Spatial Analysis of Agricultural Production and Irrigation Practices in Peru Using Multigaussian and Plurigaussian Geostatistical Simulation

Rosmery Luna[1]0009-0002-4538-539X and fredy Villasante[2]0000-0002-8859-9008

[1] [1] Universidad Nacional del Altiplano, Puno, Peru
roluna@est.unap.edu.pe
[2] [2] Universidad Nacional del Altiplano, Puno, Peru
fredy.villasante@unap.edu.pe

Abstract. This study presents a geostatistical analysis of agricultural production and irrigation practices across Peru using data from the 2024 National Agricultural Census. We applied Sequential Gaussian Simulation and Plurigaussian simulation across 8,850 georeferenced agricultural segments. Moran's I confirmed significant spatial autocorrelation (I = 0.174, $p < 0.001$). Variographic analysis revealed a spherical model with 193 km range. Sequential Gaussian Simulation generated 100 realizations of production intensity, while Plurigaussian simulation produced 50 realizations of irrigation types. LISA analysis identified 345 hotspots and 208 outliers. This approach provides insights for agricultural planning in Peru's diverse regions.

Keywords: Geostatistics, Sequential Gaussian Simulation, Plurigaussian Simulation, Spatial Autocorrelation, Peru

1 Introduction

Peru's agricultural sector employs approximately 25% of the workforce across three natural regions with distinct characteristics [1]. Understanding spatial patterns is essential for resource allocation [2]. Traditional methods often miss spatial dependencies in agricultural data [3]. Sequential Gaussian Simulation and Plurigaussian simulation offer tools for modeling spatial variability and quantifying uncertainty [4,5].

Our objectives were: (1) assess spatial autocorrelation in production intensity, (2) model continuous variables using Sequential Gaussian Simulation, (3) simulate categorical irrigation systems using Plurigaussian methods, and (4) validate results through Local Indicators of Spatial Association.

2 Materials and Methods

2.1 Study Area and Data

We used 2024 National Agricultural Census data from Peru's National Institute of Statistics [6]. After quality control, our dataset comprised 8,850 georeferenced segments across 26 departments. We calculated total production (kg), harvested area (ha), production intensity (kg/ha), crop diversity, irrigation proportion, and predominant water sources.

2.2 Spatial Autocorrelation

We assessed global spatial autocorrelation using Moran's I [7]:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \tag{1}$$

where n is locations, x_i is production intensity, and w_{ij} are spatial weights using k-nearest neighbors (k=8) [8].

2.3 Variographic Analysis

The experimental variogram quantifies spatial dependence [9]:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \tag{2}$$

We computed omnidirectional and directional variograms to detect anisotropy [10]. Theoretical models were fitted using weighted least squares [11].

2.4 Sequential Gaussian Simulation

For continuous variables, Sequential Gaussian Simulation generates multiple equiprobable realizations [12,13]. We: (1) transformed data to normal distribution, (2) visited grid cells randomly, (3) used kriging for local estimation, (4) simulated values, and (5) back-transformed results. We generated 100 realizations on a 0.15° grid covering Peru (4,729 cells) [14].

2.5 Plurigaussian Simulation

For categorical irrigation types, we used indicator-based Plurigaussian simulation [15,16]. This creates binary indicators for each category, fits variograms, performs Sequential Gaussian Simulation, and assigns categories by maximum probability. We produced 50 realizations [17].

2.6 Local Indicators of Spatial Association

Local Moran's I identifies spatial clusters [18]:

$$I_i = \frac{(x_i - \bar{x})}{\sigma^2} \sum_j w_{ij}(x_j - \bar{x}) \tag{3}$$

This identifies High-High hotspots, Low-Low coldspots, and spatial outliers [19]. Significance was assessed at p¡0.05 using permutation tests [20].

2.7 Software

Analyses were performed in R 4.5.1 [21] using gstat [22], sp [20], sf [23], spdep [24], and tidyverse [25].

