

Artículo Científico

Marco integrado de estadística y aprendizaje automático para la optimización termo-mecánica de masas compuestas de amaranto y chocho

An Integrated Statistical and Machine Learning Framework for Thermo-Mechanical Optimization of Amaranth–Lupin Composite Doughs



Rojas-Porras, Santiago Alexander¹



<https://orcid.org/0009-0004-2789-1437>



santiago.rojas@upec.edu.ec



Universidad Politécnica Estatal del Carchi, Ecuador, Tulcán.



Hidalgo-Rodríguez, Johnny Fernando²



<https://orcid.org/0000-0001-8436-7843>



johnny.hidalgo@upec.edu.ec



Universidad Politécnica Estatal del Carchi, Ecuador, Tulcán..



Burbano-Pulles, Marco Rubén³



<https://orcid.org/0000-0002-2036-7032>



marco.burbano@upec.edu.ec



Universidad Politécnica Estatal del Carchi, Ecuador, Tulcán.

Autor de correspondencia¹



DOI / URL: <https://doi.org/10.55813/gaea/rcym/v4/n2/209>

Resumen: La optimización de formulaciones farináceas compuestas requiere enfoques analíticos capaces de integrar múltiples atributos reológicos interdependientes. El objetivo de este estudio fue evaluar el comportamiento termo-mecánico de masas formuladas mediante la sustitución progresiva de harina de amaranto por harina de chocho (0–40%). Los parámetros reológicos se determinaron mediante Mixolab, considerando consistencia, debilitamiento proteico, gelatinización, estabilidad térmica y retrogradación. El análisis se desarrolló mediante un enfoque integrado que combinó estadística univariada, análisis multivariado y aprendizaje automático. Los resultados evidenciaron diferencias significativas entre tratamientos, con efectos de gran magnitud en la consistencia inicial y la retrogradación, lo que indica una fuerte influencia de la formulación sobre el comportamiento estructural de la masa. El análisis multivariado permitió identificar patrones diferenciados y explicar una alta proporción de la variabilidad total del sistema, mientras que el aprendizaje automático confirmó la discriminación entre formulaciones y la estabilidad del ranking. En conjunto, la integración metodológica mejora la confiabilidad en la toma de decisiones y constituye una estrategia robusta para la optimización de masas compuestas en sistemas alimentarios complejos.

Palabras clave: mixolab; reología; amaranto; chocho; aprendizaje automático



Check for updates

Received: 14/Abr/2026
Accepted: 08/May/2026
Published: 03/Jun/2026

Cita: Rojas-Porras, S. A., Hidalgo-Rodríguez, J. F., & Burbano-Pulles, M. R. (2026). Marco integrado de estadística y aprendizaje automático para la optimización termo-mecánica de masas compuestas de amaranto y chocho. *Revista Científica Ciencia y Método*, 4(2), 513-532. <https://doi.org/10.55813/gaea/rcym/v4/n2/209>

Revista Científica Ciencia y Método (RCyM)
<https://revistacym.com>
revistacym@editorialgrupo-aea.com
info@editoriagrupo-aea.com

© 2026. Este artículo es un documento de acceso abierto distribuido bajo los términos y condiciones de la **Licencia Creative Commons. Atribución-NoComercial 4.0 Internacional.**



Abstract:

The optimization of composite flour formulations requires analytical approaches capable of integrating multiple interdependent rheological attributes. This study aimed to evaluate the thermo-mechanical behavior of doughs formulated through the progressive substitution of amaranth flour with lupin flour (0–40%). Rheological parameters were determined using Mixolab, considering consistency, protein weakening, gelatinization, thermal stability, and retrogradation. The analysis was conducted using an integrated approach combining univariate statistics, multivariate analysis, and machine learning. The results revealed significant differences among treatments, with large effect sizes in initial consistency and retrogradation, indicating a strong influence of formulation on the structural behavior of the dough. Multivariate analysis identified distinct patterns and explained a high proportion of the total system variability, while machine learning confirmed discrimination between formulations and the stability of the performance ranking. Overall, the integrated approach enhances decision-making reliability and provides a robust strategy for optimizing composite doughs in complex food systems.

Keywords: mixolab; rheology; amaranth; lupin; machine learning.

1. Introducción

La optimización de formulaciones farináceas compuestas constituye un desafío metodológico cuando múltiples atributos reológicos interactúan de forma simultánea y no lineal. En matrices basadas en pseudocereales y leguminosas, pequeñas variaciones en la proporción de ingredientes pueden inducir reorganizaciones complejas en la red proteína–almidón, generando comportamientos emergentes que no son capturados adecuadamente mediante enfoques univariados. En este contexto, la toma de decisiones basada en comparaciones aisladas puede omitir interacciones multidimensionales críticas para el desempeño tecnológico (Sheng et al., 2025; Chuqui-Paulino et al., 2025; Bansal et al., 2025).

El amaranto (*Amaranthus caudatus* L.) y el chocho (*Lupinus mutabilis* Sweet) son materias primas de alto valor nutricional y creciente relevancia tecnológica. Su incorporación mejora el perfil proteico y funcional, pero también modifica la estructura de la masa y su respuesta termo-mecánica durante el procesamiento, afectando la hidratación, el desarrollo viscoelástico, la gelatinización y la retrogradación (Meneses & Pulles, 2025; Meng et al., 2021).

El analizador Mixolab permite evaluar de forma integrada estos fenómenos mediante parámetros como C1–C5 y las pendientes α , β y γ (Dubat, 2010; Codinã et al., 2012; Hadnađev et al., 2011; Dvořáček et al., 2019). Sin embargo, la interdependencia entre variables genera estructuras de datos correlacionadas que limitan la interpretación

basada únicamente en pruebas de hipótesis. En este sentido, la integración de metodologías multivariadas y aprendizaje automático ha demostrado mejorar la discriminación y optimización en sistemas alimentarios complejos (Granato et al., 2018; Araújo et al., 2023; Jolliffe & Cadima, 2016), siendo además relevante en contextos de innovación y sostenibilidad (Galarza-Sánchez et al., 2023).

Adicionalmente, en diseños con tamaños muestrales reducidos, la variabilidad intra-tratamiento puede afectar la identificación de la formulación óptima. Por ello, la selección de la repetición más representativa y la evaluación de la estabilidad del ranking mediante remuestreo son fundamentales para mejorar la confiabilidad inferencial (Adin et al., 2024; Efron & Tibshirani, 1993; Mougán & Nielsen, 2023).

En este contexto, el objetivo del presente estudio fue determinar el tratamiento óptimo y la repetición más representativa en formulaciones compuestas de amaranto y chocho a partir de datos de Mixolab, comparando enfoques de estadística clásica, análisis multivariado y aprendizaje automático, con el fin de proponer un marco metodológico robusto para la toma de decisiones en sistemas alimentarios complejos.

