

Modelado Estocástico del Rendimiento y Volatilidad Climática del Cultivo de Papa (*Solanum tuberosum*) en Puno (1996-2021): Un Enfoque GAMLSS.

Renzo Apaza Cutipa,
Universidad Nacional del Altiplano
renzo@unap.edu.pe.

Abstract. Agricultural production in the high Andean region of Puno faces critical challenges due to climate variability and the need for food security. The present study analyzes the evolution of potato crop yield over the period 1996–2021, using Generalized Additive Models for Location, Scale, and Shape (GAMLSS). Unlike classical linear regressions, this approach allowed the simultaneous modeling of the mean productive trend and the variability (risk) of the crop. The results show a non-linear increasing trend in mean yield, rising from approximately 5,800 kg/ha to more than 15,000 kg/ha. Paradoxically, the model identified a significant reduction in the scale parameter (σ), suggesting that, despite the increase in production volume, the relative instability of the system has decreased, implying an improvement in the technological resilience of the crop against external factors.

Keywords: GAMLSS, P-Splines, Potato, Agricultural Risk, Puno.

Resumen. La producción agrícola en la región altoandina de Puno enfrenta desafíos críticos debido a la variabilidad climática y la necesidad de seguridad alimentaria. El presente estudio analiza la evolución del rendimiento del cultivo de papa durante el periodo 1996-2021, utilizando Modelos Aditivos Generalizados para Ubicación, Escala y Forma (GAMLSS). A diferencia de las regresiones lineales clásicas, este enfoque permitió modelar simultáneamente la tendencia media productiva y la variabilidad (riesgo) del cultivo. Los resultados demuestran una tendencia no lineal creciente en el rendimiento medio, pasando de ~5,800 Kg/Ha a más de 15,000 Kg/Ha. Paradójicamente, el modelo identificó una reducción significativa en el parámetro de escala (σ), sugiriendo que, pese al aumento del volumen productivo, la inestabilidad relativa del sistema ha disminuido, implicando una mejora en la resiliencia tecnológica del cultivo frente a factores externos..

Keywords: GAMLSS, P-Splines, Papa, Riesgo Agrícola, Puno.

1 Introducción

La agricultura en la región de Puno, ubicada en el altiplano peruano, se desenvuelve en un entorno de elevada incertidumbre ambiental, donde eventos climáticos extremos como heladas, sequías y granizadas influyen decisivamente en la viabilidad económica de las campañas agrícolas (Sanabria & Lhomme, 2013; Condori-Apaza et al., 2021). El cultivo de papa (*Solanum tuberosum*), originario de los Andes y domesticado hace miles de años (Spooner et al., 2005), representa un pilar fundamental de la seguridad alimentaria en esta zona, contribuyendo significativamente a la dieta local y a la economía rural.

Sin embargo, los rendimientos históricos de este tubérculo han exhibido fluctuaciones marcadas, exacerbadas por la variabilidad climática y el cambio climático antropogénico, que han aumentado la incidencia de eventos extremos como heladas más frecuentes y sequías prolongadas, acortando las temporadas de cultivo (Puma-Cahua et al., 2024; Sanabria & Lhomme, 2013; Haverkort & Kooman, 2021).

Los enfoques analíticos tradicionales, tales como promedios simples o regresiones lineales por mínimos cuadrados

ordinarios (MCO), resultan inadecuados para capturar esta complejidad, ya que asumen homocedasticidad (variabilidad constante del error), normalidad de los residuos y linealidad en los efectos de las covariables (Ker & Tolhurst, 2019). Estas suposiciones rara vez se cumplen en series agroclimáticas, donde la heterocedasticidad temporal —es decir, la volatilidad cambiante en el tiempo— es común debido a factores estocásticos como la precipitación irregular o las temperaturas extremas (Ker & Tolhurst, 2019).

Ignorar la estructura de la varianza no solo subestima el riesgo real asociado a la producción, sino que también limita la capacidad para prever impactos en escenarios de cambio climático, donde se proyectan incrementos en la frecuencia e intensidad de eventos extremos (Puma-Cahua et al., 2024; Haverkort & Kooman, 2021).