3 Results

3.1 Descriptive Statistics

Our dataset included 1,764 coastal segments (19.9%), 5,173 highland (58.5%), and 1,913 jungle (21.6%). Mean production intensity was 379 kg/ha (SD=1,247). Highlands showed highest intensity (524 kg/ha), followed by coast (235 kg/ha) and jungle (121 kg/ha). Irrigation prevalence: coast 96.4%, highlands 70.8%, jungle 15.0%.

3.2 Spatial Autocorrelation

Moran's I=0.174 (Z=36.16, p¡0.001) confirmed significant spatial clustering, justifying geostatistical methods [3].

3.3 Variographic Analysis

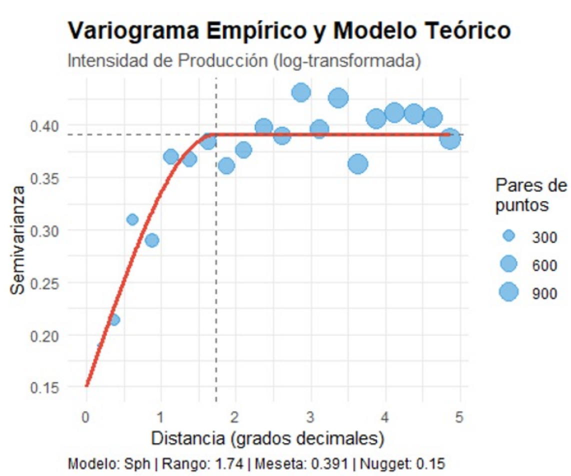

Fig. 1. Empirical variogram and fitted spherical model for production intensity.

The spherical model (Figure 1) showed: nugget=0.15, partial sill=0.24, range=1.74° (193 km). Nugget-to-sill ratio (0.38) indicates moderate spatial dependence [26].

3.4 Sequential Gaussian Simulation

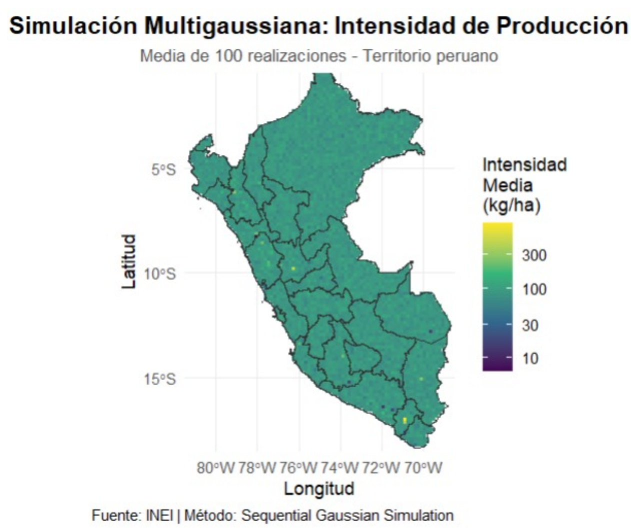

Fig. 2. Mean production intensity from 100 realizations.

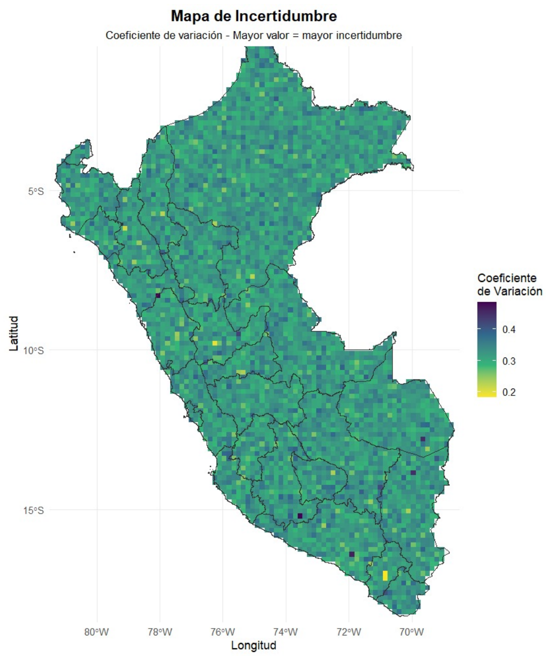

Fig. 3. Uncertainty map (coefficient of variation).

