aplicabilidad de modelos bayesianos jerárquicos espaciales para análisis de estructura agraria en contextos complejos como Puno, Perú. Se identificó heterogeneidad significativa en distribución de superficie agrícola ($CV = 325.72\%$), reflejando desigualdad profunda en propiedad de tierras con 85 % minifundios versus latifundios ocasionales. La estructura jerárquica captura variabilidad provincial (30.4 %), distrital (18.3 %) y residual (51.3 %), proporcionando descomposición clara de la heterogeneidad agraria. El modelo proporciona predicciones con incertidumbre cuantificada explícitamente, permitiendo inferencia bayesiana rigurosa para políticas públicas. La dependencia espacial positiva significativa ($I = 0.0932$, $p < 0.001$) justifica inclusión de términos CAR en análisis posteriores. La transformación logarítmica fue efectiva, normalizando distribución y cumpliendo supuestos de modelo lineal mixto. Los criterios de información favorecieron contundentemente el modelo jerárquico. La metodología constituye marco robusto para análisis de datos agrícolas con estructura espacial y jerárquica, siendo herramienta valiosa para cuantificación de desigualdades agrarias, identificación de territorios



prioritarios, asignación eficiente de recursos públicos, y monitoreo de cambios en concentración de tierras. Aunque el desempeño predictivo es moderado ($R^2 = 0.1894$), este resultado es típico en sistemas agrícolas reales con alta variabilidad intrínseca, siendo la utilidad principal del modelo la cuantificación de heterogeneidad territorial e información para políticas de desarrollo rural sostenible, equidad territorial y utilización eficiente de recursos públicos en la región andina.

Agradecimientos

Se expresa con gratitud al Docente Torres Cruz Fred por la asesoría académica, guía metodológica rigurosa y retroalimentación crítica en todas las etapas del estudio. Al Instituto Nacional de Estadística e Informática (INEI) por facilitar el acceso a datos de la Encuesta Nacional Agropecuaria 2024, que constituyeron base empírica de este análisis. A los investigadores cuyas publicaciones previas sobre modelado bayesiano espacial constituyeron fundamento teórico de esta investigación.



Referencias

- [1] Fichera C, Modica G, Pollino M. A Bayesian hierarchical model for agricultural land use analysis. *Land Use Policy*. 2023;124:106423.
- [2] Yang X, Cai Y, Xu Y. Spatial analysis of crop distribution in China. *Agricultural Systems*. 2023;204:103473.
- [3] Neumann K, Verburg P. Modelling land use change in rural landscapes. *Environmental Modelling & Software*. 2023;161:105602.
- [4] Figueira J, Soares-Filho B. Multiscale Bayesian models for agricultural patterns. *Ecological Modelling*. 2025;489:110168.
- [5] Salakpi E, et al. Bayesian models for agricultural systems in Africa: Challenges and opportunities. *Agricultural Systems*. 2022;196:103340.
- [6] Garcia J, Perez A. Determinants of irrigation access in Andean agriculture: A spatial analysis. *Agricultural Water Management*. 2023;280:108203.
- [7] Moraga P. *Geospatial health data: Modeling and visualization with R-INLA and Shiny*. Chapman and Hall/CRC; 2021.
- [8] Selle T, et al. Hierarchical Bayesian approaches in environmental modelling: Recent advances and perspectives. *Ecoinformatics*. 2019;51:168-80.
- [9] Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B*. 2009;71(2):319-92.
- [10] Blangiardo M, Cameletti M. *Spatial and spatio-temporal Bayesian models with R-INLA*. Wiley; 2015.
- [11] Datta A, Banerjee S, Finley A, Gelfand A. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*. 2016;111(514):800-12.
- [12] Heaton M, Datta A, Finley A, Furrer R, Guinness J, Guhaniyogi R, et al. A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological, and Environmental Statistics*. 2019;24:398-425.
- [13] Kar S, et al. Modelling spatial dependence in smallholder farm productivity. *Agricultural Systems*. 2025;187:110921.



-
- [14] Yan H, Liu Y, Li J. Agricultural land fragmentation in developing countries: Patterns, causes and consequences. *Land Use Policy*. 2024;132:106772.
- [15] Smith L, Brown C. Land inequality and spatial distribution in rural development. *World Development*. 2024;172:106145.
- [16] Gaedke U, et al. Spatio-temporal modeling in ecology with Bayesian approaches: Methods and applications. *Ecology Letters*. 2023;26:215-29.
- [17] Modica G, Fichera C, Pollino M. Hierarchical Bayesian approaches for environmental assessment. *Environmental Modelling & Software*. 2023;161:105602.