2. Materiales y métodos

Materias primas y preparación de harinas

Los granos de amaranto (*Amaranthus caudatus* L., var. Inti) fueron adquiridos en Imbabura (Ecuador), mientras que las semillas de chocho andino (*Lupinus mutabilis* Sweet, var. INIAP-450) se obtuvieron en el INIAP (Quito, Ecuador). Esta variedad corresponde a un genotipo mejorado ampliamente cultivado en condiciones andinas.

Las semillas fueron limpiadas, secadas a 50 °C hasta <12% de humedad (base húmeda) y molidas en molino de martillos, con tamizado $\leq 250 \mu\text{m}$. Las harinas se almacenaron a 4 °C en recipientes herméticos y se equilibraron a condiciones de laboratorio antes del análisis.

Formulación de tratamientos

Se prepararon cinco mezclas mediante sustitución progresiva de amaranto por chocho (0–40%) en base gravimétrica (T0–T4). Los incrementos del 10% permiten evaluar transiciones funcionales sin alterar abruptamente la estructura (Benkadri et al., 2021; Li et al., 2023; Duodu & Minnaar, 2011). Cada tratamiento se evaluó en triplicado ($n = 3$) bajo un diseño completamente al azar, estándar en estudios reológicos y adecuado para estimar la variabilidad experimental (Rigdon et al., 2022).

Caracterización reológica mediante Mixolab

Las propiedades termo-mecánicas se evaluaron con Mixolab 2 (Chopin Technologies), bajo protocolo Chopin+ (ICC 173) (Dubat, 2010; Chopin Technologies, 2012). Se utilizaron 50 g de harina (14% humedad), ajustando la hidratación a $C1 = 1.1 \pm 0.05$ N·m. El perfil térmico incluyó: 30 °C (8 min), calentamiento a 90 °C (4 °C/min),

mantenimiento (7 min) y enfriamiento a 50 °C. La velocidad fue de 80 rpm y el tiempo total de 45 min. Se obtuvieron C1–C5, estabilidad y pendientes α , β y γ , asociadas a procesos de debilitamiento proteico, gelatinización y retrogradación (Hadnađev et al., 2011; Codinã et al., 2012). Las mediciones se realizaron a 22 ± 1 °C y $50 \pm 5\%$ HR.

Análisis inferencial clásico

El efecto del tratamiento sobre cada parámetro reológico fue evaluado mediante análisis de varianza (ANOVA) de una vía. La normalidad de los residuos se verificó mediante la prueba de Shapiro–Wilk, mientras que la homogeneidad de varianzas se evaluó con la prueba de Levene. En casos de heterocedasticidad, se aplicó el ANOVA de Welch para garantizar inferencias robustas. Cuando se detectaron diferencias significativas ($p < 0.05$), se realizaron comparaciones múltiples mediante Tukey, aplicando corrección por tasa de descubrimientos falsos (Benjamini et al., 2001). El tamaño del efecto se estimó mediante eta cuadrado (η^2) y omega cuadrada (ω^2). Adicionalmente, se realizó un MANOVA empleando la traza de Pillai, por su robustez frente a desviaciones de normalidad multivariada.

Análisis multivariado robusto

Dada la naturaleza correlacionada de los datos, se implementó un enfoque multivariado complementario. Las variables fueron previamente estandarizadas mediante normalización tipo z. Las diferencias globales entre tratamientos se evaluaron mediante PERMANOVA basada en distancias euclidianas y 9,999 permutaciones. La homogeneidad de dispersión se verificó con PERMDISP. La reducción de dimensionalidad se realizó mediante análisis de componentes principales (PCA), reteniendo componentes con valores propios mayores a uno y varianza acumulada superior al 70% (Jolliffe & Cadima, 2016; Granato et al., 2018).

Selección de la repetición representativa

La repetición más representativa se seleccionó como aquella con menor distancia euclidiana al centroide del tratamiento en el espacio estandarizado, minimizando el efecto de valores atípicos.

Construcción de índices de decisión

Se desarrollaron dos índices de decisión: (i) un índice clásico basado en parámetros estandarizados y coeficiente de variación, y (ii) un índice multivariado basado en el espacio de componentes principales. Ambos permitieron establecer rankings de desempeño y comparar criterios.

Evaluación de estabilidad del ranking

Se aplicó bootstrap estratificado (3,000 iteraciones) para evaluar la robustez del ranking, estimando la incertidumbre sin asumir distribuciones paramétricas (Mougan & Nielsen, 2023; Adin et al., 2024).

Software estadístico

Los análisis se realizaron en R, utilizando paquetes especializados para análisis inferencial, multivariado y visualización, garantizando trazabilidad y reproducibilidad.

Análisis supervisado mediante Random Forest

Se implementó un modelo Random Forest con validación cruzada leave-one-out (LOOCV), utilizando parámetros reológicos como predictores. La importancia de variables se evaluó mediante disminución de precisión (Breiman, 2001).

3. Resultados

3.1. Perfil reológico descriptivo y variabilidad intra-tratamiento

Los estadísticos descriptivos de los parámetros reológicos obtenidos mediante Mixolab (C1–C5, estabilidad) y los coeficientes de pendiente (α , β , γ) se presentan en la Tabla 1, expresados como media \pm desviación estándar y coeficiente de variación (CV%).

Tabla 1

Media \pm desviación estándar y coeficiente de variación (CV%) de los parámetros reológicos por tratamiento

Tratamiento	Parámetro	Media \pm DE	CV (%)
T0	α	-0.0713 \pm 0.0064	9.01
	β	0.0513 \pm 0.0172	33.59
	C1	1.0850 \pm 0.0087	0.80
	C2	0.5837 \pm 0.0222	3.80
	C3	0.7300 \pm 0.0020	0.27
	C4	0.7017 \pm 0.0085	1.21
	C5	1.1853 \pm 0.0099	0.83
	Estabilidad	9.0000 \pm 0.1000	1.11
	γ	-0.0067 \pm 0.0050	75.50
	T1	α	-0.0667 \pm 0.0450
β		0.0340 \pm 0.0053	15.56
C1		1.2533 \pm 0.0147	1.18
C2		0.6810 \pm 0.0069	1.02
C3		0.7590 \pm 0.0090	1.19
C4		0.7057 \pm 0.0101	1.43
C5		1.1723 \pm 0.0093	0.79
Estabilidad		9.2667 \pm 0.4509	4.87
γ		-0.0313 \pm 0.0441	140.76
T2		α	0.0320 \pm 0.0104
	β	-0.0053 \pm 0.0099	184.98
	C1	1.0047 \pm 0.0206	2.06
	C2	0.5650 \pm 0.0095	1.69
	C3	0.6700 \pm 0.0115	1.72
	C4	0.5813 \pm 0.0105	1.81
	C5	0.8680 \pm 0.0066	0.76
	Estabilidad	8.5000 \pm 0.8718	10.26
	γ	-0.0120 \pm 0.0069	57.74
	T3	α	0.0220 \pm 0.0000
β		0.6013 \pm 0.1619	26.92
C1		0.7323 \pm 0.0555	7.58

