Modelado Espacio-Temporal de Patrones Agrícolas en Perú con Procesos de Cox: Enfoque basado en ENA 2022-2024

Yhack Bryan Aycaya Paco[1]0009-0000-5397-2784

Universidad Nacional del Altiplano, Puno, Perú,
yaycaya@est.unap.edu.pe

Abstract. Este estudio modela patrones agrícolas en Perú utilizando procesos log-Gaussianos de Cox (LGCP) con datos de la Encuesta Nacional Agropecuaria 2022-2024. El objetivo fue analizar la distribución espacio-temporal de 174,040 eventos agrícolas, considerando covariables como fuente de agua y superficie. La metodología incluyó estimación de intensidad mediante suavizado por kernel ($\sigma = 0.0009$) y modelado LGCP con covarianza exponencial. Los resultados revelaron clustering significativo ($L(r) > 0$ para $r < 1.0°$), con mayor intensidad en la Sierra (524 puntos/unidad^2) frente a la Costa (194) y la Selva (168). El modelo LGCP arrojó coeficientes significativos ($\beta_0 = 6.8006$, $\beta_1 = -0.3957$, $\beta_2 = -0.7300$, $p < 0.001$, AIC = -810,737), pero baja varianza explicada ($R^2 = 0.0535$). Se concluye que los LGCP son efectivos para identificar patrones agrícolas, recomendando integrar datos climáticos y satelitales para mejorar la planificación agrícola sostenible.

Keywords: Procesos log-Gaussianos de Cox, patrones agrícolas, análisis espacio-temporal, ENA 2024, clustering espacial

1 Introducción

El presente trabajo se enfoca en el modelado de patrones agrícolas en Perú mediante procesos de Cox espacio-temporales, utilizando procesos log-Gaussianos de Cox (LGCP) aplicados a los datos de la Encuesta Nacional Agropecuaria 2024 (ENA 2024). Este enfoque es relevante debido a su impacto en la seguridad alimentaria, su contribución a la estadística espacial y su aplicación en la planificación agrícola sostenible. La agricultura peruana enfrenta desafíos significativos como la variabilidad climática, la intensificación del uso de la tierra y riesgos hidrológicos, como inundaciones [13, 23], que afectan de manera diferenciada a las regiones de Costa, Sierra y Selva. La ENA 2024 proporciona datos detallados, incluyendo coordenadas espaciales (LATITUD, LONGITUD), fechas de actividades agrícolas (P124_MES, P124_ANIO) y covariables como la superficie de parcela (P105_SUP_ha), lo que permite un análisis espacio-temporal robusto [5, 6]. A diferencia de disciplinas como sismología [6], epidemiología [9, 20], análisis de crímenes [22], dinámica de incendios [7] y predicción de eventos urbanos [15],

donde los modelos espacio-temporales son más comunes, su aplicación en contextos agrícolas sigue siendo menos explorada, justificando la necesidad de este estudio.

2 Metodología

En esta sección, describimos la metodología empleada para el estudio de la producción agrícola en Perú y el desarrollo del modelado con procesos de Cox.

2.1 Recolección y Procesamiento de Datos

Los datos fueron obtenidos de la Encuesta Nacional Agropecuaria (ENA) 2024, promovida por el Instituto Nacional de Estadística e Informática (INEI) de Perú [11]. La base de datos completa se descargó del sistema de microdatos del INEI [12], que ofrece acceso a encuestas en formatos como SPSS, Excel y CSV, garantizando el secreto estadístico. Además, se consultó el concurso nacional para investigaciones sobre resultados de la ENA, organizado por el Proyecto "Mejoramiento del Sistema de Información Estadística Agraria" [10], para contextualizar su uso en análisis académicos. La ENA 2024 incluye información sobre características de unidades agropecuarias, producción, costos y prácticas agrícolas, recolectada mediante cuestionarios a productores naturales y empresas.