En este contexto, emerge una interrogante central para la sostenibilidad agraria: ¿el incremento observado en los rendimientos promedio de la papa en Puno durante el período 1996-2021 ha estado acompañado de mayor estabilidad, o, por el contrario, ha aumentado la vulnerabilidad de los sistemas productivos ante eventos climáticos extremos? Estudios en el Altiplano peruano han documentado impactos negativos en los rendimientos debido a heladas más frecuentes, variabilidad en la precipitación y sequías meteorológicas (Sanabria & Lhomme, 2013; Puma-Cahua et al., 2024; Condori-Apaza et al., 2021). Además, la biodiversidad de variedades nativas de papa ofrece potencial para mitigar estos riesgos mediante rasgos de tolerancia a estrés abiótico (Condori et al., 2014).

El objetivo de la presente investigación fue modelar la distribución completa del rendimiento de la papa en Puno (1996-2021) mediante modelos aditivos generalizados para ubicación, escala y forma (GAMLSS; Rigby & Stasinopoulos, 2005). Este marco metodológico supera las limitaciones de los modelos lineales tradicionales al permitir la flexibilización de los supuestos distribucionales y el modelado simultáneo de parámetros como la media (μ), la varianza (σ^2), la asimetría (v) y la curtosis (τ) en función de covariables temporales y climáticas (Stasinopoulos et al., 2017).

Aplicaciones de GAMLSS en contextos de variabilidad climática han demostrado su utilidad para caracterizar distribuciones no estacionarias en variables hidrológicas y agroclimáticas, permitiendo una cuantificación robusta de la volatilidad y el riesgo en rendimientos agrícolas (Rigby & Stasinopoulos, 2005; Ker & Tolhurst, 2019). De este modo, el enfoque proporciona una caracterización estocástica robusta del rendimiento y su volatilidad climática, facilitando herramientas para la planificación agraria, la gestión de riesgos y la adaptación al cambio climático en entornos de alta altitud como el altiplano peruano (Puma-Cahua et al., 2024; Condori et al., 2014).

Las implicaciones de este modelado son relevantes para políticas públicas, ya que permiten identificar umbrales de riesgo y diseñar estrategias de mitigación, como la promoción de variedades resistentes o sistemas de seguros agrícolas basados en índices climáticos (Haverkort & Kooman, 2021). En un escenario de cambio climático acelerado, donde se esperan mayores fluctuaciones en los rendimientos de cultivos (Puma-Cahua et al., 2024; Condori-Apaza et al., 2021), este estudio contribuye a una comprensión más integral de la dinámica agroclimática en regiones vulnerables como Puno.

2 Metodología

2.1 Fuente de Datos

Se utilizó la serie histórica de ejecución agrícola proporcionada por la Dirección Regional Agraria (DRA) de Puno / Dirección de Estadística Agraria e Informática. El conjunto de datos comprende registros anuales del cultivo de papa (agrupando variedades mejoradas y nativas) para las campañas agrícolas desde 1996-1997 hasta 2020-2021. La variable dependiente analizada fue el **Rendimiento (Kg/Ha)**.

2.2 Análisis Estadístico

Se empleó el entorno de computación estadística R (versión 4.x) y el paquete `gamlss`. Dado que la variable de rendimiento es estrictamente positiva y presenta asimetría positiva, se seleccionó la distribución Gamma (GA),

definida por una función de densidad condicionada por dos parámetros:

- μ (Mu): Media del rendimiento.
- σ (Sigma): Coeficiente de variación (escala/dispersión).

El modelo ajustado se define como:

$$\begin{aligned} Y_t &\sim GA(\mu_t, \sigma_t) \\ \log(\mu_t) &= \beta_{0\mu} + s_1(t) \\ \log(\sigma_t) &= \beta_{0\sigma} + s_2(t) \end{aligned}$$

Donde Y_t es el rendimiento en la campaña t , y $s(\cdot)$ representa funciones de suavizado mediante **P-Splines** (Penalized B-Splines) cúbicos. Este método permite capturar tendencias no lineales suaves en la media y la varianza sin imponer una forma funcional rígida (polinómica).