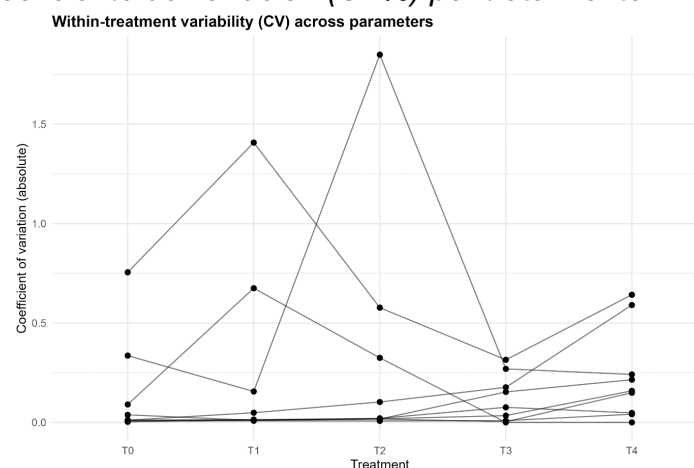
	C2	0.5417 ± 0.0830	15.32
	C3	0.8540 ± 0.0026	0.31
	C4	0.7097 ± 0.0242	3.41
	C5	0.6977 ± 0.0061	0.88
	Estabilidad	1.5000 ± 0.2646	17.64
	γ	-0.0220 ± 0.0069	31.49
	α	0.1260 ± 0.0000	0.00
	β	0.3960 ± 0.0955	24.12
	C1	0.7950 ± 0.0380	4.78
T4	C2	0.4987 ± 0.1071	21.48
	C3	0.7227 ± 0.1077	14.91
	C4	0.6210 ± 0.0985	15.87
	C5	0.5677 ± 0.0234	4.11
	Estabilidad	3.4000 ± 2.0075	59.04
		γ	-0.1480 ± 0.0950

Nota: Media ± DE: media ± desviación estándar; CV: coeficiente de variación; C1–C5: parámetros Mixolab; α, β, γ: pendientes reológicas. (Autores, 2026).

Se observó una modificación progresiva del perfil termo-mecánico con el incremento de la proporción de harina de chocho. La consistencia inicial (C1) disminuyó en los niveles altos de sustitución, pasando de 1.085 N·m en T0 a 0.732 N·m en T3 y 0.795 N·m en T4. De forma similar, C5 se redujo de 1.185 N·m en T0 a 0.568 N·m en T4, mientras que la estabilidad descendió marcadamente a partir del 30% de sustitución, de 9.00 min en T0 a 1.50 min en T3 y 3.40 min en T4.

En contraste, C3 alcanzó su valor máximo en T3 (0.854 N·m), superior al control y a las formulaciones intermedias, lo que sugiere una mayor contribución del almidón a la respuesta térmica en niveles altos de sustitución. Los coeficientes β y γ también mostraron alta variabilidad entre tratamientos, reflejando cambios en las tasas de gelatinización y retrogradación.

Figura 1
Distribución del coeficiente de variación (CV%) por tratamiento



Nota: Coeficiente de variación (CV) de los parámetros reológicos por tratamiento (T0–T4) (Autores, 2026).

Se observó varianza nula en α para T3 y T4, correspondiente a mediciones idénticas entre repeticiones. En conjunto, estos resultados muestran una modificación estructural progresiva y un incremento de la heterogeneidad reológica a niveles de sustitución superiores al 30%.

3.2. Análisis univariado y tamaño del efecto

Los resultados del análisis inferencial univariado, incluyendo la verificación de supuestos (pruebas de Shapiro–Wilk y Levene), el método estadístico aplicado (ANOVA) y las estimaciones del tamaño del efecto (η^2 y ω^2), se presentan en la Tabla 2.

Tabla 2

Resultados del análisis univariado, pruebas de supuestos y tamaños del efecto para cada parámetro reológico

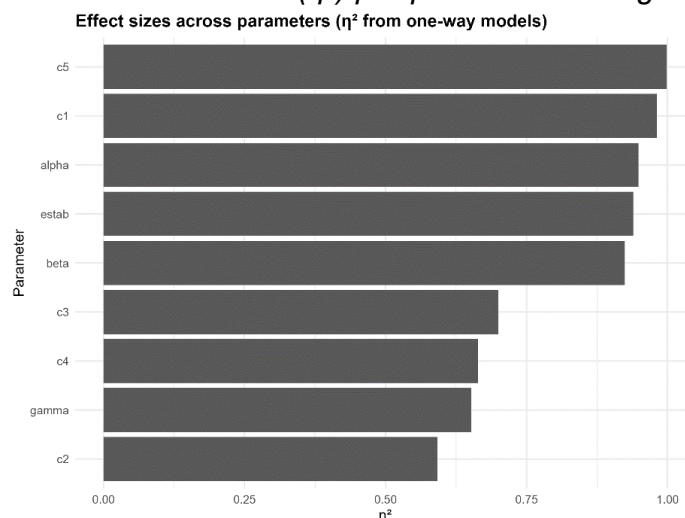
Parámetro	Método	Shapiro–Wilk (p)	Levene (p)	F	p-valor	η^2	ω^2	p (FDR)
C5	ANOVA + Tukey	0.493	0.514	1426.97	9.80×10^{-14}	0.998	0.997	8.82×10^{-13}
C1	ANOVA + Tukey	0.670	0.407	129.98	1.41×10^{-8}	0.981	0.972	6.36×10^{-8}
α	ANOVA + Tukey	0.003	0.314	45.81	2.13×10^{-6}	0.948	0.923	6.39×10^{-6}
Estabilidad	ANOVA + Tukey	0.068	0.199	38.71	4.68×10^{-6}	0.939	0.910	1.05×10^{-5}
β	ANOVA + Tukey	0.028	0.215	30.43	1.42×10^{-5}	0.924	0.887	2.55×10^{-5}
C3	ANOVA + Tukey	3.81×10^{-4}	0.257	5.83	0.0109	0.700	0.563	0.0164
C4	ANOVA + Tukey	0.0075	0.426	4.93	0.0186	0.664	0.512	0.0239
γ	ANOVA + Tukey	0.0147	0.140	4.69	0.0216	0.652	0.496	0.0244
C2	ANOVA + Tukey	0.0399	0.139	3.62	0.0449	0.592	0.412	0.0449

Nota: η^2 : eta cuadrada; ω^2 : omega cuadrada; FDR: tasa de descubrimientos falsos (Autores, 2026).

Tras la corrección por tasa de descubrimientos falsos (FDR), todos los parámetros mostraron diferencias estadísticamente significativas entre tratamientos (p ajustado < 0.05), evidenciando un efecto consistente del nivel de sustitución sobre el perfil termo-mecánico.

Los mayores tamaños del efecto se observaron en C5 ($\eta^2 = 0.998$; $\omega^2 = 0.997$) y C1 ($\eta^2 = 0.981$; $\omega^2 = 0.972$), seguidos por estabilidad y α ($\eta^2 > 0.90$), indicando que la formulación explica la mayor parte de la variabilidad en estos parámetros. En contraste, C3 ($\eta^2 = 0.700$), C4 ($\eta^2 = 0.664$) y γ ($\eta^2 = 0.652$) presentaron efectos intermedios, mientras que C2 mostró la menor sensibilidad ($\eta^2 = 0.592$).

