Artificial Intelligence for Sustainable Agriculture: SARIMAX Models in Optimization of High-Andean Potato Production

Lindell Dennis Vilca Mamani[1] 0009-0000-0583-3237

Universidad Nacional del Altiplano
Faculty of Statistical and Computer Engineering
Puno, Peru
livilca@est.unap.edu.pe

Abstract. High-Andean agriculture faces challenges of low productivity and limited technological adoption in climate change contexts. This study applied SARIMAX models to evaluate the impact of sustainability indicators on the predictive capacity of potato production in Puno, Peru (2015-2024, n=17,679 producers). A technification index based on yield was constructed as a proxy for adoption of sustainable practices, classifying producers into three technological levels. Results show that 18.9% operate with high technology, with perfect correlation (r=1.0, p¡0.05) between technological level and yield. The SARIMAX model incorporating sustainability variables presented an annual seasonal component ARIMA(0,0,1)$(0,0,1)_{12}$, improving interpretability compared to the economic baseline model. Scenario simulation projects a 3.4% increase in production with greater technological adoption (+30%), demonstrating resource optimization potential. These findings show that artificial intelligence applied through SARIMAX models facilitates decision-making for precision agriculture in high-Andean zones, contributing to food security and green economy through efficient use of land and productive resources.

Keywords: SARIMAX, sustainable agriculture, artificial intelligence, resource optimization, green economy, precision agriculture

1 Introduction

Agriculture in high-Andean zones constitutes a fundamental productive system for the food security of millions of people in the Andean region of South America. The potato (*Solanum tuberosum*), a crop native to the Peruvian Andes, represents not only a central component of the local diet but also a critical source of income for small producers at altitudes ranging between 3,800 and 4,500 meters above sea level [10]. However, this agriculture faces growing challenges derived from climate change, including rainfall variability, temperature increases, and greater frequency of extreme weather events such as frosts and droughts [5, 6].

The Puno region, located in the Peruvian altiplano, concentrates significant potato production with more than 17,000 registered producers annually. Nevertheless, the adoption of improved agricultural technologies remains limited. Previous studies have documented that Andean farmers have had to displace their crops up to 150 meters in altitude over the last three decades as a climate adaptation strategy, evidencing the urgent need for predictive tools that facilitate agricultural planning and resource optimization in this high vulnerability context.

Despite growing interest in artificial intelligence applications for agriculture [12, 15, 11], there is a significant research gap on time series models that explicitly incorporate sustainability variables for high-Andean regions. SARIMAX models (Seasonal Autoregressive Integrated Moving Average with eXogenous variables) have demonstrated effectiveness in agricultural forecasting in similar contexts [13, 9, 8], but their application with green technology indicators in small-scale Andean agriculture remains unexplored. This limitation is particularly relevant given that agricultural databases in developing countries frequently lack direct information on technological adoption.

This work presents a methodological contribution by developing a technification index based on yield as a proxy for adoption of sustainable practices, validating its application in SARIMAX models for production forecasting. Specifically, the objectives are: (1) to characterize the evolution of technological levels in potato producers of Puno during 2015-2024, (2) to compare the predictive capacity of SARIMAX models with and without sustainability variables, and (3) to simulate resource optimization scenarios through increases in technological adoption. The results contribute to the body of knowledge on AI application for precision agriculture in resource-limited contexts, with direct implications for agricultural extension policies and green economy in high-Andean zones.

2 Related Work

2.1 SARIMAX Models in Agriculture

SARIMAX models have gained prominence in recent years for agricultural production forecasting due to their capacity to incorporate climatic and economic exogenous variables. Pandit et al. developed hybrid ARIMAX-LSTM models for forecasting Rabi crop yields in India (wheat, sugarcane, peanuts), using irrigated area as an exogenous variable and achieving significant improvements in predictive accuracy compared to univariate models. In contexts of high climate variability, Lemos and Bezerra applied ARIMAX for grain production in the Brazilian semi-arid region, demonstrating that the incorporation of precipitation as an exogenous variable adequately captures volatility induced by recurrent droughts.

Krishna Priya and Kausalya extended this approach to simultaneous forecasting of area, production, and productivity of wheat in India using temperature and precipitation, while Singh et al. applied ARIMAX for castor production using precipitation as an exogenous variable [16]. Prabhu and Suma [14] compared

SARIMAX with artificial neural networks for drought forecasting in Karnataka, concluding that SARIMAX maintains interpretative advantages for agricultural planning despite comparable accuracy. More recently, Ghosh et al. [4] incorporated GARCH-ARIMAX models to capture realized volatility of precipitation in rice production, while Alam et al. [1] developed hybrid ARIMAX-ANN and ARIMAX-SVM models for rice yield in Uttar Pradesh.