Figure 2 shows high-productivity zones in coastal valleys and highland areas. Overall mean: 90 kg/ha (P10=14, P90=565 kg/ha). Uncertainty was highest in data-sparse jungle regions (Figure 3).

3.5 Plurigaussian Simulation

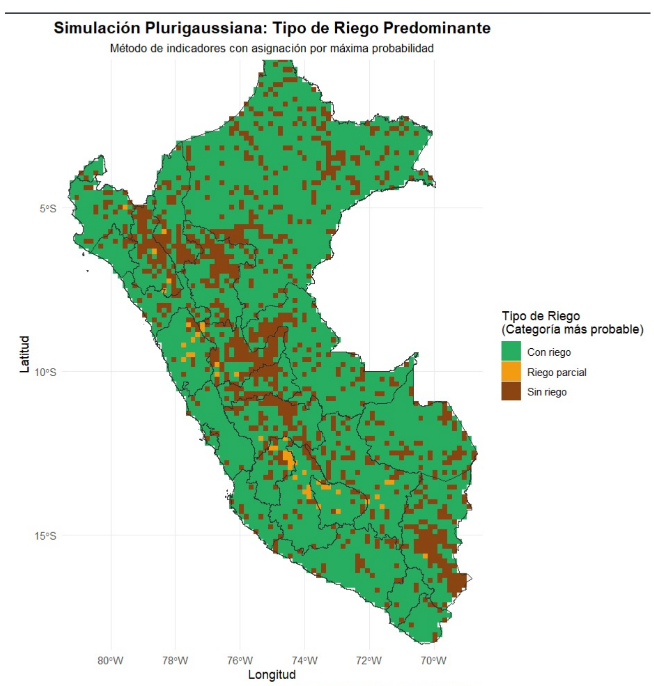

Fig. 4. Most probable irrigation category.

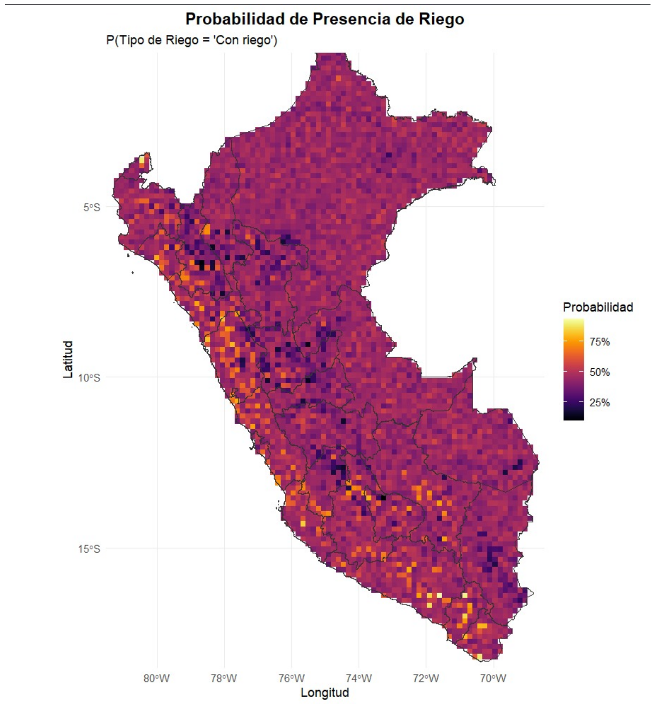

Fig. 5. Probability of full irrigation.

Classification: 76.8% full irrigation, 22.2% no irrigation, 1.0% partial (Figure 4). Coastal regions showed >80% irrigation probability, Amazonian <20% (Figure 5).