Figura 2
Distribución de los tamaños del efecto (η^2) por parámetro reológico



Nota: Tamaños del efecto (η^2) de los parámetros reológicos. Los valores reflejan la magnitud de la influencia del tratamiento (Autores, 2026).

En conjunto, el análisis univariado confirma efectos significativos del tratamiento y permite cuantificar su magnitud; no obstante, debido a la correlación entre variables, se justifica el uso de enfoques multivariados para evaluar la diferenciación estructural del sistema.

3.3. Efecto multivariado global: MANOVA y PERMANOVA

Para evaluar el efecto integrado del tratamiento sobre el perfil reológico conjunto (C1–C5, estabilidad, α , β y γ), se aplicaron enfoques multivariados complementarios.

El MANOVA evidenció un efecto global significativo del tratamiento ($F = 11.55$; $p < 0.001$; η^2 parcial = 0.822), indicando que el 82.2% de la variabilidad multivariada total del sistema se explica por diferencias entre tratamientos.

Tabla 3
Resultados del análisis multivariado de varianza (MANOVA)

Fuente	Df	Suma de Cuadrados	R ²	F	p-valor
Modelo	4	103.5821	0.8221	11.5512	0.0001
Residuo	10	22.4179	0.1779	—	—
Total	14	126.0000	1.0000	—	—

Nota: R²: proporción de varianza explicada; η^2 parcial: tamaño del efecto multivariado. (Autores, 2026).

De forma complementaria, el PERMANOVA basado en distancias euclidianas estandarizadas (9,999 permutaciones) también detectó diferencias significativas ($F = 4.87$; $p = 0.0186$), explicando el 66.1% de la variabilidad total.

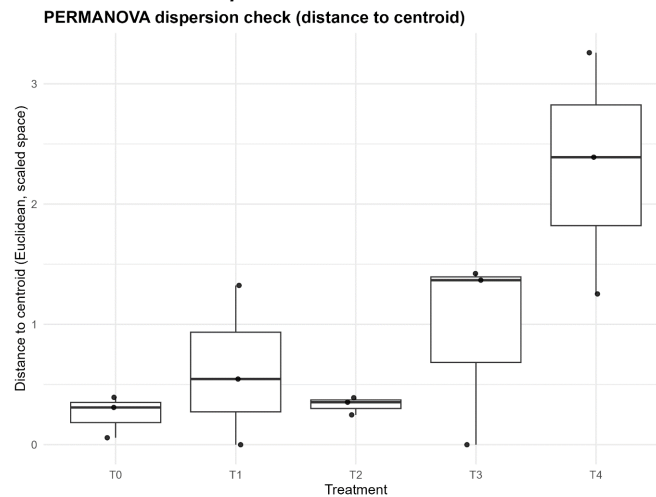
Tabla 4
Resultados del análisis de varianza multivariado permutacional (PERMANOVA)

Fuente	Df	Suma de cuadrados	Cuadrado medio	F	Nº de permutaciones	p-valor
Tratamiento	4	8.3397	2.0849	4.8726	9,999	0.0186
Residuo	10	4.2789	0.4279	—	—	—
Total	14	12.6186	—	—	—	—

Nota: PERMANOVA: análisis de varianza multivariado permutacional (Autores, 2026).

La prueba de homogeneidad de dispersión (betadisper) no fue significativa ($p > 0.05$), lo que indica que la diferenciación entre tratamientos se debe a cambios en la posición de los centroides y no a variaciones en la dispersión interna.

Figura 3
Distancia al centroide multivariado por tratamiento



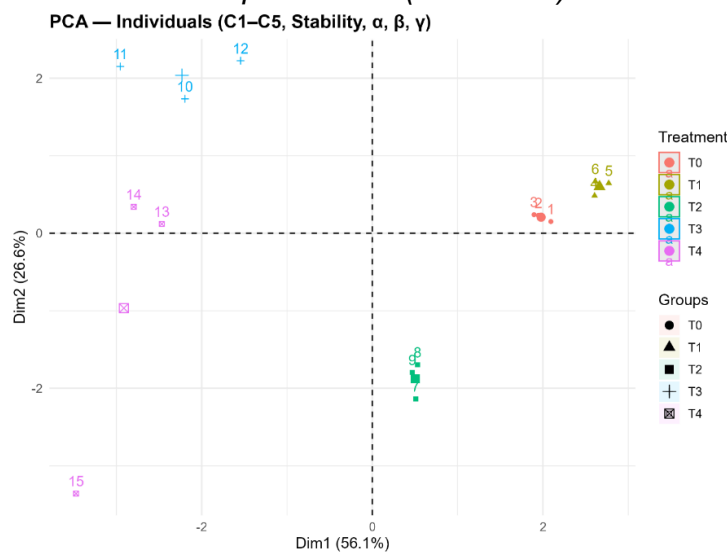
Nota: Distancia al centroide por tratamiento en el espacio multivariado (PERMDISP) (Autores, 2026).

En conjunto, ambos enfoques evidencian una diferenciación global significativa entre tratamientos, confirmando configuraciones estructurales distintas en el espacio reológico.

a. Análisis de componentes principales (PCA)

El PCA integró los nueve parámetros reológicos en un espacio de menor dimensionalidad. Los dos primeros componentes explicaron el 82.7% de la varianza total (PC1 = 56.1%; PC2 = 26.6%), indicando una adecuada representación bidimensional del sistema.

Figura 4
Proyección de tratamientos en el plano PCA (PC1–PC2)



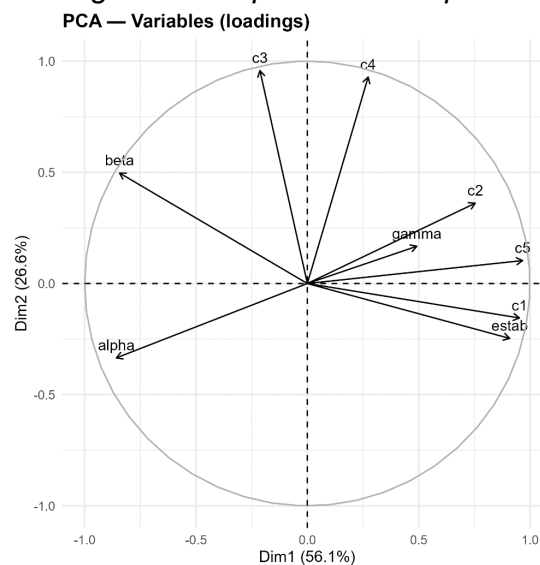
Nota: Proyección de las observaciones por tratamiento en el plano de componentes principales (PCA). (Autores, 2026).

El componente PC1 estuvo principalmente asociado con C1, C5, estabilidad y C2, mientras que PC2 estuvo influenciado por C3 y C4, reflejando la contribución de procesos térmicos del almidón.