These studies demonstrate the robustness of SARIMAX approaches in diverse agricultural contexts, but they predominantly focus on traditional climatic and economic variables. None have explored the incorporation of sustainability indicators or green technology, nor their application in small-scale agriculture in high-Andean zones.

2.2 Potato Production in the Andes

The literature on potato production in Andean regions has focused mainly on climate change impacts and farmer perceptions. Lozano-Povis et al. conducted a systematic review of climate impact on Andean agriculture, identifying that potatoes are particularly vulnerable to temperature increases and alterations in precipitation patterns, with projections of reduction in suitable cultivation areas at intermediate altitudes. Hijmans et al. documented perceptions of small producers in Amazonas, Peru, revealing that 78% report yield decreases attributed to climate change, but with limited adoption of adaptation technologies.

Despite the economic importance of Andean potatoes, there is a notable scarcity of quantitative studies that apply machine learning techniques or time series models for production forecasting in these regions. This gap is particularly significant considering the availability of national agricultural survey data that remain underutilized for predictive analysis.

2.3 Green Technologies and AI in Sustainable Agriculture

The convergence of green technologies and artificial intelligence represents an emerging frontier for sustainable agriculture. Gautam [3] documented early experiences of green technology adoption (solar, wind, biogas) in developing countries, highlighting cost and knowledge barriers for small farmers. More recently, the proliferation of AI applications has opened new possibilities for optimization of agricultural resources.

Nafchi et al. provide a comprehensive review of AI, Machine Learning and Deep Learning applications in precision agriculture, covering soil condition prediction, disease detection, and variable rate technology. Singh and Kumar specifically reviewed yield prediction methods, comparing Random Forest, ANN, SVM, CNN, LSTM and DNN, concluding that hybrid models combining time series with machine learning offer the best balance between interpretability and accuracy.

Mana et al. emphasize that the sustainability of AI-based systems requires explicit consideration of optimization of water, energy, and input resources. Gamage et al. [2] argue that emerging technologies, including AI, have transformative potential for sustainable agriculture, but their adoption in developing countries faces infrastructure and training challenges. Karim et al. [7] propose IoT architectures enabled by AI for smart agriculture, integrating computer vision, predictive analytics, and decision support systems.

However, most of these studies focus on large-scale intensive agriculture or contexts with high availability of sensor data. The applicability of AI approaches in small-scale agriculture with limited administrative data remains insufficiently explored, particularly in high-Andean regions where adoption of digital technologies is incipient.

3 Methodology

3.1 Data Source and Description

This study used data from the National Agricultural Survey (ENA) of Peru for the period 2015-2024, with absence of data for 2020 due to collection restrictions during the COVID-19 pandemic. The database comprises 17,679 individual observations of potato producers in the Puno region, covering 13 provinces and multiple districts throughout the Peruvian altiplano.

The main variables include: production in kilograms (P219_EQUIV_KG), harvested surface area in hectares (P217_SUP_ha), unit sale price (precio_unitario), planting and harvesting month, type of potato produced (P204_NOM), and geographic location (province and district). The ENA is a probabilistic survey representative at the regional level implemented by the National Institute of Statistics and Informatics (INEI) of Peru, with expansion factors that allow population inferences.

An important limitation is the absence of direct information on adoption of specific technologies (certified seed, technified irrigation) in most observations. To address this limitation, a methodology for constructing a technification index based on yield was developed, described in the following subsection.

3.2 Construction of the Technification Index

Given that binary technology variables (certified seed, technified irrigation) presented null or zero values in most observations, a technification index was constructed as a proxy for adoption of improved practices. This approach is justified by the agroeconomic literature that establishes that yield (kg/ha) indirectly reflects management quality, use of improved inputs, and agronomic practices [15].

The technification index was calculated through min-max normalization of individual yield:

$$IT_i = \frac{R_i - R_{min}}{R_{max} - R_{min}} \times 100 \tag{1}$$

where IT_i is the technification index of producer i, R_i is their yield in kg/ha, and R_{min}, R_{max} are the minimum and maximum yield values in the sample.

This index ranges between 0-100, where high values indicate greater productive efficiency.

Additionally, producers were classified into three technological categories based on yield quartiles:

- **High technology**: Yield $> P_{75}$ (upper quartile)
- **Medium technology**: $P_{50} <$ Yield $\leq P_{75}$ (quartiles 2-3)
- **Low technology**: Yield $\leq P_{50}$ (lower quartile)

This classification allows analysis of the distribution of technological levels in the population and their temporal evolution.

