Bayesian Spatial Analysis of Agricultural Yields and Climate Vulnerability in Southern Peru using Hierarchical CAR Models, 2024

Aldair Jose Maquera Andrade[1]

Faculty of Statistical and Computer Engineering,
Universidad Nacional del Altiplano de Puno (National University of the Altiplano of Puno), Peru
aj.maquera@est.unap.edu.pe

Abstract. This study quantifies spatial autocorrelation in agricultural yields in southern Peru and identifies priority provinces for climate adaptation using hierarchical Bayesian CAR models. We analyzed 1,368 observations of white potatoes and starchy corn in 20 provinces of 5 departments during 2024 (National Agricultural Survey). Results show significant spatial autocorrelation (I=0.404, p<0.001; ρ=0.402, 95% CI:[0.015,0.915]). Droughts reduced yields by 183 kg/ha (16%, CI:[-355,-3]), the only climatic event with a robust effect. Potatoes outperformed corn by 780 kg/ha (104%). Four hotspots were identified within Puno department (1,605-2,345 kg/ha) and four coldspots in Cusco (538-1,229 kg/ha). Agricultural practices showed no effects (low adoption: terraces 1.6%, certified seed 0.4%). The model explained 18.4% of the variance. Four critical provinces in Cusco plus La Unión (Arequipa, 461 kg/ha) require urgent intervention. First estimate of spatial autocorrelation parameter (ρ) in Peruvian agriculture, establishing a methodological framework for evidence-based territorial prioritization.

Keywords: spatial autocorrelation, Bayesian CAR models, climate vulnerability, Andean agriculture, Getis-Ord Gi*

1 Introduction

Southern Peru accounts for 40% of national agricultural production and supports 2.3 million rural families [15], facing increasing vulnerability to extreme weather events that reduce yields by 20-60% [14, 6] and annual losses of USD 3.4 billion [22]. Despite the adoption of adaptive practices, there is little quantitative evidence on their effectiveness against specific events or spatial variation in impacts [17, 24]. Traditional models ignore spatial dependence between geographic units, assuming independence and underestimating uncertainty [21, 10].

This research applies Leroux's Bayesian CAR models for the first time in Peruvian agriculture to: (i) estimate the magnitude of the impact of climatic events by controlling hierarchical effects and spatial autocorrelation; (ii) evaluate the effectiveness of agricultural practices; (iii) identify statistically significant clusters; (iv) generate a territorial prioritization system. **Hypotheses:** (H1) yields exhibit positive spatial autocorrelation with geographic clusters; (H2) droughts have a robust adverse effect; (H3) there are persistent clusters of low yields in Cusco and high yields in Puno; (H4) potatoes show higher yields than corn.

2 Methodology

Design and data. Cross-sectional study with a spatial hierarchical structure. Data: 2024 National Agricultural Survey [11] (production, area, climatic events, agricultural practices). Sample: 1,368 observations, 20 provinces, 5 departments (Puno, Cusco, Arequipa, Tacna, Moquegua). Crops: white potatoes and starchy corn.

Variables. Dependent: Yield (kg/ha). Climatic (binary): droughts, frosts, hailstorms. Practices (binary): organic matter, terraces, rotation, water management, certified seed. Control: crop type. Spatial: geographic coordinates.

Spatial autocorrelation. Moran's I index [16]: $I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}$. Weight matrix: k-neighbors (k=5). Getis-Ord Gi* [9] for hotspots/coldspots at 90%, 95%, 99%.

Bayesian hierarchical model (brms/Stan [3, 4]):

$$\text{Yield}_{ijt} \sim \text{Normal}(\mu_{ijt}, \sigma^2), \quad \mu_{ijt} = X_{ijt}\beta + u_j + v_{jk} \tag{1}$$

where X includes climatic events, crop, practices, frost×crop interaction; $u_j \sim \text{N}(0, \sigma_u^2)$ (department); $v_{jk} \sim \text{N}(0, \sigma_v^2)$ (province).

Spatial CAR model (CARBayes [12]):

$$\text{Yield}_k \sim \text{Normal}(X_k\beta + \phi_k, \nu^2), \quad \phi_k | \phi_{-k} \sim \text{Normal}\left(\frac{\rho \sum_j w_{kj}\phi_j}{\rho \sum_j w_{kj} + 1 - \rho}, \frac{\tau^2}{\rho \sum_j w_{kj} + 1 - \rho}\right) \tag{2}$$

$\rho \in [0, 1]$ control autocorrelation: $\rho = 0$ independence, $\rho = 1$ complete dependence.