En el plano PCA, T0 y T1 se ubicaron en la región positiva de PC1, mientras que T4 se desplazó hacia la región negativa, evidenciando una diferenciación estructural marcada. El tratamiento T3 mostró una separación relevante a lo largo de PC2, asociada a sus mayores valores de C3.

Figura 5

Cargas de las variables reológicas en el plano de componentes principales.



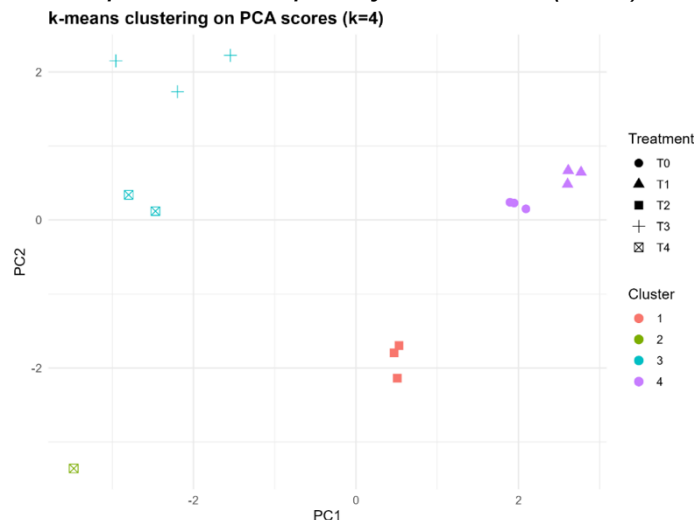
Nota: Cargas de variables en PCA. Dim1 (56.1%) y Dim2 (26.6%) corresponden a la varianza explicada (Autores, 2026).

Estos resultados confirman patrones estructurales diferenciados consistentes con MANOVA y PERMANOVA.

3.4. Agrupamiento no supervisado en el espacio PCA

Se aplicó un análisis de clustering no supervisado mediante k-means sobre los puntajes de los dos primeros componentes principales (PC1–PC2). El número óptimo de grupos ($k = 4$) se determinó mediante el criterio de silueta, basado en la cohesión interna y la separación entre clusters.

Figura 6
Agrupamiento *k*-means aplicado a los puntajes del PCA (*k* = 4)



Nota: Agrupamiento *k*-means aplicado a los puntajes del PCA (*k* = 4) (Autores, 2026).

T0 y T1 se agruparon en un mismo cluster, indicando alta proximidad estructural, mientras que T2 formó un grupo independiente. T3 se diferenció en PC2 y T4 ocupó una región separada en PC1.

La concordancia entre PCA y *k*-means confirma la estabilidad de la estructura multivariada.

3.5. Discriminación supervisada de tratamientos mediante Random Forest

Se evaluó la separabilidad entre tratamientos mediante un modelo Random Forest con validación cruzada leave-one-out (LOOCV). El modelo alcanzó una precisión global del 93.3% (IC 95%: 68.1–99.8%) y un coeficiente kappa de 0.917, indicando una concordancia casi perfecta. La precisión fue significativamente superior a la tasa de no información (NIR = 0.20; *p* < 0.001), evidenciando alta capacidad discriminativa del perfil reológico.

Tabla 5
Desempeño del modelo Random Forest bajo validación cruzada LOOCV

Predicción \ Referencia	T0	T1	T2	T3	T4
T0	3	0	0	0	0
T1	0	3	0	0	0
T2	0	0	3	0	0
T3	0	0	0	3	1
T4	0	0	0	0	2

Nota: T0: Tratamiento 0; T1: Tratamiento 1; T2: Tratamiento 2; T3: Tratamiento 3; T4: Tratamiento 4 (Autores, 2026).

Tabla 6
Desempeño global del modelo

Métrica	Valor
Exactitud	0.933
IC 95%	0.681–0.998
Índice Kappa	0.917
Tasa de no información (NIR)	0.20
p (Exactitud > NIR)	< 0.001

Nota: NIR: tasa de no información; IC: intervalo de confianza (Autores, 2026).

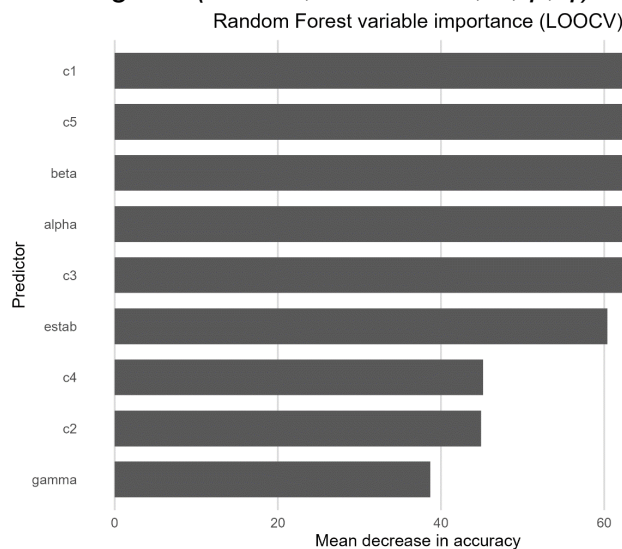
Tabla 7
Métricas de desempeño por clase

Métrica	T0	T1	T2	T3	T4
Sensibilidad	1.00	1.00	1.00	1.00	0.667
Especificidad	1.00	1.00	1.00	0.917	1.00
Valor Predictivo Positivo	1.00	1.00	1.00	0.75	1.00
Valor Predictivo Negativo	1.00	1.00	1.00	1.00	0.923
Exactitud Balanceada	1.00	1.00	1.00	0.958	0.833

Nota: Métricas calculadas mediante validación LOOCV (Autores, 2026).

La matriz de confusión mostró clasificación perfecta para T0, T1 y T2 (sensibilidad = 1.00), con una única clasificación errónea entre T3 y T4. En particular, una muestra de T4 fue clasificada como T3, resultando en una precisión balanceada de 0.833 para T4. No obstante, el resto de métricas se mantuvieron elevadas, evidenciando un desempeño robusto del modelo.

Figura 7
Importancia de variables del modelo Random Forest bajo validación LOOCV, basada en los parámetros reológicos (C1–C5, estabilidad, α , β , γ)



Nota: Importancia de variables en el modelo Random Forest (LOOCV), basada en la disminución de la precisión (Autores, 2026).

El análisis de importancia identificó a C1 y C5 como los predictores más influyentes, seguidos por β y α , mientras que γ y C2 mostraron menor contribución. Este patrón es consistente con el análisis univariado, reforzando la relevancia estructural de estos parámetros.