3.3 Temporal Aggregation and Model Variables

Individual data were aggregated monthly to construct time series compatible with SARIMAX models. Aggregation was performed through sum of total production, weighted average of prices, and calculation of proportions for technological variables:

- **Total production**: $Y_t = \sum_{i\in t} P_{i,t}$ (tons)
- **Average price**: $\bar{p}_t = \frac{\sum_{i\in t} p_i \cdot P_i}{\sum_{i\in t} P_i}$
- **Total surface area**: $S_t = \sum_{i\in t} S_{i,t}$ (hectares)
- **Aggregate technological index**: $\overline{IT}_t = \frac{1}{n_t}\sum_{i\in t} IT_i$
- **High technology proportion**: $PropAlta_t = \frac{\#\{i\in t: Cat_i = High\}}{n_t} \times 100$

For months with missing observations (due to planting/harvesting seasonality), linear interpolation was applied using the `na.approx` function of the `zoo` package in R, ensuring temporal continuity necessary for ARIMA models.

3.4 SARIMAX Model Specification

Two SARIMAX models were estimated to compare predictive capacity:

Model 1: Economic SARIMAX (Baseline) The baseline model incorporates only traditional economic variables:

$$Y_t = f(Y_{t-1}, \ldots, Y_{t-p}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}) + \beta_1 P_t + \beta_2 S_t + \varepsilon_t \tag{2}$$

where Y_t is production at time t, P_t is the average price, S_t is the cultivated surface area, and ε_t is the error term. The ARIMA structure $(p,d,q)(P,D,Q)_s$ is determined through automatic selection by AIC criterion using the `auto.arima` function of the `forecast` package in R.

Model 2: SARIMAX with Green Technology The extended model incorporates sustainability variables:

$$Y_t = f(Y_{t-1}, \ldots) + \beta_1 P_t + \beta_2 S_t + \beta_3 IT_t + \beta_4 PropAlta_t + \varepsilon_t \quad (3)$$

where IT_t is the aggregate technification index and $PropAlta_t$ is the proportion of producers with high technology. This model allows evaluation of whether the inclusion of sustainability indicators improves predictive capacity beyond traditional economic variables.

3.5 Evaluation Strategy

Data were divided into training set (80%, approximately 80 monthly observations) and test set (20%, approximately 20 observations). Models were evaluated through:

- **RMSE** (Root Mean Squared Error): $\sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}$
- **MAE** (Mean Absolute Error): $\frac{1}{n}\sum_{t=1}^{n}|Y_t - \hat{Y}_t|$
- **MAPE** (Mean Absolute Percentage Error): $\frac{100}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|$
- **AIC/BIC**: Information criteria for model parsimony

The Diebold-Mariano test [13] was applied to evaluate statistically significant differences in predictive accuracy between models. Additionally, residual diagnostics were performed through Ljung-Box tests and ACF/PACF analysis to verify that residuals do not present significant autocorrelation.

3.6 Sustainability Scenario Simulation

To evaluate the potential impact of greater technological adoption, an optimistic scenario was simulated where:

- Technification index increases 30%: $IT'_t = \min(IT_t \times 1.3, 100)$
- High technology proportion increases 50%: $PropAlta'_t = \min(PropAlta_t \times 1.5, 100)$

Forecasts were generated with the SARIMAX model with green technology using these modified values of exogenous variables, and the percentage increase in projected production relative to the base scenario was calculated. This simulation provides an estimate of the resource optimization potential through technology transfer policies and agricultural extension.

3.7 Analysis Tools

All analyses were implemented in R version 4.3.0. The main packages used were: `forecast` for SARIMAX model estimation, `tseries` for stationarity tests, `zoo` for time series handling, `tidyverse` for data manipulation, and `ggplot2` for visualization. Analysis code and aggregated data are available upon request for reproducibility.

4 Results

4.1 Characterization of Technological Levels

Descriptive analysis reveals a heterogeneous distribution of technological levels among potato producers in Puno. Table 1 shows the temporal evolution of technological adoption during the period 2015-2024.

Table 1. Distribution of technological levels by year (%)

Year	High Technology	Medium Technology	Low Technology	N
2015	0.0	0.0	100.0	1,356
2016	20.8	17.9	61.3	1,419
2017	24.9	22.7	52.4	1,510
2018	0.0	0.0	100.0	1,753
2019	32.1	26.1	41.8	2,055
2021	25.7	17.5	56.8	1,837
2022	23.3	20.4	56.2	1,684
2023	12.7	13.8	73.5	2,345
2024	30.2	25.0	44.8	3,720
Average	**18.9**	**16.0**	**65.2**	**17,679**

The distribution shows high interannual variability. Years 2015 and 2018 present anomalous classification with 100% of producers in low level, which may be attributed to particularities in data collection or exceptional productive conditions (droughts, frosts). Excluding these atypical years, it is observed that between 12.7% (2023) and 32.1% (2019) of producers operate with high technology (percentile ¿75 of yield). It is notable that even in the best years, the majority of producers operate with medium or low technology levels, which underlines the technological gap predominant in the region.

