



Concurso de Jóvenes Biometristas

“Susana Filippini”

Posgrado

Caracterización química del grano de colza y su relación con variables ambientales

Autores

Ingeniero Agrónomo Fiore Juan Manuel

D.N.I: 39.445.424

Carrera de posgrado:

- Doctorado en Ciencias Agropecuarias (FCA-UNC)

Correo electrónico:

- juanmfio@mi.unc.edu.ar

Ingeniero Ambiental Rodríguez Nuñez Martin

D.N.I: 39.693.313

Carrera de posgrado:

- Doctorado en Ciencias de la Ingeniería (FCEFYN-UNC)
- Maestrando en Estadística Aplicada (UNC)

Correo electrónico:

- martinrodrigueznunez@mi.unc.edu.ar

Introducción

El cultivo de colza (*Brassica napus*) ocupa un lugar destacado como el segundo productor mundial de aceite vegetal (Friedt et al., 2018). Este aceite se destaca por su alto contenido de ácido graso alfa linolénico (~9%), su baja proporción de ácidos grasos saturados (~7%) y una óptima relación de omega 6 y omega 3 a 2:1. Estos atributos confieren al aceite de colza un valor nutricional excepcional, asociado con beneficios para la salud (Jahreis & Schäfer, 2011). Además, el aceite de colza posee un contenido significativo de ácido graso oleico, lo que lo convierte en una opción versátil y apreciada en diversas aplicaciones culinarias. La calidad del aceite para fritura está directamente relacionada con el índice de estabilidad oxidativa y la cantidad de ácidos grasos saturados presentes (Raboanatahiry et al., 2021).

La comparación del rendimiento y evaluación de la estabilidad de múltiples genotipos en experimentos diseñados en diferentes locaciones desempeña un papel importante en los programas de mejora de variedades de cultivos y en la recomendación de nuevas variedades para la producción. Este sistema suele denominarse ensayos multiambientales (MET) (Smith et al., 2001). Aunque el análisis de la varianza (ANOVA) es el método más ampliamente utilizado para comparar el rendimiento basado en MET, estos asumen que todos los genotipos tienen la misma varianza entre sí, lo que restringe su uso para evaluar la estabilidad del genotipo.

Un enfoque alternativo al ANOVA es el modelo de efectos principales aditivos e interacción multiplicativa (AMMI). Originalmente propuesto como un modelo de efectos fijos, el AMMI permite analizar la interacción $G \times E$ en un marco de modelos mixtos, considerando los entornos (o genotipos) como aleatorios mediante una estructura de covarianza (Piepho, 1998). Múltiples autores señalan que las herramientas estándar basadas en ANOVA no son apropiadas para los datos MET, ya que la heterogeneidad de la varianza de los efectos de interacción $G \times E$ es frecuente en estos tipos de ensayos (Kidula et al., 2000) (Coe, 2007). Por este motivo, se ha destacado la utilidad de los modelos mixtos para analizar los datos MET utilizando la metodología de máxima verosimilitud restringida (REML), considerando los efectos del entorno como aleatorios y los de genotipo como fijos (Piepho, 1994, 1998).

Con la creciente popularidad de los modelos lineales mixtos en el análisis práctico, se ha desarrollado una gama de modelos mixtos que van desde el simple componente de varianza, que proporcionan información similar al ANOVA, hasta modelos con estructuras de varianza-covarianza más complejas para modelar la interacción $G \times E$ y la variación del error. Denis et al. (1997) señalaron que diferentes métodos analíticos, incluido el ANOVA y la mayoría de las medidas de estabilidad se pueden enmarcar en un modelo mixto unificado, en el que los entornos son un factor aleatorio. Dentro de este marco, los diferentes modelos de estabilidad pueden tratarse como un modelo mixto con estructuras de varianza-covarianza, lo que puede dar lugar a mayor precisión en la evaluación del rendimiento del genotipo. Los componentes de varianza estimados de un modelo mixto adecuado sirven como medidas de estabilidad.

