

1 Introducción

La papa (*Solanum tuberosum* L.) es esencial para la seguridad alimentaria peruana, con más de 4,000 variedades nativas [Devaux et al., 2021, Condori-Huanca et al., 2023]. En 2023, la producción cayó 10% debido a El Niño Costero, sequías y plagas [LatinAmerican Post, 2024, Save the Children International, 2023]. La región Puno experimentó descenso de 998,000 a 596,000 toneladas, y 64% de cultivos dependen exclusivamente de lluvia natural [ProducePay, 2024, Maldonado-Quispe et al., 2025].

Los modelos predictivos son cruciales para planificación agrícola [El-Kenawy et al., 2024, Khan et al., 2024b]. La Encuesta Nacional Agropecuaria (ENA) proporciona datos georeferenciados para modelamiento espacio-temporal. ARIMA ha sido ampliamente aplicado en agricultura [Alhussan et al., 2024, Towfek et al., 2024, Khan et al., 2024a], pero ignora dependencias espaciales. Machine learning y deep learning alcanzan errores MAPE ¡15% [Brzeziński et al., 2025, Ebrahimi et al., 2023, Farhadi et al., 2025], pero pocos estudios integran dimensiones espaciales y temporales en países en desarrollo. Los modelos espacio-temporales capturan autocorrelación espacial [Ben Hen and Rabin, 2024, Papadopoulos et al., 2021], aplicándose en forecasting agrícola [Leng and Hall, 2020, Paudel et al., 2022, Xu et al., 2024], pero su aplicación en papa peruana con tres regiones naturales diferenciadas permanece inexplorada.

Research Statement

Este estudio desarrolla y evalúa modelos STARIMA para predecir la producción de papa en el Perú durante el periodo 2017–2024, utilizando como fuente primaria las bases de datos de la Encuesta Nacional Agropecuaria (ENA). La investigación aborda la brecha metodológica en el análisis de producción agrícola al incorporar simultáneamente dependencias espaciales y temporales en un contexto de alta heterogeneidad geográfica.

Datos y variables. Se emplea la variable producción total de papa en kilogramos

(P219_EQUIV_KG), agregada a nivel departamental con frecuencia anual. Esta agregación genera una serie espacio-temporal estructurada de $N = 25$ departamentos (unidades espaciales) y $T = 7$ períodos temporales (2017–2024), resultando en 200 observaciones totales. La agregación departamental permite capturar la heterogeneidad entre las tres regiones naturales del Perú (costa, sierra y selva) mientras mantiene suficiente granularidad espacial para el análisis.

Hipótesis central. La incorporación de matrices de pesos espaciales que representan relaciones de proximidad y contigüidad entre departamentos mejorará significativamente la precisión predictiva frente a modelos puramente temporales que asumen independencia espacial.[Leng and Hall, 2020, Paudel et al., 2022].

Modelos comparativos. Se implementarán y compararán tres enfoques metodológicos mediante métricas estándar de error predictivo (RMSE: Root Mean Square Error, MAE: Mean Absolute Error, MAPE: Mean Absolute Percentage Error):

1. **ARIMA univariado por departamento:** Modelo base que trata cada departamento de forma independiente, ignorando completamente las dependencias espaciales. Este modelo servirá como benchmark para evaluar la ganancia en precisión al incorporar información espacial.
2. **STARIMA con matriz de contigüidad (W_1):** Matriz binaria donde $w_{ij} = 1$ si los departamentos i y j comparten frontera física, y $w_{ij} = 0$ en caso contrario. Esta especificación captura efectos de derrame directo entre departamentos adyacentes.
3. **STARIMA con matriz de distancia inversa (W_2):** Matriz continua donde $w_{ij} = d_{ij}^{-\alpha}$, siendo d_{ij} la distancia geodésica entre centroides departamentales y α un parámetro de decaimiento espacial a estimar. Esta especificación permite que la influencia de una región sobre otra disminuya gradualmente con la distancia, capturando patrones de difusión más complejos.

Extensión SARIMAX espacial. Adicionalmente, se evaluará la incorporación de

variables exógenas espacialmente estructuradas para construir modelos SARIMAX (Seasonal ARIMA with eXogenous variables) espaciales. Las variables exógenas incluyen:

- Sistema de riego (P119): Proporción de superficie bajo riego vs. seco
- Semilla certificada (P214): Proporción de productores usando semilla certificada
- Eventos climáticos adversos: sequía (P223B_1), heladas (P223B_3), plagas (P223B_7)
- Región natural (REGION): Variables dummy para Costa, Sierra y Selva

Estas variables permiten capturar la heterogeneidad espacial no observable en los datos de producción por sí solos, potencialmente mejorando tanto la precisión predictiva como la interpretabilidad del modelo.

Hipótesis estadística formal. Se contrastará la hipótesis nula de igualdad en desempeño predictivo:

$$H_0 : RMSE_{ARIMA} = RMSE_{STARIMA-W_1} = RMSE_{STARIMA-W_2} = RMSE_{SARIMAX_{espacial}}$$

Contribuciones esperadas. Este estudio contribuye en dos dimensiones:

Metodológica: Adapta y valida modelos STARIMA en el contexto agrícola peruano, considerando la heterogeneidad geográfica de las tres macro-regiones mediante matrices de pesos espaciales apropiadas [Ben Hen and Rabin, 2024, Papadopoulos et al., 2021].

Aplicada: Genera información accionable para el diseño de políticas agrícolas territorialmente diferenciadas, asignación eficiente de recursos públicos, y mitigación de riesgos asociados a variabilidad climática [Khan et al., 2024b, Brzeziński et al., 2025]. Dado que Perú necesita S/. 50 mil millones adicionales para adaptación climática [Global Landscapes Forum, 2024].