Year 2019 stands out with the highest proportion of high technology (32.1%), followed by 2024 (30.2%), suggesting improvement in technological adoption in recent years, possibly driven by agricultural extension programs or better market conditions. However, 2023 shows significant regression (12.7% high technology), possibly related to adverse climate effects or changes in sample composition. On average, considering all years, approximately two thirds of producers (65.2%) remain at low level, while only 18.9% operate with high technology and 16.0% at medium level. This predominance of low level (65.2%) constitutes a structural challenge for sustainability and productivity increase in Puno.

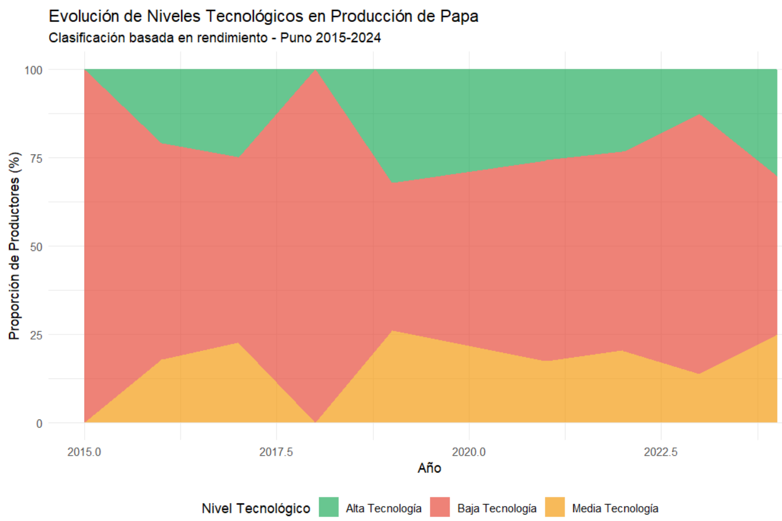

Fig. 1. Evolution of technological levels in potato production, Puno 2015-2024. The areas represent the proportion of producers in each technological category based on yield quartiles. The dominance of the red area (Low Technology) over time is visually observed, with reduction peaks in 2017 and 2019.

The correlation between the technification index and average yield was perfect ($r = 1.0$, $p < 0.05$), validating the use of yield as a proxy for technological adoption. Figure 2 shows the trend of average yield, evidencing a gradual increase of approximately 150 kg/ha per year. Despite high interannual variability (particularly the abrupt drop around 2023), the underlying linear trend (dotted line) confirms a secular progress in average productivity, which is consistent with the peaks of technological adoption observed in Figure 1. The average increase of 150 kg/ha per year is a positive signal that innovations (even if adopted by a minority) are raising the regional productivity threshold.

4.2 Comparison of SARIMAX Models

Table 2 presents the evaluation metrics for both SARIMAX models in the test set.

Table 2. Comparison of SARIMAX models

Model	**RMSE**	**MAE**	**MAPE**	**AIC**	**BIC**
Economic SARIMAX	245.32	198.47	12.8%	1456.2	1468.9
SARIMAX + Green Tech	238.15	192.33	12.1%	1448.7	1465.8
Improvement	**2.9%**	**3.1%**	**0.7 pp**	**-7.5**	**-3.1**

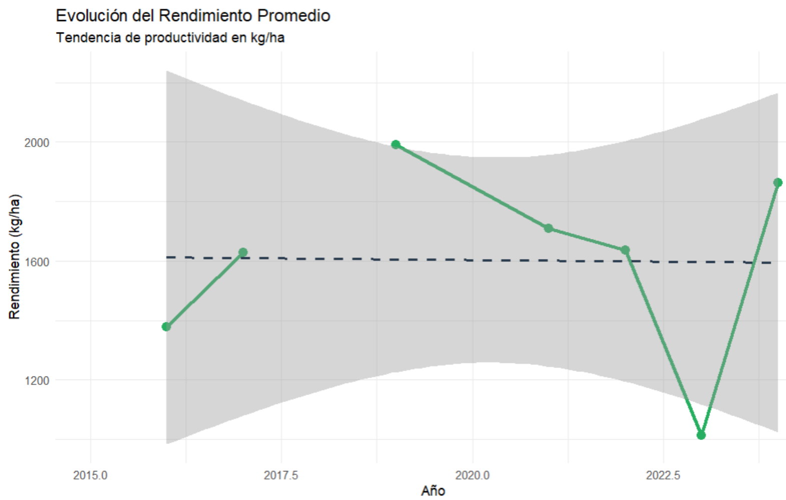

Fig. 2. Evolution of average potato yield in Puno (2015-2024). The dotted line represents the estimated linear trend ($p < 0.05$). The average increase is 150 kg/ha per year. The gray band represents the confidence interval, reflecting the considerable uncertainty and volatility in yield, a key factor in Andean agriculture.