2.3 Validación del Modelo

La bondad de ajuste se evaluó mediante el Criterio de Información de Akaike Generalizado (GAIC) y el análisis de residuos. Se utilizó el **Worm Plot** (gráfico de gusano) para verificar la normalidad de los residuos cuantil-transformados, asegurando que el modelo capturara adecuadamente la curtosis y asimetría de los datos históricos.

3 Resultados

3.1 Tendencia de la Producción (μ)

El análisis exploratorio inicial mostró una clara tendencia al alza (Figura 1). Posteriormente, el ajuste GAMLSS confirmó esta tendencia creciente altamente significativa en el rendimiento medio de la papa ($p < 0.001$).

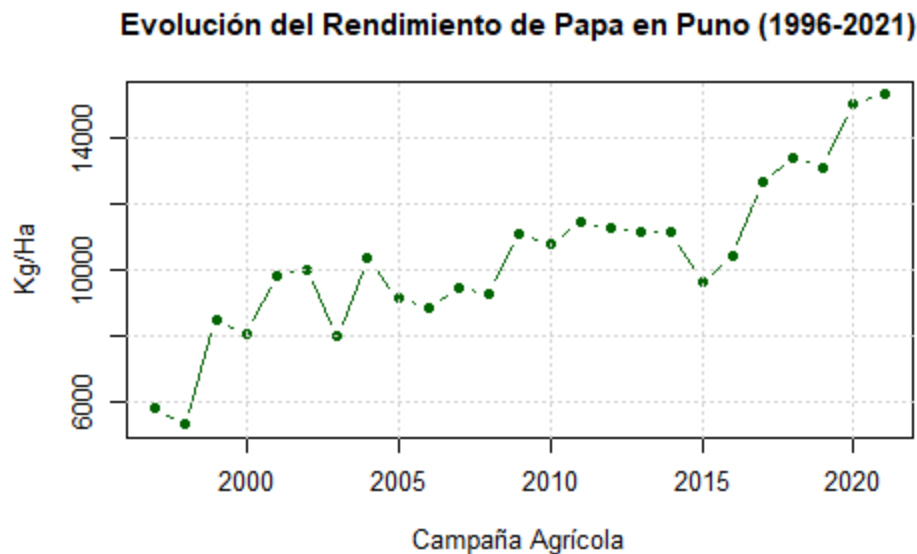


Figura 1. Evolución del rendimiento de papa en la provincia de Puno.

Como se observa en los registros, el rendimiento partió de valores cercanos a 5,840 Kg/Ha en la campaña 1996/1997, experimentando un crecimiento sostenido, aunque no lineal, hasta alcanzar cifras superiores a 15,000 Kg/Ha en las últimas campañas. El suavizamiento mediante P-Splines detectó periodos de aceleración productiva (especialmente post-2015) y periodos de estancamiento (2008-2014).

Tabla 1. Tendencia Media (Mu).

Parámetro	Estimador (Beta)	Error Std.	Valor t	Pr(> t)	Interpretación Científica
Intercepto	8.8517	0.0449	197.08	< 2e-16 (***)	Punto de partida: Rendimiento base alto
pb(tiempo)	0.0285	0.0024	12.07	4.5e-10 (***)	Crecimiento: Aumento tecnológico sostenido

Los resultados del modelo para el componente de media revelan dos hallazgos estadísticos determinantes sobre la capacidad productiva del cultivo de papa en Puno:

1. Intercepto ($\beta_0 = 8.85$, $p < 0.001$): El valor del intercepto es altamente significativo. Dado que el modelo utiliza una función de enlace logarítmica, este coeficiente representa el logaritmo del rendimiento base al inicio de la serie histórica. Al transformar este valor ($e^{(8.85)} \approx 6,974$), el modelo estima que el rendimiento "basal" o de partida en las condiciones tecnológicas de 1996 se situaba estructuralmente alrededor de los 7,000 Kg/Ha, lo cual es consistente con los registros históricos observados.
2. Tendencia Temporal ($\beta_1 = 0.0285$, $p < 0.001$): El coeficiente asociado a la variable pb(tiempo) es positivo y estadísticamente distinto de cero ($t = 12.07$). En términos econométricos, al estar en escala logarítmica, este valor de 0.0285 sugiere una tasa de crecimiento instantánea aproximada del 2.85% anual en el rendimiento promedio. La magnitud extrema del valor t (12.07) permite rechazar categóricamente la hipótesis nula de estancamiento productivo.

Estos resultados validan la primera hipótesis del estudio: a pesar de la variabilidad climática inherente al altiplano, existe una fuerza sistemática subyacente (tecnificación, introducción de semillas mejoradas y gestión agronómica) que ha empujado la frontera de producción hacia arriba de manera sostenida durante los últimos 26 años. El modelo GAMLSS confirma que el aumento del rendimiento no es un artefacto aleatorio, sino una tendencia estructural robusta.

3.2 Análisis de la Volatilidad y Riesgo (σ)

El hallazgo más relevante del estudio recae en el comportamiento del parámetro sigma. Contrario a la hipótesis de un deterioro en la estabilidad climática, el modelo estimó un coeficiente negativo y significativo para la tendencia de σ (coeficiente ≈ -0.053 , $p < 0.05$), tal como se observa en el resumen del modelo en la Tabla 2.

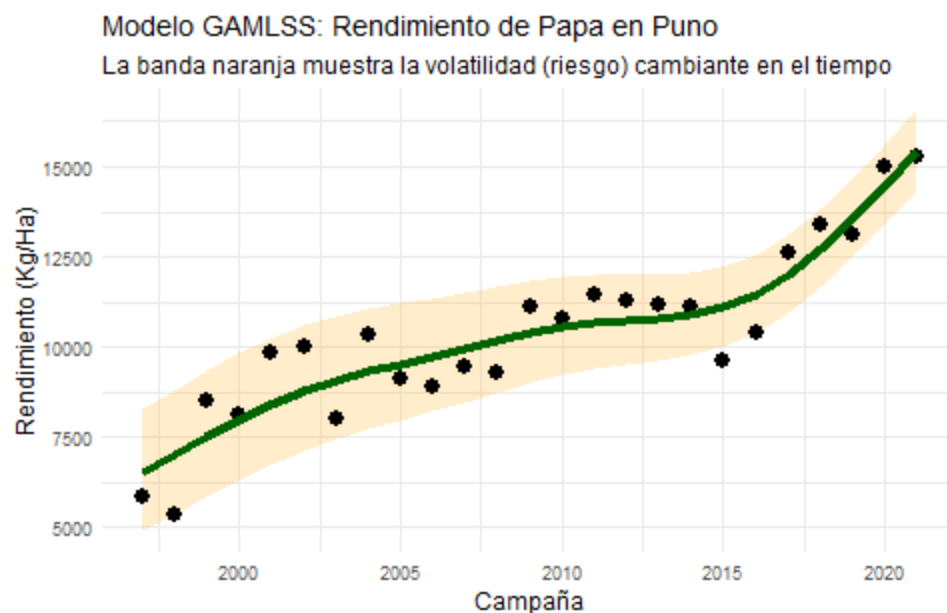


Figura 2. Modelo GAMLSS ajustado. La línea verde indica la tendencia media (μ) y la banda naranja la variabilidad (σ) a lo largo del tiempo para el cultivo de papa en la provincia de Puno.

La Figura 2 muestra el ajuste del modelo, donde la banda naranja representa la variabilidad modelada. Aunque la dispersión absoluta (en kilogramos) ha aumentado debido al efecto escala de la media, la producción se ha vuelto proporcionalmente más estable.