Inference. Hierarchical: 4 MCMC chains, 2,000 iterations (1,000 warm-up). CAR: 1 chain, 20,000 iterations (10,000 warm-up), thinning=10. Weakly informative priors: $\beta \sim \text{N}(0, 10)$; $\sigma_u, \sigma_v, \sigma \sim \text{Student-t}^+(3, 0, 2.5)$; $\rho \sim \text{U}(0, 1)$; $\tau^2 \sim \text{Gamma}^{-1}(1, 0.01)$. Convergence: $\hat{R} < 1.1$, ESS> 400 [8]. Goodness of fit: LOO-CV [20], DIC, R^2 Bayesian. Software: R 4.4.3, brms, CARBayes, spdep [18, 1].

3 Results

3.1 Descriptive Statistics

The final sample included 1,368 observations distributed across 20 provinces in 5 departments during 2024. The average yield was 1,147 kg/ha (SD=1,234 kg/ha), with marked heterogeneity between provinces (range: 461-2,345 kg/ha, coefficient of variation: 107.6%). White potatoes had an average yield of 1,526 kg/ha, surpassing starchy corn (747 kg/ha), a difference of 104.3%.

Table 1. Descriptive statistics by department

Dept.	N	Yield (kg/ha)	Frosts (%)	Droughts (%)	CV
Puno	151	1897 (987)	62.9	14.6	52.0
Cusco	811	789 (1043)	55.1	28.9	132.2
Arequipa	132	612 (845)	31.8	0.8	138.1
Moquegua	195	1296 (1178)	60.0	7.7	90.9
Tacna	79	1364 (1256)	34.1	17.1	92.1
Total	**1368**	**1147 (1234)**	**54.1**	**23.8**	**107.6**

Note: DE in parentheses. CV = Coefficient of variation (%).

Puno exhibited the highest average yield (1,897 kg/ha) with the lowest variability (CV=52.0%), while Cusco showed the lowest (789 kg/ha) with high variability (CV=132.2%). The prevalence of climatic events varied substantially: frost affected 54.1% of producers (departmental range: 31.8-62.9%), droughts 23.8% (range: 0.8-28.9%), and hailstorms 18.3%. Cusco had the highest exposure to droughts (28.9%), while Arequipa had the lowest (0.8%).

3.2 Spatial Autocorrelation Analysis

Global Autocorrelation Moran's I test revealed significant positive spatial autocorrelation ($I=0.404$, $z=4.13$, $p<0.001$), rejecting the null hypothesis of random spatial distribution ($p<0.001$). This value indicates moderate-to-high spatial clustering: provinces with similar yields tend to be geographically grouped. Geary's C index complemented these findings ($C=0.594$, $z=3.36$, $p<0.001$), confirming positive clustering. Values of $C<1$ indicate that neighboring provinces have more similar yields than expected under spatial randomness, validating the need for models that incorporate explicit spatial structure.

3.3 Local Analysis: Hotspots and Coldspots

The Getis-Ord Gi* statistic identified statistically significant spatial clusters at multiple confidence levels. Four hotspots (high yield) were detected in Puno at the 99% confidence level: Yunguyo (1,605 kg/ha, Gi*=2.87), Puno capital (1,676 kg/ha, Gi*=2.84), Huancané (1,961 kg/ha, Gi*=2.76), and El Collao (2,345 kg/ha, Gi*=2.70). These clusters represent areas of agricultural excellence with yields 40-104% above the national average.

Conversely, four coldspots (low yields) were identified: Canas in Cusco (1,229 kg/ha, Gi*=-2.15, $p<0.05$), Acomayo in Cusco (666 kg/ha, Gi*=-1.82, $p<0.10$), Canchis in Cusco (538 kg/ha, Gi*=-1.75, $p<0.10$), and La Unión in Arequipa

(461 kg/ha, Gi*=-1.66, $p<0.10$). Notably, four of the five cold spots were concentrated in Cusco, evidencing marked spatial polarization of agricultural yield not captured by aggregate departmental analyses.

3.4 Hierarchical Bayesian Modeling

Convergence and Diagnostics The hierarchical model exhibited satisfactory convergence for all parameters: $\hat{R} < 1.01$ (all variables), effective sample sizes ESS $>$ 1,200 (fixed effects) and ESS $>$ 800 (random effects). The MCMC chains showed adequate mixing without systematic trends. The posterior predictive check indicated that the model adequately captures the general shape of the yield distribution, although with a slight overestimation of dispersion at high values.