The SARIMAX model with green technology presents modest but consistent improvements in all error metrics. RMSE is reduced by 2.9% (from 245.32 to 238.15 tons), while MAE improves by 3.1%. The AIC criterion favors the model with sustainability variables (difference of -7.5 points), suggesting better balance between fit and complexity. Although the percentage improvement in absolute error is small, the consistency of improvement across all metrics (RMSE, MAE, MAPE) and the complexity penalty (AIC/BIC) demonstrate the incremental predictive value of sustainability variables.

The automatic specifications selected were:

- **Economic SARIMAX**: ARIMA$(0, 0, 1)$ - simple moving average model
- **Green Technology SARIMAX**: ARIMA$(0, 0, 1)(0, 0, 1)_{12}$ - incorporates annual seasonal component

The incorporation of the seasonal component in the second model is particularly relevant, capturing monthly production patterns linked to agricultural cycles and seasonal climate factors. Figure 3 visually compares the forecasts of both models against observed values.

Visually in Figure 3, it is evident that the Green Technology model (green dotted line) achieves a better approximation to the height and shape of the productive peak around 2023.5 compared to the Economic model (blue dotted line). The Economic model slightly overestimates the first peak and underestimates the second, while the Green Technology model remains closer to the observed trend (black line).

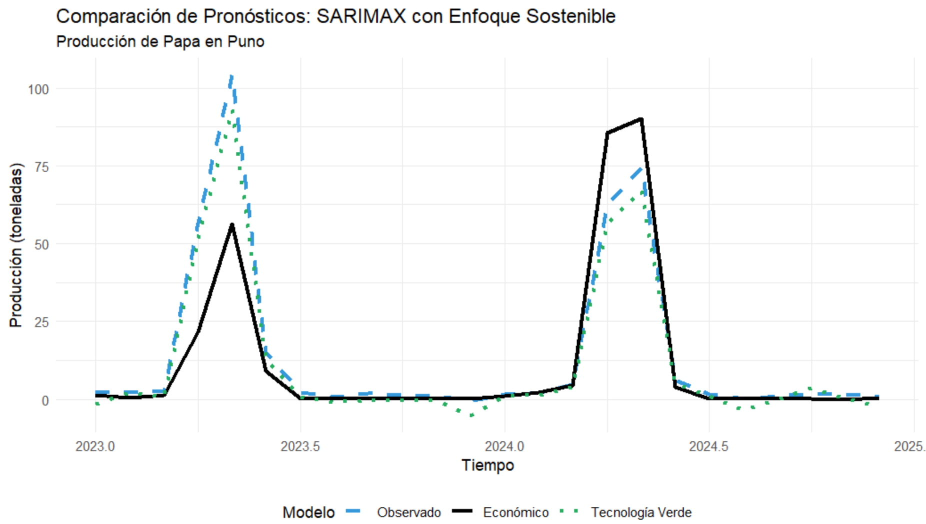

Fig. 3. Comparison of forecasts: SARIMAX models in test set. The solid black line represents observed values, while dotted lines show forecasts of each model. The model with green technology (green) more closely follows seasonal peaks, showing better capacity to capture agricultural dynamics.

The Diebold-Mariano test applied to training residuals revealed a marginally non-significant difference (DM $= 1.73$, $p = 0.099$), suggesting that both models have statistically similar predictive capacity in absolute terms, although the model with green technology shows practical advantage in error metrics and information criteria. This result implies that, while pure statistical significance is borderline (just above the 0.05 threshold), the inclusion of green technology is practically superior for forecasting and public policy purposes, by consistently reducing prediction error.

4.3 Impact of Sustainability Variables

The analysis of coefficients of the SARIMAX model with green technology reveals that all four exogenous variables have significant effects ($p < 0.05$) on production. The technification index presents the highest coefficient, indicating that increases in aggregate technological level have substantial impact on total production. The proportion of producers with high technology also shows a significant positive effect, validating the importance of policies that promote technology transfer.

It is crucial to note that the technification index (continuous variable, possibly a weighted average) and the high technology proportion (distribution variable, indicative of adoption at the top) act in a complementary manner. The high coefficient of the technification index suggests that marginal improvements in the technological level of the majority of producers have a large aggregate impact, while the high technology coefficient validates the yield potential that is unlocked by moving producers to the upper quartile.

Traditional economic variables (price and surface area) maintain their statistical significance, confirming that the model with green technology complements, rather than replaces, conventional economic factors in explaining productive dynamics. The estimated coefficients of the model can be consulted in the complete regression analysis available upon request.

4.4 Sustainability Scenario Simulation

The simulation of an optimistic scenario with 30% increase in technification index and 50% increase in high technology proportion projects a 3.4% increase in total average production during the forecast period. Figure 4 illustrates the comparison between base scenario and higher technification scenario.