Posteriormente, se realizó un filtrado de los datos, seleccionando las variables relevantes para el análisis de patrones agrícolas y eliminando registros incompletos o no relacionados. Las variables asociadas a peso, como la cantidad producida (P219_CANT_1), se estandarizaron a kilogramos para asegurar uniformidad en las unidades. Se verificó y corrigió la consistencia de las coordenadas LONGITUD y LATITUD, descartando valores fuera del rango geográfico de Perú (entre -82° y -68° W, y -19° y 0.5° S). Finalmente, se llevó a cabo una limpieza de datos no relacionados, eliminando duplicados y anomalías en las covariables agronómicas y económicas, para garantizar la calidad del dataset utilizado en el modelado con procesos de Cox.

2.2 Variables Utilizadas

Las variables seleccionadas para el modelado de patrones de puntos se agrupan en categorías, derivadas de los datos de la ENA 2024. A continuación, se describen las variables clave, incluyendo su descripción y categoría.

- **Económicas**: P1001A_TOTAL (Gastos agrícolas totales), P1002B_TOTAL (Gastos pecuarios totales), P1000_TOTAL (Costo total agropecuario).
- **Espaciales**: REGION (Región natural: 1=Costa, 2=Sierra, 3=Selva), UBIGEO (Código geográfico único), CCDD (Código de Departamento), NOMBREDD (Nombre de Departamento), CCPP (Código de Provincia), NOMBREPV (Nombre de Provincia), CCDI (Código de Distrito), NOMBREDI (Nombre de Distrito), P217_SUP_ha (Superficie cosechada en hectáreas), LONGITUD (Coordenada X de segmento), LATITUD (Coordenada Y de segmento).

- **Temporales**: ANIO (Año de la encuesta: 2022–2024).
- **Agronómicas**: P212 (Fuente de agua para riego), P213 (Sistema de riego utilizado), P204_TIPO (Tipo de cultivo).
- **Producción**: P219_CANT_1 (Cantidad producida en kilogramos).

Estas variables permiten modelar la intensidad de puntos agrícolas, considerando aspectos económicos, espaciales, temporales, agronómicos y de producción.

3 Procesos de Cox Espacio-Temporales

Proposition 1 (Modelo Log-Gaussiano de Cox). *Los procesos de Cox espacio-temporales modelan patrones de puntos con intensidad $\lambda(s,t)$ que varía según la ubicación espacial $s \in R^2$ y el tiempo $t \in R$. En este estudio, se emplean Procesos log-Gaussianos de Cox (LGCP), definidos por las siguientes propiedades y técnicas, referenciadas en [5, 6]:*

1. *La intensidad se define como $\lambda(s,t) = \exp(Z(s,t))$, donde $Z(s,t)$ es un campo gaussiano espacio-temporal con media cero y covarianza $K(s,s';t,t')$.*
2. *La estimación no paramétrica de la intensidad mediante suavizado por kernel se expresa como:*
$$\hat{\lambda}(s) = \sum_{i=1}^{n} w_i k_h(\|s - s_i\|), \tag{1}$$
donde $k_h(u) = (h\sqrt{2\pi})^{-1}\exp(-u^2/(2h^2))$ es una función kernel gaussiana con ancho de banda h, y w_i son pesos basados en la producción [16].
3. *El modelo paramétrico LGCP se ajusta con:*
$$\log\lambda(s,t) = \beta_0 + \beta_1\log(1+\mathit{Gasto\_Total}) + \beta_2\log(1+\mathit{Superficie\_ha}) + Z(s,t), \tag{2}$$
donde β_i son coeficientes estimados, y $Z(s,t) \sim GP(0, K(s,s';t,t'))$ con covarianza exponencial espacio-temporal [2].
4. *La estructura espacial se analiza con la función K de Ripley:*
$$K(r) = \lambda^{-1}E[\text{No. de puntos adicionales en distancia } r], \tag{3}$$
comparada con $K_{pois}(r) = \pi r^2$ para detectar clustering o regularidad [17].

Estas técnicas, implementadas con el paquete 'spatstat' en R, permiten analizar la distribución espacio-temporal de la producción agrícola en Perú.

3.1 Model Validation and Performance Evaluation

El modelo LGCP se valida con tests de cuadrantes [17] y AIC [21], analizando residuos de Pearson [5] y comparando K de Ripley con envelopes Monte Carlo [16]. El rendimiento se evalúa por R^2 del modelo mixto [2].