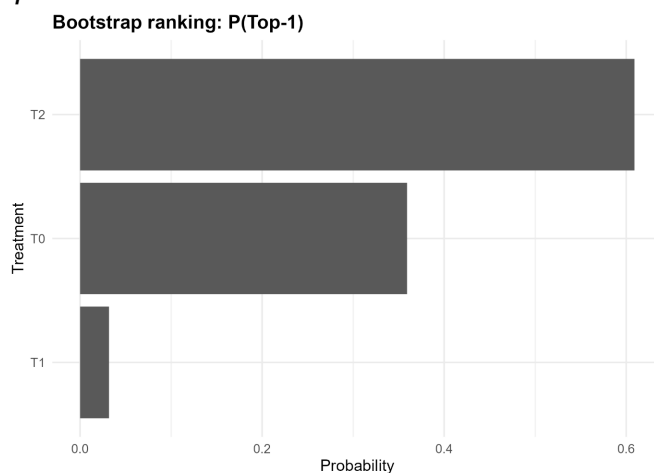
En conjunto, los resultados confirman una diferenciación estructural consistente en el espacio reológico multidimensional.

3.6. Evaluación de la estabilidad del ranking mediante bootstrap

Se aplicó un procedimiento de remuestreo bootstrap estratificado (3,000 iteraciones) sobre el índice de decisión compuesto, integrando la proximidad al perfil tecnológico objetivo y la variabilidad intra-tratamiento.

Figura 8

Probabilidad de ocupar la primera posición del ranking ($P[\text{Top-1}]$) obtenida mediante remuestreo bootstrap



Nota: Probabilidad de ocupar la primera posición del ranking ($P[\text{Top-1}]$) estimada mediante bootstrap (Autores, 2026).

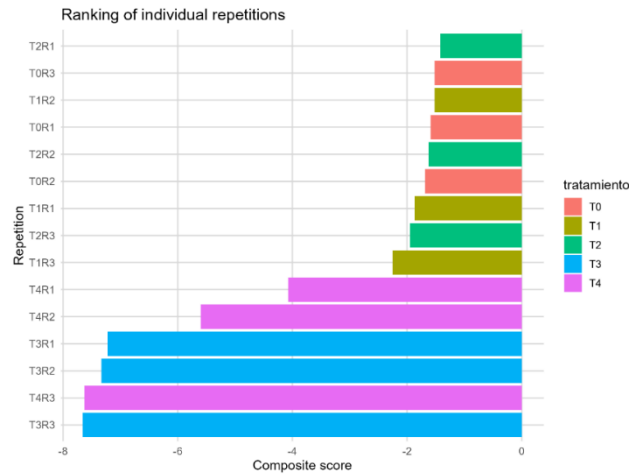
El tratamiento T2 presentó la mayor probabilidad de ocupar la primera posición (≈ 0.61), seguido por T0 (≈ 0.35), mientras que T1 mostró una contribución marginal. Los tratamientos T3 y T4 no evidenciaron presencia relevante como formulaciones óptimas.

La distribución observada indica estabilidad del ranking frente a perturbaciones de muestreo, destacando el desempeño consistente de T2.

3.7. Selección de la mejor repetición individual mediante el índice compuesto

El índice compuesto se calculó a nivel de repetición individual con el fin de identificar tanto el tratamiento óptimo como la unidad experimental de mayor desempeño estructural, integrando proximidad al perfil tecnológico objetivo y variabilidad intra-observación.

Figura 9
Ranking de repeticiones individuales basado en el índice compuesto



Nota: Ranking de repeticiones individuales basado en el índice compuesto. Valores más altos indican mejor desempeño estructural (Autores, 2026).

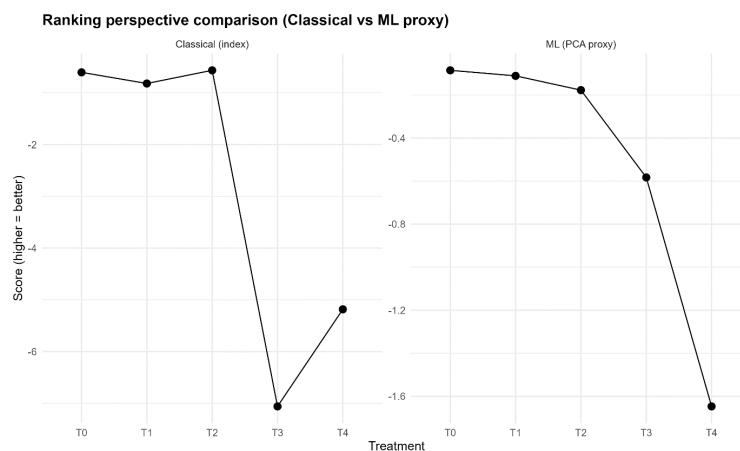
La repetición T2R1 alcanzó el mayor puntaje, seguida por T0R3 y T1R2, mientras que las repeticiones de T3 y T4 se ubicaron en posiciones inferiores, evidenciando menor desempeño estructural.

El ranking individual fue consistente con el análisis de estabilidad por bootstrap (Figura 8), donde T2 presentó la mayor probabilidad de ocupar la primera posición, lo que refuerza la robustez del proceso de selección.

3.8. Comparación entre el enfoque clásico y el enfoque multivariado

El ranking de tratamientos obtenido a partir del índice compuesto clásico fue contrastado con el ranking derivado de la proyección estructural multivariada en el espacio de componentes principales (PCA).

Figura 10
Comparación del ranking de tratamientos bajo el enfoque clásico y el enfoque multivariado.



Nota: Comparación del ranking de tratamientos bajo el enfoque clásico y el enfoque multivariado (PCA). (Autores, 2026).

Ambos enfoques identificaron al tratamiento T2 entre las formulaciones mejor posicionadas, evidenciando convergencia en la selección del tratamiento óptimo. Sin embargo, se observaron diferencias en la posición relativa de T3 y T4; en particular, el enfoque multivariado mostró una mayor diferenciación estructural para T4 en comparación con el índice clásico.

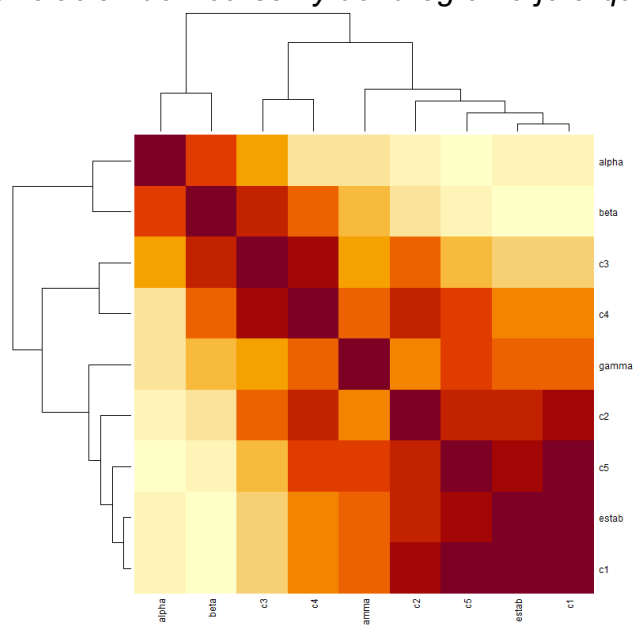
Estas diferencias se explican por la naturaleza de cada enfoque. El índice clásico sintetiza la información en una métrica escalar, mientras que el enfoque multivariado preserva la estructura conjunta de variables correlacionadas en un espacio reducido, capturando relaciones más complejas.