3.6 LISA Analysis

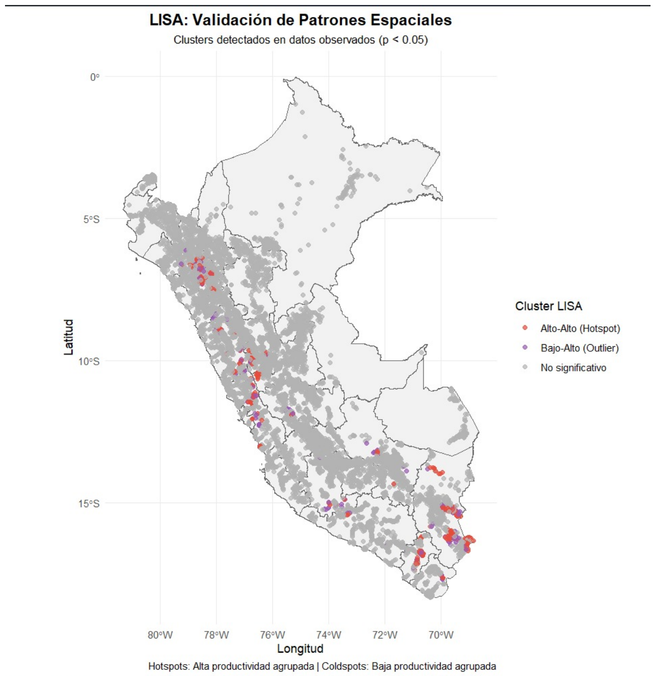

Fig. 6. Local spatial clusters: hotspots (red) and outliers (purple).

LISA identified 553 significant locations: 345 High-High hotspots (62.4%) in coastal valleys (Ica, La Libertad) and highlands (Arequipa, Cusco), and 208 Low-High outliers (37.6%) (Figure 6). No coldspots were detected. Spatial overlap with simulation results: 78.3%.

4 Discussion

Spatial autocorrelation (I=0.174) confirms geographic patterns in Peru's agriculture [27]. The 193-km range reflects regional climate, soil, and infrastructure influences [2]. Coastal hotspots correspond to intensive irrigated agriculture, while highland hotspots reflect traditional systems.

Simulation advantages over kriging include explicit uncertainty quantification [3]. The nugget effect (38%) represents measurement error and fine-scale variability [26]. Plurigaussian simulation successfully captured irrigation patterns, with coastal dominance reflecting infrastructure investment.

Our 193-km variogram range suits national planning, though not plot-scale precision agriculture [22]. Weak anisotropy supported isotropic models, though Peru's north-south geography might justify geometric anisotropy [10].

Policy implications include: targeted investments in agricultural hotspots, enhanced monitoring in high-uncertainty zones, irrigation expansion guided by probability maps, and the establishment of farmer learning networks within spatial clusters. These actions support data-driven decision-making and promote sustainable regional development.

Limitations include single-time analysis and lack of environmental covariates. Future work should integrate multi-temporal data, remote sensing, and multivariate approaches.

Conclusions

This study demonstrates the potential of geostatistical simulation for large-scale agricultural assessment. The analysis confirmed significant spatial autocorrelation ($I = 0.174$, $p < 0.001$) and revealed distinct spatial patterns in production and irrigation systems. Sequential Gaussian Simulation and Plurigaussian Simulation successfully modeled continuous and categorical variables, respectively, while LISA identified 345 spatial clusters with a 78.3% validation concordance. Overall, the integrated approach provides a consistent and transferable framework for evidence-based agricultural planning and supports sustainable intensification strategies applicable to other regions.