Los resultados ofrecerán una base empírica para sistemas de alerta temprana sobre pérdidas de rendimiento, permitiendo anticipar políticas de subsidio, estrategias de riego y medidas de control fitosanitario.

References

- Amel Ali Alhussan, Abdelhameed A Battah, SK Towfek, Abdelaziz A Ibrahim, and Abdelaziz A Abdelhamid. Global potato production forecasting based on time series analysis and advanced waterwheel plant optimization algorithm. *Potato Research*, 67:1625–1651, 2024. doi: 10.1007/s11540-024-09728-x.
- Neta Ben Hen and Neta Rabin. Spatio-temporal time series forecasting using an iterative kernel-based regression. *INFORMS Journal on Data Science*, 4(1):20–32, 2024. doi: 10.1287/ijds.2023.0019.
- Szymon Brzeziński, Agnieszka Klimek-Kopyra, Florian Gambuś, Jadwiga Wierzbowska, and Tomasz Zajac. Review of methods and models for potato yield prediction. *Agriculture*, 15(4):367, 2025. doi: 10.3390/agriculture15040367.
- Veronica Condori-Huanca, Jan Kreuze, Fernando Guzmán, Luz Gómez Oliva, David Ellis, Willmer García, Manuel Gastelo, and Merideth Bonierbale. Screening potato landraces to cope with climate change in the central Andes of Peru. *International Journal of Plant Biology*, 14(4):1167–1179, 2023. doi: 10.3390/ijpb14040085.
- André Devaux, Jean-Pierre Goffart, Peter Kromann, Jorge Andrade-Piedra, Vivian Polar, and Guy Hareau. The potato of the future: opportunities and challenges in sustainable agri-food systems. *Potato Research*, 64:681–720, 2021. doi: 10.1007/s11540-021-09501-4.
- Himan Ebrahimi, Yang Wang, and Zhanglin Zhang. Utilization of synthetic minority over-sampling technique for improving potato yield prediction using remote sensing data and machine learning algorithms with small sample size of yield data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 201:12–25, 2023.
- El-Sayed M El-Kenawy, Amel Ali Alhussan, Nima Khodadadi, Seyedali Mirjalili, and Marwa M Eid. Predicting potato crop yield with machine learning and deep learning for sustainable agriculture. *Potato Research*, 2024. doi: 10.1007/s11540-024-09753-w.

- Hanieh Farhadi, Ehsan Najafi, Aitazaz A Farooque, and Yang Wang. Advanced machine learning for regional potato yield prediction: analysis of essential drivers. *npj Sustainable Agriculture*, 3:52, 2025. doi: 10.1038/s44264-025-00052-6.
- Global Landscapes Forum. In Peru, will potatoes survive climate change? Online article, 2024. Published October 9, 2024.
- Yasir Khan, Vijay Kumar, Pankaj Setiya, and Arpan Satpathi. Forecasting soybean yield: a comparative study of neural networks, principal component analysis and penalized regression models using weather variables. *Theoretical and Applied Climatology*, 155:2937–2952, 2024a.
- Yasir Khan, Vijay Kumar, Pankaj Setiya, and Arpan Satpathi. Optimizing potato yield predictions in Uttar Pradesh, India: a comparative analysis of machine learning models. *Scientific Reports*, 14:17419, 2024b. doi: 10.1038/s41598-025-12719-8.
- LatinAmerican Post. Potato production in Peru declines due to climate vulnerability. Online article, 2024. Accessed May 31, 2024.
- Guoyong Leng and Jim W Hall. Predicting spatial and temporal variability in crop yields: an inter-comparison of machine learning, regression and process-based models. *Environmental Research Letters*, 15:044027, 2020.
- Gianella Maldonado-Quispe, Javier Cabrera-Sánchez, Augusto León-Cruz, and Nilton B Rojas-Briceño. Perception of climate change among smallholder potato producers in northern Peru. *Climate Services*, 2025. doi: 10.1080/27658511.2025.2521945.
- Antonios Papadopoulos, Anuj Karpatne, Wei-Keng Cheng, and Ilias Bilionis. Spatiotemporal forecasting in earth system science: methods, uncertainties, predictability and future directions. *Earth-Science Reviews*, 222:103828, 2021. doi: 10.1016/j.earscirev.2021.103828.

- Dilli Paudel, Hendrik Boogaard, Allard de Wit, Sander Janssen, Sjeff Ozinga, Christos Pyliandis, and Ioannis N Athanasiadis. Machine learning for regional crop yield forecasting in Europe. *Field Crops Research*, 276:108377, 2022.
- ProducePay. Climate challenges pose a threat to agricultural production in Peru. Industry report, 2024. Published April 4, 2024.
- Save the Children International. Drought by day, ice by night: Extreme weather in Peru’s Andes killing crops and leaving families hungry. Press release, 2023. Published December 15, 2023.
- SK Towfek, Amel Ali Alhussan, Reham M Saad, and El-Sayed M El-Kenawy. Potato consumption forecasting based on a hybrid stacked deep learning model. *Potato Research*, 2024. doi: 10.1007/s11540-024-09764-7.
- Xinlei Xu, Mingsheng Yang, Yizhe Wang, Lijie Song, Zhanxiang Qi, and Di Wu. Assessing climate change impacts on crop yields and exploring adaptation strategies in Northeast China. *Earth’s Future*, 12:e2023EF004063, 2024. doi: 10.1029/2023EF004063.