3.9. Estructura de correlación y agrupamiento jerárquico de variables

Se realizó un análisis de correlación de Pearson y un agrupamiento jerárquico con el objetivo de evaluar la estructura de asociación interna entre los parámetros reológicos.

Figura 11

Mapa de calor de correlación de Pearson y dendrograma jerárquico



Nota: Mapa de calor de correlación de Pearson y dendrograma jerárquico de las variables reológicas. (Autores, 2026).

El análisis evidenció grupos de variables altamente asociadas. Los parámetros C1, estabilidad y C5 conformaron un clúster fuertemente correlacionado, relacionado con la consistencia estructural y el comportamiento durante el enfriamiento. Por su parte, C3 y C4 se agruparon en un subgrupo independiente, asociado a procesos térmicos del almidón.

Los coeficientes α y β mostraron patrones de agrupamiento diferenciados, indicando que representan dimensiones complementarias con menor dependencia respecto a las variables estructurales dominantes.

El dendrograma identificó dos ramas principales: una asociada a la consistencia estructural (C1, estabilidad, C5) y otra a transiciones térmicas (C3, C4). La posición de α y β fuera de estos grupos sugiere una relativa independencia funcional.

En conjunto, estos resultados confirman una interdependencia estructurada entre variables y la naturaleza multidimensional del sistema, respaldando el uso de enfoques multivariados para su análisis.

4. Discusión

Las transiciones estructurales observadas reflejan modificaciones no lineales en la red proteína–almidón que gobierna el comportamiento termo-mecánico de las masas. En sistemas viscoelásticos, pequeñas variaciones composicionales pueden generar respuestas macroscópicas desproporcionadas (Rico et al., 2020). En este contexto, los elevados tamaños del efecto en C1 y C5 evidencian que la proporción de formulación actúa como variable de control dominante, en concordancia con estudios en matrices cereal–leguminosa (Dvořáček et al., 2019).

La sustitución progresiva de amaranto por chocho indujo una reorganización sistemática del perfil reológico. La disminución de C1, C5 y estabilidad en los niveles altos de sustitución sugiere menor resistencia estructural y alteraciones en la retrogradación, asociadas a cambios en la interacción proteína–almidón en sistemas enriquecidos con leguminosas (Meng et al., 2021; Kharbach, 2025). Estos resultados evidencian que la formulación modula simultáneamente la integridad mecánica y la respuesta térmica de la masa.

Los análisis multivariados confirmaron que los parámetros del Mixolab operan como un sistema interdependiente. La diferenciación global detectada mediante MANOVA y PERMANOVA demuestra una influencia multidimensional del tratamiento, lo que limita la interpretación basada solo en análisis univariados y respalda el uso de enfoques integrados (Granato et al., 2018; Araújo et al., 2023). En la misma línea, el PCA identificó un eje asociado a estabilidad estructural (C1, C5, estabilidad) y otro vinculado a procesos térmicos (C3, C4), mientras que la concordancia con el clustering no supervisado sugiere que la separación entre tratamientos emerge de la estructura intrínseca de los datos (Atudorei et al., 2021).

El modelo Random Forest validó la separabilidad en un espacio de alta dimensionalidad, alcanzando alta precisión bajo LOOCV. La relevancia de C1 y C5 como predictores coincide con sus tamaños del efecto y cargas en el PCA, evidenciando consistencia entre enfoques inferencial, multivariado y supervisado (Breiman, 2001; Hadnađev et al., 2011). No obstante, el tamaño muestral utilizado fue limitado ($n = 3$ por tratamiento), lo que puede restringir la generalización del modelo. Aunque LOOCV y bootstrap maximizan el uso de los datos disponibles, existe riesgo potencial de sobreajuste y optimismo en las métricas de desempeño; por tanto, los

resultados deben interpretarse como evidencia exploratoria robusta, cuya estabilidad predictiva debe confirmarse en futuros estudios con mayor número de muestras.

El análisis bootstrap aportó robustez al proceso de selección, mostrando predominio de T2 y reduciendo la dependencia de resultados puntuales, especialmente en contextos de tamaño muestral limitado (Adin et al., 2024). La comparación entre el enfoque clásico y el multivariado evidenció convergencia parcial en la identificación del tratamiento óptimo, aunque con diferencias asociadas a la incorporación de la estructura de covarianza. Mientras el índice clásico resume la información en una métrica escalar, el enfoque multivariado preserva relaciones estructurales, fortaleciendo la toma de decisiones en sistemas reológicos complejos (Araújo et al., 2023; Sheng et al., 2025).

Desde el punto de vista tecnológico, los niveles intermedios de sustitución mostraron un equilibrio entre estabilidad estructural y respuesta termo-mecánica, mientras que los niveles altos se asociaron con mayor heterogeneidad y menor consistencia. En particular, T2 representa un punto de equilibrio tecnológico, ya que mantiene una consistencia adecuada, menor degradación estructural durante el calentamiento y valores intermedios de C3 y C4, lo que sugiere gelatinización controlada y retrogradación moderada. Estas condiciones son favorables para procesos de panificación y formulación de productos farináceos, al integrar funcionalmente las propiedades del amaranto y el chocho sin comprometer la continuidad estructural de la masa.

Finalmente, aunque el diseño experimental en triplicado es estándar en estudios reológicos, futuras investigaciones con mayores tamaños muestrales, validación externa y evaluación en productos finales permitirán fortalecer la generalización del enfoque propuesto. En conjunto, la integración de análisis univariado, multivariado, aprendizaje supervisado y bootstrap mejora la confiabilidad en la toma de decisiones y constituye un marco reproducible para la optimización de sistemas farináceos multidimensionales.

5. Conclusiones

La integración de estadística clásica, modelación multivariada, aprendizaje supervisado y evaluación mediante bootstrap constituye un enfoque robusto para la optimización de harinas compuestas en sistemas reológicos multidimensionales, superando las limitaciones de los métodos basados únicamente en comparaciones de medias.

En particular, el tratamiento T2 mostró el desempeño más consistente dentro del marco integrado, destacándose tanto en el análisis bootstrap como en la selección de la repetición individual.

La estrategia permitió identificar de forma coherente el tratamiento óptimo y la repetición con mejor desempeño, evidenciando concordancia entre análisis univariado, multivariado, aprendizaje supervisado y validación por remuestreo.

Desde el punto de vista tecnológico, los niveles intermedios de sustitución mostraron un equilibrio entre estabilidad estructural y respuesta termo-mecánica, mientras que niveles altos se asociaron con mayor heterogeneidad y menor consistencia.

En conjunto, el marco propuesto constituye una arquitectura reproducible y transferible para la optimización de sistemas farináceos y otros sistemas multidimensionales, fortaleciendo la toma de decisiones en ingeniería de alimentos.