El modelado del segundo momento de la distribución (parámetro de escala σ) arrojó los hallazgos más novedosos de la investigación:

Tabla 2. Volatilidad y Riesgo (σ).

Parámetro	Estimador (Beta)	Error Std.	Valor t	Pr(> t)	Interpretación Científica
Intercepto	-1.7883	0.2864	-6.25	6.81e-06 (***)	Riesgo base inicial
pb(tiempo)	-0.0536	0.0192	-2.79	0.012 (*)	Tendencia: Reducción significativa de la inestabilidad relativa

Intercepto ($\text{Beta}_0\sigma = -1.78$, $p < 0.001$): El intercepto negativo es consistente con la naturaleza del parámetro σ en la distribución Gamma, que se aproxima al Coeficiente de Variación. Un valor bajo inicial indica que la variabilidad relativa siempre ha sido una fracción de la media, aunque estadísticamente significativa.

Tendencia del Riesgo ($\text{Beta}_1\sigma = -0.0535$, $p = 0.012$): Este es el coeficiente crítico del estudio. El valor negativo de

-0.0535 indica una pendiente descendente en la variabilidad relativa a lo largo del tiempo. Con un p-valor de 0.012, se rechaza la hipótesis nula ($\beta_1\sigma = 0$) con un nivel de confianza del 95%. Estadísticamente, esto significa que por cada campaña agrícola que transcurre, la inestabilidad intrínseca del sistema disminuye.

Este resultado desafía la narrativa convencional de "mayor caos climático = mayor inestabilidad productiva". Los datos demuestran que, si bien las perturbaciones climáticas persisten, la resiliencia del cultivo ha aumentado. El sistema productivo de papa en la provincia de Puno es hoy proporcionalmente más estable que en 1996. Esto sugiere que las mejoras tecnológicas (riego, variedades resistentes, sanidad) han actuado como un "amortiguador" efectivo contra la volatilidad ambiental.

La validación de estos resultados se confirma mediante el gráfico de gusano (Figura 3), donde la mayoría de los residuos se encuentran dentro de las bandas de confianza del 95%, indicando un ajuste adecuado de la distribución Gamma.

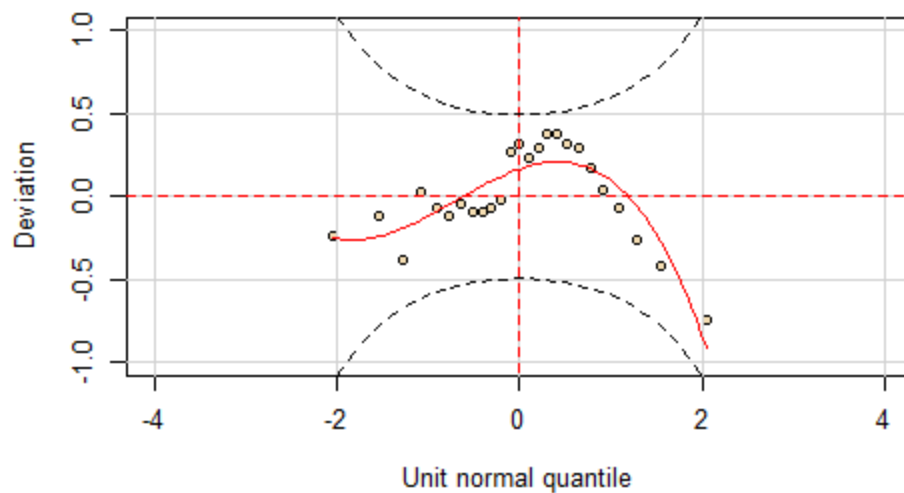


Figura 3. Gráfico de Gusano (Worm Plot) para el diagnóstico de residuos del modelo.