Fixed Effects **Weather events:** Droughts exhibited the only statistically robust adverse effect, reducing yields by 183 kg/ha (95% CI: [-355, -3]), equivalent to 16.0% of the average yield. This result is consistent with global literature reporting reductions of 15-30% [14]. Frosts showed a non-significant positive effect (20 kg/ha, CI: [-165, 206]), suggesting local adaptation through tolerant varieties. Hail did not exhibit significant effects (-43 kg/ha, CI: [-198, 112]).

Differences between crops: White potatoes outperformed starchy corn by 780 kg/ha (95% CI: [594, 965]), the most robust effect in the model with posterior probability $P(\beta_{potato} > 0) = 1.00$. This result represents a difference of 104.4% and confirms the adaptive advantage of potatoes in high Andean conditions, consistent with the literature on the resilience of Andean tubers [7].

Agricultural practices: No practice showed significant effects at 95% credibility. Terracing (+113 kg/ha, CI: [-365, 587], $P(\beta > 0)=0.68$), organic matter (+15 kg/ha, CI: [-129, 163], $P(\beta > 0)=0.54$), crop rotation (-79 kg/ha, CI: [-227, 66], $P(\beta < 0)=0.86$), water management (-217 kg/ha, CI: [-582, 141], $P(\beta < 0)=0.89$), and certified seed (-352 kg/ha, CI: [-1067, 391], $P(\beta < 0)=0.82$) exhibited wide intervals crossing zero, reflecting high uncertainty.

Hierarchical Structure The variability between departments (SD=249 kg/ha, 95% CI: [26, 501]) and between provinces within departments (SD=222 kg/ha, 95% CI: [42, 421]) was substantial, justifying a multilevel structure. Departmental variance accounts for 4.2% of total variance, while provincial variance accounts for 3.4% and residual variance accounts for 92.4%. The model explained 18.4% of total variance (Bayesian $R^2 = 0.184$, 95% CI: [0.153, 0.217]), indicating that unobserved factors (soil quality, microclimate, management capacity, credit access) explain 81.6% of the variability.

3.5 Spatial CAR Modeling

Autocorrelation Parameter Leroux's CAR model estimated a spatial autocorrelation parameter $\rho=0.402$ (95% CI: [0.015, 0.915]), indicating moderate-high spatial dependence. The lower limit of the interval barely excludes zero

Table 2. A posteriori estimates: hierarchical model versus spatial CAR

Variable	Hierarchical			Spatial CAR		
	Mean	**95% CI**	$\hat{R}$	**Mean**	**95% CI**	**ESS**
Intercept	758	[442, 1074]	1.00	-29	[-732, 640]	2000
Droughts	-183	[-355, -3]	1.00	619	[-231, 1427]	2000
Frosts	20	[-165, 206]	1.00	190	[-649, 1006]	2000
Hailstorms	-43	[-198, 112]	1.00	485	[-404, 1331]	2071
White potato	780	[594, 965]	1.00	1455	[752, 2125]	2201
Organic matter	15	[-129, 163]	1.00	192	[-570, 943]	2000
Terraces	113	[-365, 587]	1.00	10102	[2829, 17181]	1791
Rotation	-79	[-227, 66]	1.00	-594	[-1345, 188]	2000
Water management	-217	[-582, 141]	1.00	939	[-4048, 6008]	2000
Certified seeds	-352	[-1067, 391]	1.00	-24861	[-51568, 4951]	2000
Frost × Potato	62	[-192, 316]	1.00	–	–	–
SD (Department)	249	[26, 501]	1.01	–	–	–
SD(Prov.)	222	[42, 421]	1.01	–	–	–
SD(Residual)	1198	[1154, 1246]	1.00	–	–	–
ρ (autocorr.)	–	–	–	0.402	[0.015, 0.915]	1406
τ^2 (spec. var.)	–	–	–	0.018	[0.002, 0.091]	672
ν^2 (res. var.)	–	–	–	104328	[44822, 243733]	1752
Bayesian R^2 = 0.184; LOOIC = 23324.2				DIC = 297.5; LMPL = -153.1		

(0.015), suggesting moderate but not overwhelming evidence of spatial structure. The posterior probability P(ρ > 0)=0.976 indicates 97.6% certainty of positive autocorrelation. This result validates hypothesis H1 and confirms that ignoring spatial structure underestimates the uncertainty of estimates.

Changes in Estimates By incorporating explicit spatial autocorrelation, the point estimates changed substantially compared to the hierarchical model: droughts reversed sign (+619 kg/ha vs. -183 kg/ha), frosts increased in magnitude (+190 kg/ha vs. +20 kg/ha), and white potatoes increased in effect (+1,455 kg/ha vs. +780 kg/ha). However, all these climatic effects had wide intervals crossing zero in the CAR model, reflecting the difficulty in precise causal identification with n=20 spatial units and high collinearity between spatially aggregated climatic events.