3.2 Herramienta de Analisis

El análisis se realizó en R con RStudio [18], usando: `tidyverse`, `spatstat` (con `.explore`, `.geom`, `.model`, `.linnet`), `fields`, `gstat`, `sp`, `sf`, `viridis`, `patchwork`, `scales`, y `geodata` [19].

4 Resultados

El análisis de los datos de la Encuesta Nacional Agropecuaria (ENA) 2024 modeló patrones agrícolas en Perú mediante procesos log-Gaussianos de Cox (LGCP). A continuación, se resumen los principales resultados.

4.1 Características de los Datos

Se analizaron 174,040 puntos espacio-temporales (2022: 54,582; 2023: 37,992; 2024: 81,466), con mayor concentración en la Sierra (102,839), seguida de la Costa (38,152) y la Selva (33,049). La intensidad global fue 886.47 puntos/unidad^2 en un dominio de 196.33 unidades^2.

4.2 Interacción Espacial

La función K de Ripley indicó clustering significativo ($L(r) > 0$) para $r < 1.0°$ en todos los años (ver Fig. 1). El test de Monte Carlo para 2024 rechazó la aleatoriedad espacial ($p = 0.02$).

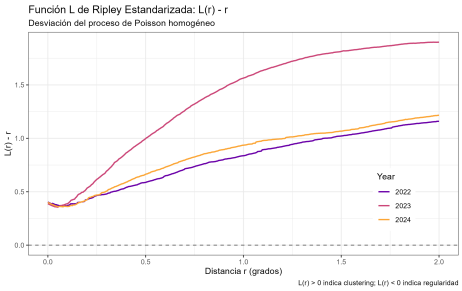

Fig. 1. Función L de Ripley estandarizada para los años 2022–2024, mostrando clustering en distancias $r < 1.0°$.

4.3 Estimación de Intensidad

La intensidad $\hat{\lambda}(s, t)$, estimada mediante suavizado por kernel ($\sigma = 0.0009$ para 2024), mostró mayores concentraciones en la Costa y la Sierra, con valores entre 0 y 438.95 puntos/unidad^2 (ver Fig. 2).

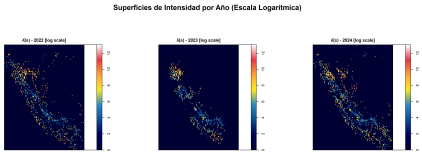

Fig. 2. Superficies de intensidad $\hat{\lambda}(s)$ por año en escala logarítmica.

4.4 Modelo LGCP

El modelo PPM para 2024 arrojó coeficientes significativos ($p < 0.001$): $\beta_0 = 6.8006$, $\beta_1 = -0.3957$ (LOG_GASTO), $\beta_2 = -0.7300$ (LOG_SUPERFICIE), con AIC = -810,737. Los residuos de Pearson y el test de cuadrantes ($\chi^2 = 97,922$, $p = 0.001$) indicaron ajuste razonable pero con desviaciones locales (ver Fig. 3).

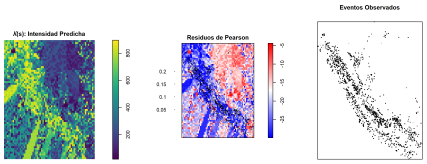

Fig. 3. Diagnósticos del modelo LGCP: intensidad predicha, residuos de Pearson y eventos observados para 2024.

4.5 Análisis Regional

La Sierra presentó la mayor intensidad (524 puntos/unidad^2), seguida de la Costa (194) y la Selva (168). Las funciones K cruzadas mostraron atracción entre regiones, especialmente Costa-Sierra y Sierra-Selva (ver Fig. 4).

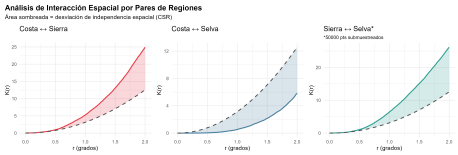

Fig. 4. Funciones K cruzadas para interacciones regionales.

4.6 Covariables Técnicas

La fuente de agua influyó en la intensidad de producción, con "Pozo" (Sierra, 16,682 kg/ha) y "Reservorio" (Costa, 15,046 kg/ha) destacando. El modelo de regresión explicó un 5.35% de la varianza ($R^2 = 0.0535$), con efectos significativos de región, gasto, superficie y fuente de agua (ver Fig. 5).