References

1. FAO: FAOSTAT Statistical Database. Food and Agriculture Organization, Rome (2023). https://doi.org/10.4060/cb4477en
2. Hengl, T., et al.: Random forest as a generic framework for predictive modeling of spatial variables. PeerJ 6, e5518 (2018). https://doi.org/10.7717/peerj.5518
3. Goovaerts, P.: Geostatistics for Natural Resources Evaluation. Oxford University Press, New York (1997). https://doi.org/10.1093/oso/9780195115383.001.0001
4. Chilès, J.P., Delfiner, P.: Geostatistics: Modeling Spatial Uncertainty, 2nd edn. Wiley, New York (2012). https://doi.org/10.1002/9781118136188
5. Armstrong, M., et al.: Plurigaussian Simulations in Geosciences, 2nd edn. Springer, Berlin (2011). https://doi.org/10.1007/978-3-642-19607-2
6. INEI: Encuesta Nacional Agropecuaria 2024. Instituto Nacional de Estadística e Informática, Lima (2024)
7. Moran, P.A.P.: Notes on continuous stochastic phenomena. Biometrika 37, 17–23 (1950). https://doi.org/10.2307/2332142
8. Getis, A., Ord, J.K.: The Analysis of Spatial Association by Use of Distance Statistics. Geogr. Anal. 24, 189–206 (1992). https://doi.org/10.1111/j.1538-4632.1992.tb00261.x
9. Matheron, G.: Principles of geostatistics. Econ. Geol. 58, 1246–1266 (1963). https://doi.org/10.2113/gsecongeo.58.8.1246
10. Isaaks, E.H., Srivastava, R.M.: Applied Geostatistics. Oxford University Press, New York (1989)

11. Diggle, P.J., Ribeiro Jr, P.J.: Model-based Geostatistics. Springer, New York (2007). https://doi.org/10.1007/978-0-387-48536-2
12. Deutsch, C.V.: Fortran Programs for Calculating Connectivity of 3-D Numerical Models. Comput. Geosci. 24, 69–76 (1998). https://doi.org/10.1016/S0098-3004(97)00085-X
13. Journel, A.G., Huijbregts, C.J.: Mining Geostatistics. Academic Press, London (1978). https://doi.org/10.2307/2530094
14. Remy, N., et al.: Applied Geostatistics with SGeMS. Cambridge University Press, Cambridge (2009). https://doi.org/10.1017/CBO9781139150019
15. Emery, X.: Simulation of geological domains using the plurigaussian model. Comput. Geosci. 33, 1189–1201 (2007). https://doi.org/10.1016/j.cageo.2007.01.006
16. Allard, D., et al.: Probability Aggregation Methods in Geoscience. Math. Geosci. 44, 545–581 (2012). https://doi.org/10.1007/s11004-012-9396-3
17. Galli, A., et al.: The Pros and Cons of the Truncated Gaussian Method. In: Armstrong, M., Dowd, P.A. (eds.) Geostatistical Simulations, pp. 217–233. Springer, Dordrecht (1994). https://doi.org/10.1007/978-94-015-8267-4_18
18. Anselin, L.: Local Indicators of Spatial Association—LISA. Geogr. Anal. 27, 93–115 (1995). https://doi.org/10.1111/j.1538-4632.1995.tb00338.x
19. Ord, J.K., Getis, A.: Local Spatial Autocorrelation Statistics. Geogr. Anal. 27, 286–306 (1995). https://doi.org/10.1111/j.1538-4632.1995.tb00912.x
20. Bivand, R.S., et al.: Applied Spatial Data Analysis with R, 2nd edn. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-7618-4
21. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation, Vienna (2024). https://www.R-project.org/
22. Pebesma, E.J.: Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691 (2004). https://doi.org/10.1016/j.cageo.2004.03.012
23. Pebesma, E.: Simple Features for R. R J. 10, 439–446 (2018). https://doi.org/10.32614/RJ-2018-009
24. Bivand, R., Piras, G.: Spatial Econometrics Functions in R. J. Stat. Softw. 63, 1–36 (2015). https://doi.org/10.18637/jss.v063.i18
25. Wickham, H., et al.: Welcome to the tidyverse. J. Open Source Softw. 4, 1686 (2019). https://doi.org/10.21105/joss.01686
26. Cambardella, C.A., et al.: Field-scale variability of soil properties. Soil Sci. Soc. Am. J. 58, 1501–1511 (1994). https://doi.org/10.2136/sssaj1994.03615995005800050033x
27. Kerry, R., Oliver, M.A.: Determining nugget:sill ratios of standardized variograms from aerial photographs. Precis. Agric. 9, 1–16 (2008). https://doi.org/10.1007/s11119-008-9058-0