Terraces showed an extremely high effect (10,102 kg/ha, CI: [2,829, 17,181]) but implausible, possibly reflecting confusion with altitude, soil type, or economic capacity of producers. Certified seed had an implausible negative effect (-24,861 kg/ha), indicating severe identification problems due to extremely low prevalence (0.4% of sample). These results suggest that the CAR model with n=20 faces limitations in estimating the effects of variables with low spatial variability.

3.6 Goodness of Fit

The CAR model presented DIC=297.5 with an effective number of parameters p.d=9.3, and marginal predictive likelihood log LMPL=-153.1. Direct comparison with hierarchical models using information criteria is not possible (different implementations: Stan vs. classical MCMC), but the parameter ρ significantly different from zero suggests that the spatial model captures structure not modeled by simple hierarchical random effects. Moran's I test applied to CAR model residuals did not detect significant residual autocorrelation (I=-0.082, p=0.68), validating the adequacy of the spatial model.

3.7 Territorial Prioritization System

Integrating observed yield, estimated departmental effects, climate vulnerability (composite index: frost + 2×droughts + hailstorms), and spatial classification (hotspot/coldspot), a prioritization system for public policies was generated.

Table 3. Priority provinces for territorial intervention

Province	**Dept.**	**Yield (kg/ha)**	**Effect dept.**	**Vuln. (0-3)**	**Gi***	**Prior.**
Canchis	Cusco	538	-358	1.05	Coldspot	Critical
Acomayo	Cusco	666	-358	1.05	Coldspot	Critical
Cusco	Cusco	2077	-358	1.38	Coldspot	Critical
Canas	Cusco	1229	-358	1.96	Coldspot	Critical
La Unión	Arequipa	461	-535	0.06	Coldspot	High
Urubamba	Cusco	848	-358	1.58	Normal	High
El Collao	Puno	2345	+750	2.07	Hotspot	High
Chumbivilcas	Cusco	634	-358	0.36	Normal	Medium
Caylloma	Arequipa	729	-535	0.51	Normal	Medium
Tacna	Tacna	1364	+217	0.51	Hotspot	Low

Dept. effect = Random departmental effect of the hierarchical model (kg/ha).
Vuln. = Composite climate vulnerability index.

Critically prioritized provinces: The four provinces of Cusco (Canchis, Acomayo, Cusco, Canas) require urgent intervention based on three criteria: (i) low absolute yield (538-1,229 kg/ha, 53-107% below the national average), (ii) high climate vulnerability (0.36-1.96), (iii) belonging to statistically significant spatial coldspots. Canchis presents the most critical situation: yield 53% below the national average, negative departmental effect of -358 kg/ha, and classification as a 90% coldspot.

High-priority provinces: La Unión (Arequipa) requires immediate attention despite low climate vulnerability (0.06) due to extremely low yield (461 kg/ha, 60% below average), the most negative departmental effect of the entire sample (-535 kg/ha), and classification as a coldspot. El Collao (Puno), although a hotspot with exceptional yield (2,345 kg/ha, 104% above average), shows very high climate vulnerability (2.07) that threatens the sustainability of current productivity, justifying preventive investments in adaptation.

Spatial implications: The concentration of 4 out of 5 coldspots in a single department (Cusco) suggests systemic structural factors beyond climate: advanced soil degradation, limited access to markets due to poor road infrastructure, low agricultural extension coverage, or credit restrictions. This justifies coordinated interventions at the departmental level to complement provincial actions. Potential spatial spillovers (I=0.404) imply that improvements in critical provinces can spread to neighboring ones through social networks and peer learning [5].

4 Discussion

Significant spatial autocorrelation (I=0.404, ρ=0.402) is consistent with CAR applications in agriculture [23], although lower than previous studies in Europe that reported (0.35-0.55) [21]. Hotspots/coldspots confirm Andean subnational heterogeneity [19]. Sample size (n=20) introduces uncertainty [10], requiring higher district resolution.

Droughts reduce yields by 16%, in line with the global 15-30% [14]. The absence of frost effects suggests local adaptation [24], although binary variables do not capture intensity [6]. Potato +104% vs corn reflects high Andean advantages [7, 13]. Low adoption of practices (terraces 1.6%, certified seed 0.4%) reflects structural limitations rather than ineffectiveness [17].

Limitations: (i) cross-sectional design prevents robust causality; (ii) n=20 affects precision; (iii) binary variables underestimate intensity; (iv) R^2=0.184 indicates dominant unobserved factors [2]; (v) k=5 simplifies spatial interactions [21].