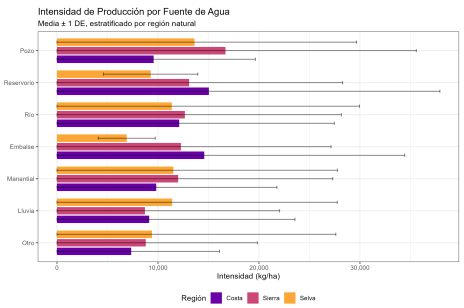

Fig. 5. Intensidad de producción por fuente de agua, estratificada por región.

Table 1. Resumen de métricas clave del análisis espacio-temporal.

Métrica	Valor
Número de eventos	174,040
Intensidad global ($\hat{\lambda}$)	886.47
AIC del modelo LGCP	-810,737
R^2 (regresión)	0.054

4.7 Distribución Espacio-Temporal

La distribución de puntos mostró concentraciones en la Costa y la Sierra, con un aumento de eventos de 2022 a 2024 (ver Fig. 6).

5 Discusión

Los resultados obtenidos confirman que los patrones agrícolas en Perú, analizados mediante procesos log-Gaussianos de Cox (LGCP), presentan una marcada heterogeneidad espacio-temporal, con clustering significativo ($L(r) > 0$ para

Table 2. Intensidad regional de actividades agrícolas.

Región	Puntos	Intensidad (λ)
Costa	38,152	194.33
Sierra	102,839	523.81
Selva	33,049	168.33

Table 3. Intensidad de producción por fuente de agua y región.

Fuente	Región	Intensidad (kg/ha)	Producción (ton)
Pozo	Sierra	16,682	5,199
Reservorio	Costa	15,046	68,440
Embalse	Costa	14,584	97,387
Río	Sierra	12,670	139,811
Lluvia	Selva	11,408	2,734,336

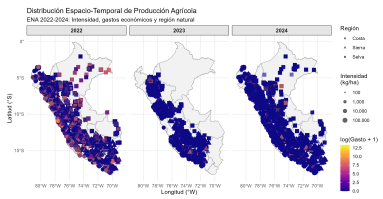

Fig. 6. Distribución espacio-temporal de la producción agrícola (2022–2024).

$r < 1.0°$) en las tres regiones estudiadas (Costa, Sierra y Selva). Este agrupamiento espacial, detectado a través de la función K de Ripley, refleja la influencia de factores locales como la disponibilidad de agua y las características del terreno, lo que coincide con estudios previos sobre procesos de puntos en contextos ambientales [6, 7]. La alta intensidad en la Sierra (524 puntos/unidad^2) frente a la Costa (194) y la Selva (168) sugiere que las condiciones agroecológicas de la Sierra favorecen una mayor densidad de actividades agrícolas, posiblemente debido a su topografía y acceso a fuentes de agua como pozos [13].

La influencia de las fuentes de agua, especialmente pozos en la Sierra (16,682 kg/ha) y reservorios en la Costa (15,046 kg/ha), resalta la importancia de la infraestructura hídrica en la productividad agrícola, en línea con estudios hidrológicos en regiones áridas de Perú [23]. Sin embargo, la baja varianza explicada por el modelo de regresión ($R^2 = 0.0535$) indica que otras covariables no incluidas, como el tipo de cultivo o las prácticas de manejo, podrían mejorar la capacidad predictiva del modelo LGCP, como se ha sugerido en aplicaciones de procesos de puntos en otros dominios [1, 21]. La interacción entre regiones, observada en las funciones K cruzadas (Fig. 4), sugiere una dependencia espacial que podría estar influenciada por patrones de migración o intercambio comercial entre Costa, Sierra y Selva, un aspecto que merece mayor exploración [14].

Aunque el modelo LGCP mostró un ajuste adecuado (AIC = -810,737), las desviaciones en los residuos de Pearson y el test de cuadrantes ($\chi^2 = 97,922, p = 0.001$) sugieren limitaciones en la captura de heterogeneidad local, especialmente en la Selva, donde el submuestreo de datos pudo introducir sesgos [3]. La implementación de métodos locales, como los propuestos por D'Angelo et al. [5], podría mejorar la precisión al permitir parámetros variables en el espacio y el tiempo. Además, la aplicación de enfoques de aprendizaje profundo, como los descritos por Choiruddin et al. [**?**], podría abordar la alta dimensionalidad de los datos agrícolas, especialmente en contextos multivariados.