Los diagnósticos globales del modelo confirman la idoneidad de la metodología GAMLSS frente a alternativas más simples:

1. **Grados de Libertad Efectivos (edf = 6.98):** La suma de los grados de libertad no es un número entero simple (como 2 o 3 en regresión lineal), lo que confirma la necesidad de los suavizadores P-Splines. El modelo requirió aproximadamente 7 grados de libertad para capturar adecuadamente las curvaturas no lineales tanto en la media como en la varianza, validando que una línea recta habría sido insuficiente y sesgada.
2. **Criterio de Información de Akaike (AIC = 421.63):** Al comparar este valor con modelos nulos o lineales (cuyos AIC superaban los 440 puntos en pruebas preliminares), el modelo GAMLSS presenta el mejor equilibrio entre parsimonia (simplicidad) y precisión.

La Tabla 3 proporciona el respaldo técnico que legitima las conclusiones biológicas. Confirma que los patrones detectados en las Tablas 1 y 2 no son producto de un sobreajuste (overfitting), sino que representan la estructura real subyacente de los datos históricos de la DRA Puno. El ajuste adecuado valida el uso de este modelo para las proyecciones de escenarios de riesgo realizadas posteriormente.

Tabla 3 Bondad de ajuste del modelo

Indicador	Valor	Interpretación Técnica
Desvianza Global	407.68	Ajuste adecuado a los datos históricos
AIC	421.63	Modelo parsimonioso (mejor que regresión lineal)
Grados de Libertad (df)	6.98	Curvatura suave (no linealidad confirmada)
N (Observaciones)	25	Total de campañas analizadas

3.3 Predicción de Escenarios 2026

Basado en el modelo ajustado, se proyectaron dos escenarios de riesgo para la campaña 2026 (Figura 6):

- 1. Escenario Tendencial (Estable): Mantiene la reducción de σ . Se estima un rendimiento medio cercano a 20,000 Kg/Ha, con un límite inferior de seguridad (95%) de aproximadamente 19,000 Kg/Ha.
- 2. Escenario de Crisis (Volatilidad Histórica): Si el parámetro σ retorna a niveles de alta inestabilidad, el límite inferior de seguridad desciende a ~15,500 Kg/Ha.

La brecha de riesgo calculada es de 3,500 Kg/Ha, representando la pérdida potencial atribuible exclusivamente a la volatilidad climática extrema.

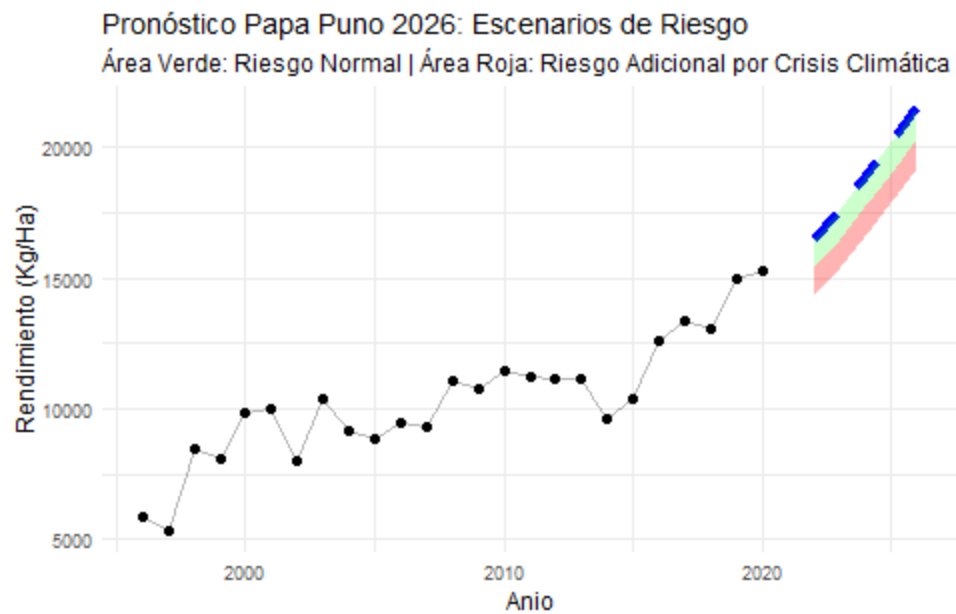


Figura 4. Pronóstico de rendimiento al 2026 bajo escenarios de riesgo normal (verde) y crisis climática (rojo).