5 Conclusions

First quantitative evidence of spatial autocorrelation (I=0.404, ρ=0.402) in agriculture in southern Peru using Bayesian CAR models. Droughts reduce yields by 16%—the only robust event. Potatoes are 104% more productive than corn. Five critical provinces (4 in Cusco, 1 in Arequipa) require urgent intervention in drought management. Low adoption of practices (1.6-0.4%) reflects structural barriers.

Recommendations: (i) pilot drought management programs in critical provinces; (ii) Puno-Cusco exchange networks; (iii) spatially differentiated subsidies; (iv) provincial early warning systems; (v) infrastructure investment in Cusco.

Future work: spatiotemporal models with 2019-2024 panels, satellite data (NDVI, CHIRPS) at the district level (n≈100), randomized experimental studies, incorporation of soil/topography/market variables.

Agricultural vulnerability has a pronounced spatial structure, requiring territorially differentiated policies based on quantitative evidence. Bayesian CAR models provide rigorous tools for identifying priority areas and efficiently allocating public resources in climate change.

Bibliography

[1] Bivand, R., Piras, G.: Comparing implementations of estimation methods for spatial econometrics. Journal of Statistical Software 63(18), 1–36 (2015)

[2] Burke, M., et al.: Global non-linear effect of temperature on economic production. Nature 527, 235–239 (2015)

[3] Bürkner, P.C.: brms: An r package for bayesian multilevel models using stan. Journal of Statistical Software 80(1), 1–28 (2017)

[4] Carpenter, B., et al.: Stan: A probabilistic programming language. Journal of Statistical Software 76(1), 1–32 (2017)

[5] Conley, T.G., Udry, C.R.: Learning about a new technology. American Economic Review 100, 35–69 (2010)

[6] Dell, M., Jones, B.F., Olken, B.A.: What do we learn from the weather? the new climate-economy literature. Journal of Economic Literature 52(3), 740–798 (2014)

[7] Devaux, A., et al.: Global food security from sustainable potato systems. The Potato Crop pp. 3–35 (2021)

[8] Gelman, A., Rubin, D.B.: Inference from iterative simulation using multiple sequences. Statistical Science 7(4), 457–472 (1992)

[9] Getis, A., Ord, J.K.: The analysis of spatial association by use of distance statistics. Geographical Analysis 24(3), 189–206 (1992)

[10] Hodges, J.S., Reich, B.J.: Adding spatially-correlated errors can mess up the fixed effect you love. The American Statistician 64, 325–334 (2010)

[11] Instituto Nacional de Estadística e Informática: Encuesta Nacional Agropecuaria 2024. INEI, Lima, Perú (2024)

[12] Lee, D.: Carbayes: An r package for bayesian spatial modeling. Journal of Statistical Software 55(13), 1–24 (2013)

[13] Lipper, L., et al.: Climate-smart agriculture for food security. Nature Climate Change 4, 1068–1072 (2014)

[14] Lobell, D.B., et al.: Nonlinear heat effects on african maize as evidenced by historical yield trials. Nature Climate Change 1, 42–45 (2011)

[15] Ministerio de Agricultura y Riego del Perú: Anuario estadístico de producción agrícola 2023. Tech. rep., MINAGRI (2023)

[16] Moran, P.A.P.: Notes on continuous stochastic phenomena. Biometrika 37, 17–23 (1950)

[17] Posthumus, H., et al.: Cost-benefit analysis of terraces in peru. Land Degradation & Development 21, 427–439 (2010)

[18] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2023), https://www.R-project.org/

[19] Sietz, D., et al.: Landuse-based adaptation to global change. Environmental Management 50, 87–99 (2012)

[20] Vehtari, A., Gelman, A., Gabry, J.: Practical bayesian model evaluation using leave-one-out cross-validation. Statistics and Computing 27(5), 1413–1432 (2017)

[21] Wall, M.M.: A close look at the spatial structure implied by the car and sar models. Journal of Statistical Planning and Inference 121, 311–324 (2004)
[22] World Bank Group: Climate-smart agriculture in peru. Tech. rep., World Bank Group (2019), https://climateknowledgeportal.worldbank.org/sites/default/files/2019-06/CSA-in-Peru.pdf
[23] Zhang, B., et al.: Spatial analysis of climate change impacts on yield. Journal of Geographical Sciences 28, 1547–1567 (2018)
[24] Zimmerer, K.S.: Biological diversity in agriculture and global change. Annual Review of Environment and Resources 35, 137–166 (2010)