Dado que los ensayos en MET se realizan bajo diferentes circunstancias, la varianza del error puede variar de un entorno a otro debido a las diferencias en las condiciones naturales, como el suelo y el clima, así como las operaciones experimentales. Además, es difícil garantizar una varianza de error residual homogénea en el MET práctico. Smith et al. (2001) señalaron que la heterogeneidad de la varianza de error entre ensayos es común y, si no se tiene en cuenta, puede resultar en ineficiencias en la evaluación de los efectos de variedad. Recientemente, la contabilización de la heterogeneidad de las varianzas de error residual ha recibido atención en el análisis de datos MET basados en el enfoque de modelos mixtos (Casanoves et al., 2005).

La estimación más natural y común de los efectos del genotipo es la media aritmética simple de un genotipo en todas las réplicas en una localización. A pesar de su simplicidad, este enfoque no aprovecha plenamente toda la información contenida en el conjunto completo de datos (Piepho, 1994). Se ha demostrado que estimadores alternativos, como los estimadores por mínimos cuadrados y el mejor estimador lineal insesgado (BLUE), pueden ser predictivamente más precisos que la media aritmética. El BLUP permite realizar inferencias específicas para cada localización utilizando la información de todo el ensayo para las distintas localizaciones simultáneamente (Piepho et al., 2008; Windhausen et al., 2012)

El contenido de aceite de la colza oleaginosa varía entre el 45% y el 50% (Wittkop et al., 2009), dependiendo de factores como la fertilización, la variedad y las condiciones meteorológicas. La temperatura desempeña un papel destacado en la modificación de la cantidad de colza oleaginosa y el equilibrio entre ácidos grasos saturados, monoinsaturados y poliinsaturados (Gauthier et al., 2017). Numerosos informes indican que tanto el frío como el calor pueden influir en la composición de ácidos grasos en la colza, especialmente en el contenido de ácido linolénico (Baux et al., 2013; Canvin, 2011; Deng & Scarth, 1998; Trémolières et al., 1978).

Sin embargo, a pesar de sus numerosos beneficios, el desarrollo del cultivo de colza en Argentina ha sido limitado, y la diversidad y caracterización de cultivares se encuentran por debajo de los estándares de otros cultivos. Por ende, el propósito de este trabajo es llevar a cabo una caracterización exhaustiva de la calidad del aceite de cultivos de colza sometidos a ensayos multiambientales. Mediante esta investigación, se busca promover el avance y la adopción de la colza como un cultivo estratégico y nutritivo, que contribuya tanto a la seguridad alimentaria como al manejo sostenible de los recursos agrícolas.

Objetivos

1. Identificar los cultivares de colza con el mayor contenido de ácido linolénico y el menor contenido de ácidos grasos saturados.
2. Realizar agrupaciones de cultivares basadas en la similitud de perfiles de ácidos grasos, tocoferoles e índice de estabilidad.

3. Analizar cómo la temperatura durante el período desde el inicio de floración hasta la madurez fisiológica (IF-MF) afecta el contenido de ácido linolénico.
4. Evaluar la contribución relativa de los factores (cultivar o ambiente) en la variación de cada componente del grano, determinar si existe interacción genotipo \times ambiente y recomendar cultivares en un contexto general o específico por sitio.
5. Investigar qué factores climáticos explican la interacción entre cultivar y ambiente en relación al contenido de ácido linolénico.

Materiales y métodos

Experimento

Los datos provienen de un ensayo conducido realizado para evaluar el comportamiento de 13 cultivares de colza conducido en cuatro localidades de Argentina. Cada localidad fue considerada un ambiente específico dentro del contexto geográfico de Argentina. Para cada uno de los 13 cultivares, se realizaron dos repeticiones, lo que significa que se tomaron muestras de grano de dos parcelas experimentales diferentes en cada localidad. Dado que se trata de un experimento del cual solo se cuenta con datos recolectados el mismo año, se usará indistintamente los términos localidad y ambiente. Se determinó la composición de ácidos grasos en estas muestras, proporcionando información importante sobre las características nutricionales de la colza. Se registró: el contenido total de aceite (g/100g grano, base seca), el porcentaje sobre el total de aceite de los ácidos grasos saturados e insaturados más relevantes, el contenido de tocoferoles ($\mu\text{g/g}$ de aceite de colza) y la estabilidad del índice oxidativo (OSI).