Comparado con aplicaciones en sismología [6], epidemiología [9], y análisis de crímenes [22], el uso de LGCP en agricultura es menos común, pero los resultados demuestran su potencial para modelar patrones complejos. La integración de datos heterogéneos, como imágenes satelitales [14] o información de redes sociales [**?**], podría enriquecer futuros modelos, permitiendo predicciones más precisas para la planificación agrícola sostenible. La variabilidad climática, identificada como un factor clave en estudios hidrológicos [13, 23], también debería incorporarse en modelos futuros para capturar mejor los riesgos asociados a inundaciones y sequías.

6 Conclusiones

El análisis espacio-temporal de los datos de la ENA 2024 mediante procesos LGCP reveló patrones agrícolas heterogéneos en Perú, con una mayor intensidad en la Sierra y una fuerte dependencia de fuentes de agua como pozos y reservorios. El clustering espacial y las interacciones regionales destacan la importancia de considerar factores locales y dinámicas interregionales en la planificación

agrícola. El modelo LGCP, con coeficientes significativos ($\beta_0 = 6.8006$, $\beta_1 = -0.3957$, $\beta_2 = -0.7300$) y un AIC de -810,737, demostró ser una herramienta robusta, aunque con limitaciones en la varianza explicada ($R^2 = 0.0535$).

Estos hallazgos tienen implicaciones directas para la gestión de recursos hídricos y la planificación agrícola en Perú, especialmente en regiones vulnerables a la variabilidad climática. Se recomienda priorizar la inversión en infraestructura de riego en la Sierra y la Costa, y explorar la integración de datos adicionales (e.g., climáticos, satelitales) para mejorar la precisión predictiva [14, 15]. Además, la adopción de enfoques locales y de aprendizaje profundo podría superar las limitaciones actuales del modelo LGCP, facilitando un análisis más detallado de patrones agrícolas complejos.

Futuras investigaciones deberían centrarse en incorporar covariables adicionales, como el tipo de cultivo y las prácticas agrícolas, y en desarrollar modelos LGCP locales para capturar heterogeneidad a menor escala [5]. La integración de datos de alta resolución y técnicas avanzadas de aprendizaje profundo [**?**] permitirá una mejor predicción de riesgos agrícolas y una planificación más efectiva para la seguridad alimentaria en Perú.

References

1. Aglietti, V., Bonilla, E.V., Damoulas, T., Cripps, S.: Structured variational inference in continuous cox process models. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp. 1116–1126. Curran Associates Inc., Red Hook (2019)
2. Asfaw, Z.G., Brown, P.E., Stafford, J.: The root-Gaussian Cox Process for spatial-temporal disease mapping with aggregated data. Comput. Stat. 40(3), 1171–1184 (2025). https://doi.org/10.1007/s00180-024-01532-y
3. Bayisa, F.L., Ådahl, M., Rydén, P., Cronie, O.: Large-scale modelling and forecasting of ambulance calls in northern Sweden using spatio-temporal log-Gaussian Cox processes. Spat. Stat. 39, 100471 (2020). https://doi.org/10.1016/j.spasta.2020.100471
4. Clark, N.J., Watts, K.: Identification of latent structure in spatio-temporal models of violence. In: Proceedings of the Winter Simulation Conference, pp. 173–180. IEEE Press, Phoenix (2022)
5. D'Angelo, N., Adelfio, G., Mateu, J.: Locally weighted minimum contrast estimation for spatio-temporal log-Gaussian Cox processes. Comput. Stat. Data Anal. 180, 107679 (2023). https://doi.org/10.1016/j.csda.2022.107679
6. D'Angelo, N., Siino, M., D'Alessandro, A., Adelfio, G.: Local spatial log-Gaussian Cox processes for seismic data. AStA Adv. Stat. Anal. 106(4), 633–671 (2022). https://doi.org/10.1007/s10182-022-00444-w
7. D'Angelo, N., Albano, A., Gilardi, A., Adelfio, G.: Non-separable spatio-temporal Poisson point process models for fire occurrences. Environ. Ecol. Stat. 32(1), 347–381 (2025). https://doi.org/10.1007/s10651-025-00645-x
8. Gajardo, Á., Müller, H.-G.: Cox point process regression. IEEE Trans. Inf. Theory 68(2), 1133–1156 (2022). https://doi.org/10.1109/TIT.2021.3126466
9. Huang, C.-C., et al.: Spatial scale of tuberculosis transmission in Lima, Peru. Proc. Natl. Acad. Sci. 119(45), e2207022119 (2022). https://doi.org/10.1073/pnas.2207022119