CONFLICTO DE INTERESES

“Los autores declaran no tener ningún conflicto de intereses”.

Referencias Bibliográficas

- Adin, A., Krainski, E. T., Lenzi, A., Liu, Z., Martínez-Minaya, J., & Rue, H. (2024). Automatic cross-validation in structured models: Is it time to leave out leave-one-out? *Spatial Statistics*, 62, Artículo 100843. <https://doi.org/10.1016/j.spasta.2024.100843>
- Araújo, S. O., Peres, R. S., Ramalho, J. C., Lidon, F., & Barata, J. (2023). Machine learning applications in agriculture: Current trends, challenges, and future perspectives. *Agronomy*, 13(12), Artículo 2976. <https://doi.org/10.3390/agronomy13122976>
- Atudorei, D., Atudorei, O., & Codină, G. G. (2021). Dough rheological properties, microstructure and bread quality of wheat-germinated bean composite flour. *Foods*, 10(7), Artículo 1542. <https://doi.org/10.3390/foods10071542>
- Bansal, S., Rodriguez, C. Z., Thompson-Witrick, K. A., Wang, Y., Taft, D. H., & Zhang, B. (2025). Machine learning-powered multi-omics for food microbiology and smarter food safety. *Trends in Food Science & Technology*, 163, Artículo 105145. <https://doi.org/10.1016/j.tifs.2025.105145>
- Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research*, 125(1–2), 279–284. [https://doi.org/10.1016/S0166-4328\(01\)00297-2](https://doi.org/10.1016/S0166-4328(01)00297-2)
- Benkadri, S., Salvador, A., Sanz, T., & Zidoune, M. N. (2021). Optimization of xanthan and locust bean gum in a gluten-free infant biscuit based on rice-chickpea flour using response surface methodology. *Foods*, 10(1), Artículo 12. <https://doi.org/10.3390/foods10010012>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

- Chopin Technologies. (2012). *Mixolab applications handbook: Rheological and enzyme analyses, analysis methods, studies and applications*.
- Chuqui-Paulino, F. J., Hidalgo Chávez, D. W., Ramírez Ascheri, J. L., Grassi Mellinger, C., Vargas-Solorzano, J. W., & Carvalho, C. W. P. (2025). Impact of whole cereal–pulse flours on the functionality and antioxidant properties of gluten-free extruded flours. *Foods*, 14(20), Artículo 3515. <https://doi.org/10.3390/foods14203515>
- Codină, G. G., Mironeasa, S., Mironeasa, C., Popa, C. N., & Tamba-Berehoiu, R. (2012). Wheat flour dough Alveograph characteristics predicted by Mixolab regression models. *Journal of the Science of Food and Agriculture*, 92(3), 638–644. <https://doi.org/10.1002/jsfa.4623>
- Dubat, A. (2010). A new AACC International approved method to measure rheological properties of a dough sample. *Cereal Foods World*, 55(3), 150–153. <https://doi.org/10.1094/CFW-55-3-0150>
- Duodu, K. G., & Minnaar, A. (2011). Legume composite flours and baked goods: Nutritional, functional, sensory, and phytochemical qualities. En V. R. Preedy, R. R. Watson, & V. B. Patel (Eds.), *Flour and breads and their fortification in health and disease prevention* (pp. 193–203). Academic Press. <https://doi.org/10.1016/B978-0-12-380886-8.10018-2>
- Dvořáček, V., Bradová, J., Sedláček, T., & Šárka, E. (2019). Relationships among Mixolab rheological properties of isolated starch and white flour and quality of baking products using different wheat cultivars. *Journal of Cereal Science*, 89, Artículo 102801. <https://doi.org/10.1016/j.jcs.2019.102801>
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall/CRC. <https://doi.org/10.1201/9780429246593>
- Galarza-Sánchez, P. C., Erazo-Luzuriaga, A. F., & Boné-Andrade, M. F. (2023). Uso de computación cuántica en la mejora de algoritmos de aprendizaje automático. *Revista Científica Ciencia Y Método*, 1(4), 16–30. <https://doi.org/10.55813/gaea/rcym/v1/n4/25>
- Granato, D., Santos, J. S., Escher, G. B., Ferreira, B. L., & Maggio, R. M. (2018). Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends in Food Science & Technology*, 72, 83–90. <https://doi.org/10.1016/j.tifs.2017.12.006>
- Hadnađev, T. D., Torbica, A., & Hadnađev, M. (2011). Rheological properties of wheat flour substitutes/alternative crops assessed by Mixolab. *Procedia Food Science*, 1, 328–334. <https://doi.org/10.1016/J.PROFOO.2011.09.051>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), Artículo 20150202. <https://doi.org/10.1098/rsta.2015.0202>

- Kharbach, M. (2025). AI-Powered Advances in Data Handling for Enhanced Food Analysis: From Chemometrics to Machine Learning. *Foods* 2025, Vol. 14, Page 3415, 14(19), 3415. <https://doi.org/10.3390/FOODS14193415>
- Li, M., Zhang, Y., You, X., Wang, Y., Zhou, K., Wei, P., & Wei, L. (2023). Assessment of functional properties of wheat–cassava composite flour. *Foods*, 12(19), Artículo 3585. <https://doi.org/10.3390/foods12193585>
- Meneses Quelal, O., & Pulles, M. B. (2025). Nutritional, functional and microbiological potential of Andean *Lupinus mutabilis* and *Amaranthus* spp. in the development of healthy foods—a review. *Foods*, 14(12), Artículo 2059. <https://doi.org/10.3390/foods14122059>
- Meng, K., Gao, H., Zeng, J., Zhao, J., Qin, Y., Li, G., & Su, T. (2021). Rheological and microstructural characterization of wheat dough formulated with konjac glucomannan. *Journal of the Science of Food and Agriculture*, 101(10), 4373–4379. <https://doi.org/10.1002/jsfa.11078>
- Mougan, C., & Nielsen, D. S. (2023). Monitoring model deterioration with explainable uncertainty estimation via non-parametric bootstrap. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(12), 15037–15045. <https://doi.org/10.1609/aaai.v37i12.26755>
- Rico, D., González-Paramás, A. M., Brezmes, C., & Martín-Diana, A. B. (2020). Baking optimization as a strategy to extend shelf-life through the enhanced quality and bioactive properties of pulse-based snacks. *Molecules*, 25(16), Artículo 3716. <https://doi.org/10.3390/molecules25163716>
- Rigdon, S. E., Pan, R., Montgomery, D. C., & Freeman, L. J. (2022). *Design of experiments for reliability achievement*. Wiley. <https://doi.org/10.1002/9781119237754>
- Sheng, X., Cui, Q., Yin, H., Xi, Z., Yi, J., Zhang, H., Xu, X., & Ma, Y. (2025). Application of hydrocolloids in the quality improvement of flour-based products. *Trends in Food Science & Technology*, 164, Artículo 105211. <https://doi.org/10.1016/j.tifs.2025.105211>