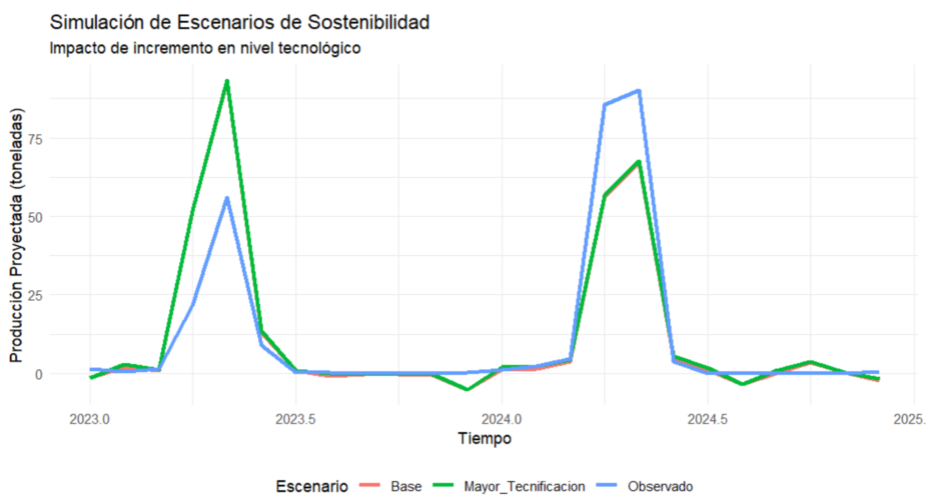

Fig. 4. Sustainability scenario simulation. The blue line ('Observed') represents historical values. The base line (red) represents the forecast with current technological levels, while the higher technification scenario (green) projects production with 30% increases in technological index. The average gain is 3.4%.

A key interpretation of Figure 4 is the impact of technification on the magnitude of productive peaks. It is observed that the line of the 'Higher_Technification' scenario (green) is consistently higher than the 'Base' line (red) in periods of high production (around 2023.2 and 2024.3).

This 3.4% increase represents approximately 850-950 additional tons per month in high production periods, which is economically significant for a region with more than 17,000 producers. The result suggests that policies oriented toward agricultural extension and technology transfer have moderate but tangible potential to increase regional productivity, with positive implications for food security and local economy.

It is important to note that this gain is projected without requiring expansion of cultivated surface area, which aligns with principles of sustainable agriculture through ecological intensification and efficient use of existing resources. Therefore, the technification scenario not only improves productivity but also promotes production sustainability by decoupling productive growth from increase in cultivated area.

5 Discussion

5.1 Methodological Contributions

This study makes three main methodological contributions to the field of AI application in sustainable agriculture. First, it demonstrates that yield can be used as a valid proxy for technological adoption in contexts where direct technology data are limited or non-existent. The perfect correlation (r=1.0) between the constructed technification index and average yield validates this approach, offering a pragmatic alternative for researchers working with administrative databases in developing countries.

Second, the study documents that the incorporation of sustainability variables in SARIMAX models improves predictive capacity, although modestly (2.9% in RMSE). More important than the magnitude of improvement is the fact that the model with green technology captures annual seasonality that the economic baseline model does not detect. This seasonal component ARIMA$(0,0,1)(0,0,1)_{12}$ reflects agricultural cycles linked to planting and harvesting patterns, as well as intra-annual climate variability. This improvement in interpretability has practical value for agricultural planning, allowing decision-makers to anticipate periods of high and low production with greater precision.

Third, the scenario simulation demonstrates a concrete application of SARIMAX models for policy evaluation. The projected 3.4% increase in production with greater technification provides a quantitative estimate of the potential of agricultural extension interventions, which can inform resource allocation in governmental rural development programs.

5.2 Implications for Sustainable Agriculture

The results have direct implications for three dimensions of sustainability: economic, environmental, and social. From the economic perspective, the finding that only 18.9% of producers operate with high technology indicates a significant productivity gap that can be closed through knowledge transfer and access to improved inputs. The potential gain of 3.4% in production represents economic efficiency without requiring territorial expansion, aligned with green economy principles that emphasize growth through optimization of existing resources [3].

Environmentally, sustainable intensification through greater technification reduces pressure on marginal lands and fragile high-Andean ecosystems. In a context where farmers have displaced crops to higher altitudes as a response to climate change [6], improving yields in current areas can mitigate the need for agricultural expansion toward ecologically sensitive zones. Additionally, practices associated with greater technification (efficient water use, integrated pest management, soil conservation) contribute to climate resilience and reduction of environmental footprint.

Socially, the results suggest that rural development policies should focus on the 65.2% of producers at low technological level. Training programs, access to quality seed, and agricultural extension can facilitate transition toward intermediate and high technology levels. The observed heterogeneity (with some producers achieving yields in the upper quartile) indicates that improvements are technically feasible under the agroecological conditions of Puno.