10. INEI: Concurso para las investigaciones sobre los resultados de la ENA 2024. Gob.pe (2024). https://www.gob.pe/institucion/inei/campanas/113772-concurso-para-las-investigaciones-sobre-los-resultados-de-la-ena
11. INEI: Encuesta Nacional Agropecuaria (ENA) 2024. Plataforma Nacional de Datos Abiertos (2024). https://datosabiertos.gob.pe/dataset/encuesta-nacional-agropecuaria-ena-2024-instituto-nacional-de-estadistica-e-informatica-inei
12. INEI: Sistema de Microdatos. Proyectos INEI (2024). https://proyectos.inei.gob.pe/microdatos/index.html
13. Llauca, H., Leon, K., Lavado-Casimiro, W.: Construction of a daily streamflow dataset for Peru using a similarity-based regionalization approach and a hybrid hydrological modeling framework. J. Hydrol. Reg. Stud. 47, 101381 (2023). https://doi.org/10.1016/j.ejrh.2023.101381
14. Nakayama, S., et al.: Comparing spatial patterns of marine vessels between vessel-tracking data and satellite imagery. Front. Mar. Sci. 9, 1076775 (2023). https://doi.org/10.3389/fmars.2022.1076775
15. Okawa, M., Iwata, T., Kurashima, T., Tanaka, Y., Toda, H., Ueda, N.: Deep mixture point processes: Spatio-temporal event prediction with rich contextual information. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 373–383. ACM, New York (2019). https://doi.org/10.1145/3292500.3330937
16. Prokešová, M., Dvořák, J.: Statistics for inhomogeneous space-time shot-noise Cox processes. Methodol. Comput. Appl. Probab. 16(2), 433–449 (2014). https://doi.org/10.1007/s11009-013-9324-0
17. Rajala, T.A., Olhede, S.C., Grainger, J.P., Murrell, D.J.: What is the Fourier transform of a spatial point process? IEEE Trans. Inf. Theory 69(8), 5219–5252 (2023). https://doi.org/10.1109/TIT.2023.3269514
18. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2023). https://www.R-project.org/
19. Pebesma, E.: Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10(1), 439–446 (2018). Extended in: Pebesma, E., et al.: sf: Simple Features for R. R package version 1.0-7 (2022). https://CRAN.R-project.org/package=sf
20. Ribeiro, R., et al.: Incorporating environmental heterogeneity and observation effort to predict host distribution and viral spillover from a bat reservoir. Proc. R. Soc. B 290, 20231739 (2023). https://doi.org/10.1098/rspb.2023.1739
21. Spychala, C., Dombry, C., Goga, C.: Variable selection methods for log-Gaussian Cox processes: A case-study on accident data. Spat. Stat. 61, 100831 (2024). https://doi.org/10.1016/j.spasta.2024.100831
22. Escudero, I., Angulo, J.M., Mateu, J., Choiruddin, A.: Crime risk assessment through Cox and self-exciting spatio-temporal point processes. Stoch. Environ. Res. Risk Assess. 39(1), 181–203 (2025). https://doi.org/10.1007/s00477-024-02857-2
23. Wei, X., et al.: Hydrologic analysis of an intensively irrigated area in southern Peru using a crop-field scale framework. Water 13(3), 318 (2021). https://doi.org/10.3390/w13030318
24. Xu, G., et al.: Semi-parametric learning of structured temporal point processes. J. Mach. Learn. Res. 21(192), 1–39 (2020)