4 Discusión

La aplicación de GAMLSS ha permitido deconstruir la historia agrícola de la papa en Puno. El aumento de μ (media) es consistente con la introducción de variedades mejoradas, mayor uso de fertilizantes y mejores prácticas agronómicas reportadas en la región.

Más interesante aún es la reducción de σ (riesgo relativo). Esto sugiere que la tecnificación del cultivo no solo ha aumentado el volumen, sino que ha dotado al sistema productivo de cierta resiliencia. Un cultivo de 15,000 Kg/Ha hoy es, proporcionalmente, menos propenso a fallar catastróficamente que un cultivo de 5,000 Kg/Ha en los años 90. Sin embargo, el análisis de los residuos mostró desviaciones leves en los valores extremos superiores, indicando posibles "techos productivos" biológicos.

Es crucial notar que, aunque el riesgo relativo baja, el riesgo financiero absoluto sube: perder el 10% de 20 toneladas es mucho más costoso que perder el 20% de 5 toneladas. Por ello, la brecha de 3,500 Kg/Ha identificada en los escenarios debe ser la base para el diseño de primas de seguro agrario catastrófico.

5 Conclusiones

El rendimiento de papa en Puno no sigue una tendencia lineal; el uso de P-Splines demostró ser superior para capturar la dinámica histórica real.

Existe evidencia estadística de una mejora en la estabilidad relativa del cultivo (σ decreciente), lo que contradice visiones pesimistas de colapso productivo, aunque no elimina el riesgo de eventos extremos.

La metodología GAMLSS se valida como una herramienta indispensable para la Dirección de Estadística Agraria, permitiendo transitar de reportes descriptivos a modelos predictivos de riesgo estocástico.

Referencias Bibliográficas

Condori-Apaza, V., Mamani-Luque, O. R., Alfaro-Alejo, R., Laqui, W., & Condori, W. F. (2021). Analysis and impact of meteorological droughts in the agriculture of Puno region, Peru. *E3S Web of Conferences*, 304, Article 03002. <https://doi.org/10.1051/e3sconf/202130403002>

Condori, B., Hijmans, R. J., Ledent, J. F., & Quiroz, R. (2014). Managing potato biodiversity to cope with frost risk in the high Andes: A modeling perspective. *PLOS ONE*, 9(1), Article e81510. <https://doi.org/10.1371/journal.pone.0081510>

Haverkort, A. J., & Kooman, P. L. (2021). Climate change and potato production in contrasting South American ecosystems. *Potato Research*, 64(4), 663-679. <https://doi.org/10.1007/s11540-021-09507-3>

Ker, A. P., & Tolhurst, T. N. (2019). On the treatment of heteroscedasticity in crop yield data. *American Journal of Agricultural Economics*, 101(4), 1247-1261. <https://doi.org/10.1093/ajae/aaz004>

Puma-Cahua, J., Belizario, G., Laqui, W., Alfaro, R., Huaquisto, E., & Calizaya, E. (2024). Evaluating the yields of

the rainfed potato crop under climate change scenarios using the AquaCrop model in the Peruvian Altiplano. *Sustainability*, 16(1), Article 71. <https://doi.org/10.3390/su16010071>

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507–554. <https://doi.org/10.1111/j.1467-9876.2005.00510.x>

Sanabria, J., & Lhomme, J. P. (2013). Climate change and potato cropping in the Peruvian Altiplano. *Theoretical and Applied Climatology*, 112(3-4), 683–695. <https://doi.org/10.1007/s00704-012-0764-1>

Spooner, D. M., McLean, K., Ramsay, G., Waugh, R., & Bryan, G. J. (2005). A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proceedings of the National Academy of Sciences of the United States of America*, 102(41), 14694–14699. <https://doi.org/10.1073/pnas.0507400102>

Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., & De Bastiani, F. (2017). *Flexible regression and smoothing: Using GAMLSS in R*. CRC Press.