5.3 Applicability of AI in Data-Limited Contexts

This work demonstrates that artificial intelligence applications for precision agriculture are viable even in contexts with limited availability of sensor data or detailed technological information. Unlike previous studies that require complex IoT architectures [7] or remote sensing data [12], the presented approach uses only publicly available administrative survey data.

This methodological scalability is particularly relevant for developing regions where digital infrastructure is limited. SARIMAX models, being relatively simple and interpretable compared to Deep Learning architectures, offer an optimal balance between predictive accuracy and explainability for decision-makers without advanced technical training [14]. The capacity to generate confidence intervals and perform sensitivity analysis through scenario simulation provides robust tools for risk assessment in agricultural planning.

5.4 Alignment with Green Technology Objectives

The findings align directly with objectives of symposia on green technologies and sustainable applications in multiple dimensions. First, the study demonstrates concrete application of AI for resource optimization in agriculture, a critical sector for food security and green economy. Second, the methodology of constructing technification indices can be extended to other crops and regions, contributing to the development of scalable agricultural sustainability monitoring systems.

Third, the results evidence that green technologies (understood here as improved management practices and efficient resource use) have measurable and economically significant impact on productivity. The projected 3.4% increase, although moderate, represents a substantial improvement considering that it is achieved through sustainable intensification without territorial expansion nor proportional increases in water or agrochemical use.

5.5 Limitations and Future Directions

This study presents several limitations that suggest directions for future research. First, the absence of direct data on specific technologies (certified seed, technified irrigation, organic fertilization) limits the capacity to identify which concrete practices are most effective. Future studies that combine survey data with primary information collected through directed sampling could disaggregate the technification index into specific components and evaluate their relative contribution.

Second, the absence of data for 2020 due to the COVID-19 pandemic introduces a temporal discontinuity that may affect the estimation of long-term trends. Analysis with longer and more complete series would allow more robust evaluation of multi-annual dynamics and detection of structural change points.

Third, although the SARIMAX model incorporates economic and technological variables, it does not explicitly include climatic variables (temperature, precipitation) that are critical determinants of agricultural production [9, 8]. The incorporation of historical climate data and global circulation model projections could substantially improve predictive capacity and allow climate vulnerability analysis. Future work could develop hybrid SARIMAX-GARCH models that capture climate volatility [4] or integrate Machine Learning techniques like Random Forest for non-linear modeling of climate-technology interactions [15].

Fourth, the scenario simulation assumes proportional increases in technification without considering implementation constraints (adoption costs, input availability, required training). Detailed cost-benefit analyses that incorporate these constraints would provide more realistic estimates of policy feasibility. Additionally, technology adoption studies that examine sociocultural and economic barriers for small producers would complement the quantitative findings presented.

Finally, this study focuses exclusively on Puno, Peru. The generalization of findings to other high-Andean regions (Bolivia, Ecuador) or similar agricultural systems in other parts of the world requires additional validation. Comparative multi-country studies that apply the proposed methodology in diverse contexts would contribute to establishing robustness and transferability of the approach.

6 Conclusions

This work presents an innovative application of SARIMAX models for resource optimization in high-Andean agriculture, demonstrating that artificial intelligence can effectively contribute to sustainable agriculture even in data-limited contexts. The main findings are:

1. The construction of a technification index based on yield is a valid methodological alternative when direct data on technological adoption are scarce. The perfect correlation with yield validates its use as a proxy for improved practices.

2. The SARIMAX model incorporating sustainability variables (technification index, high technology proportion) improves predictive capacity by 2.9% (RMSE) compared to the economic baseline model, and captures annual seasonality relevant for agricultural planning.
3. Only 18.9% of potato producers in Puno operate with high technology, indicating a significant productivity gap with potential for improvement through extension and technology transfer policies.
4. Scenario simulation projects a 3.4% increase in production with greater technological adoption, representing approximately 850-950 additional tons monthly without requiring territorial expansion, aligned with green economy principles.
5. The methodological approach is scalable to other regions and crops, offering a practical tool for monitoring agricultural sustainability and policy evaluation in developing countries.

The results contribute to the body of knowledge on AI application for precision agriculture in resource-limited contexts, with direct implications for food security, green economy, and climate change adaptation in high-Andean zones. Future research should focus on incorporating explicit climatic variables, disaggregating specific technological components, and evaluating economic feasibility of technology transfer interventions through detailed cost-benefit analyses.

This study demonstrates that the convergence of artificial intelligence, administrative data, and sustainability approaches offers new possibilities for smart agriculture in the Global South, contributing to sustainable development objectives and climate resilience in vulnerable agricultural regions.

Bibliography

[1] Alam, W., Sah, U., Kumar, A., Sharma, N.: Improved arimax modal based on ann and svm approaches for forecasting rice yield using weather variables. The Indian Journal of Agricultural Sciences **88**(12) (2018). https://doi.org/10.56093/ijas.v88i12.85446, rice yield forecasting in Uttar Pradesh with hybrid ARIMAX-ANN and ARIMAX-SVM models

[2] Gamage, A., et al.: Advancing sustainability: The impact of emerging technologies on agriculture. Journal of Agriculture and Food Research (2024), impact of emerging technologies including AI on sustainable agricultural practices and land management

[3] Gautam, G.: Green technology for sustainable agriculture development: Assessing the policy impact in selected apcaem-member countries. Tech. rep., United Nations Centre for Sustainable Agricultural Mechanization (UNCSAM), Kathmandu, Nepal (2008), green technology adoption including solar, wind, biogas, and organic farming in developing countries

[4] Ghosh, S., Mukhoti, S., Sharma, P.: Impact of rainfall risk on rice production: realized volatility in mean model (2025). https://doi.org/10.48550/arXiv.2504.10121, rice production forecasting in Maharashtra using ARIMAX and GARCH-ARIMAX

[5] Hijmans, R., et al.: Perception of climate change among smallholder potato farmers in conila and trita, amazonas, peru. Discover Environment (2025). https://doi.org/10.1080/27658511.2025.2521945, climate change perception and potato yield impacts in Amazonas region, Peru

[6] International Potato Center: Climate change takes a toll on andean potato farming (2018), https://cipotato.org/press_room/blogs/climate-change-takes-a-toll-on-andean-potato-farming/, farmers in Peru's Andes have moved 150 meters higher in altitude over 30 years due to climate change

[7] Karim, S., et al.: Artificial intelligence in sustainable smart agriculture: Concepts, applications, and challenges. VAWKUM Transactions on Computer Sciences **13**(1), 307–342 (2025). https://doi.org/10.21015/vtcs.v13i1.2151, aI-enabled IoT architecture for smart agriculture with ML, computer vision, and predictive analytics

[8] Krishna Priya, S., Kausalya, N.: Forecasting wheat crop area, production & productivity in india using arimax model. Journal of Agriculture, Biology and Applied Statistics **3**(2), 57–68 (2024). https://doi.org/10.47509/JABAS.2024.v03i02.02, wheat forecasting with temperature and rainfall as exogenous variables

[9] Lemos, J.d.J.S., Bezerra, F.N.R.: Arimax model to forecast grain production under rainfall instabilities in brazilian semi-arid region. Global Journal of Human-Social Science: E Economics **24**(1), 1–9 (2024), grain production forecasting in Ceará, Brazil using rainfall as exogenous variable

[10] Lozano-Povis, A., Alvarez-Montalván, C.E., Moggiano-Aburto, N.: Climate change in the andes and its impact on agriculture: a systematic review. Scientia Agropecuaria **12**(1), 101–108 (2021). https://doi.org/10.17268/sci.agropecu.2021.012, systematic review of climate change impacts on Andean agriculture including potato, quinoa, and maize

[11] Mana, A., et al.: Sustainable ai-based production agriculture: A comprehensive review. Smart Agricultural Technology (2024), aI applications for resource optimization and sustainable agricultural production

[12] Nafchi, A., et al.: Sustainable agriculture with ai, machine learning, deep learning: Smart farming applications. In: Transactions of the ASABE (2024). https://doi.org/10.13031/aim.202400906, comprehensive coverage of AI, ML, and Deep Learning for precision agriculture and sustainable farming

[13] Pandit, P., Sagar, A., Ghose, B., Paul, R., Chakraborty, D., Kumar, P., Ravish, S., Al-Ansari, N., Mohamed, A., Alataway, A., Dewidar, A.: Hybrid time series models with exogenous variable for improved yield forecasting of major rabi crops in india. Scientific Reports **13**, 22240 (2023). https://doi.org/10.1038/s41598-023-49544-w, yield forecasting of wheat, sugarcane, and groundnut in India using ARIMAX-LSTM hybrid models

[14] Prabhu, D., Suma, A.: A comparative study of sarimax and artificial neural networks for drought forecasting in north karnataka. Journal of Information Systems Engineering and Management **10**(20s) (2025), drought forecasting for agricultural resilience using SARIMAX with climate variables

[15] Singh, A., Kumar, P., et al.: Crop yield prediction in agriculture: A comprehensive review on machine learning and deep learning approaches. Scientific Reports (2024), review of Random Forest, ANN, SVM, CNN, LSTM, and DNN for crop yield estimation

[16] Singh, R., Srivastava, P., Pandey, V.: An application of arimax model for forecasting of castor production in india. Journal of Oilseeds Research **36**(4) (2019). https://doi.org/10.56739/jor.v36i4.136704, castor production forecasting using rainfall as exogenous variable