Además, se registraron condiciones ambientales en el período comprendido entre el inicio de la floración hasta la madurez fisiológica de las plantas. Las condiciones ambientales monitoreadas incluyeron la temperatura máxima, la temperatura media, la temperatura mínima, la cantidad de precipitación y la evapotranspiración. Estos datos fueron fundamentales para comprender cómo las condiciones climáticas influyeron en el rendimiento y en las características de los cultivares de colza evaluados.

Metodología estadística aplicada

Herramientas utilizadas para responder el objetivo 1

Para identificar los cultivares de colza que presentan el contenido más alto de ácido linolénico y menor contenido de ácidos grasos saturados, implementamos un análisis de la varianza (ANOVA) basado en modelos lineales mixtos (MLM) considerando un efecto aleatorio de ambiente. Para el MLM de ácido graso linolénico se implementó una función de varianza constante *varIdent* para cumplir el supuesto de homogeneidad de las mismas. La selección de los modelos se basó en la minimización simultánea de los criterios de información de Akaike (*AIC*) y de información Bayesiana (*BIC*). Se realizó un análisis a posteriori

para inferir respecto de estas diferencias. Utilizamos el método de comparaciones múltiples *post hoc* de Di Rienzo, Guzmán y Cassanoves (DGC) permitiendo generar grupos mutuamente excluyentes.

Herramientas utilizadas para responder el objetivo 2

Para determinar cómo los cultivares se agrupan en función de la similitud de sus perfiles de ácidos grasos, tocoferoles e índice de estabilidad, se empleó un protocolo estadístico robusto que involucró múltiples herramientas estadísticas.

Inicialmente, se utilizó el índice de Silhouette y el índice de Calinski-Harabasz para determinar el número óptimo de grupos a crear a partir de los datos. Con esta información se procedió a la aplicación del algoritmo de *clustering* jerárquicos, con el objetivo de poder analizar el dendrograma de conglomeración.

Los resultados obtenidos de estos clústeres jerárquicos se compararon con los resultados de otro método de agrupación, el algoritmo *KMeans*, y con los obtenidos a partir de la visualización de un *biplot* derivado de un análisis de componentes principales (ACP). Esta comparación permitió validar la robustez de los grupos identificados.

Se decidió implementar un MLM introduciendo las variables localidad y repetición como efectos aleatorios, lo que permitió reducir la variabilidad asociada a estos factores. El objetivo fue aislar la variabilidad relacionada con el cultivar, a través de la eliminación de la vinculada tanto a localidad como repetición. De esta forma, los algoritmos de *clustering* también se ejecutaron sobre los residuos del modelo previamente mencionado.

Herramientas utilizadas para responder el objetivo 3

Para determinar cómo la temperatura afectó el contenido de ácido linolénico en los cultivos de colza, se implementó un análisis basado en un modelo de regresión lineal múltiple que incluyó todas las variables ambientales relevantes. Inicialmente, se examinaron las correlaciones entre las diferentes variables ambientales, seleccionando aquellas cuya correlación fue inferior a 0.7 y que se encontrasen más fuertemente correlacionadas con la variable respuesta (contenido de ácido linolénico). Se realizó un ajuste iterativo de modelos de regresión lineal múltiple en búsqueda del que minimizase conjuntamente los criterios de *AIC* y *BIC*, al mismo tiempo que maximizase los valores de R^2 ajustado y minimizase el error cuadrático medio de predicción. Durante este proceso, se aseguró el cumplimiento de los supuestos del modelo y se llevó a cabo un análisis exhaustivo de los residuos parciales para determinar las relaciones entre la variable respuesta y las variables independientes.

Herramientas utilizadas para responder el objetivo 4

Para evaluar la contribución relativa de los factores (cultivar o localidad) en la variación de cada componente del grano medido se utilizó un MLM, donde se introdujeron como componentes aleatorias los

factores localidad, cultivar y su interacción. Esto permitió inferir sobre el porcentaje de variabilidad explicado por cada uno de estos factores en cada componente de interés.

Posteriormente se procedió a realizar un ajuste iterativo de MLM variando tanto los factores fijos como los aleatorios e introduciendo una función *varIdent* de varianza constante que permitiese cumplir el supuesto de homogeneidad de las mismas en los casos en que fue necesario. A partir de la minimización simultánea tanto de los criterios de *AIC* y *BIC* fue que se determinó el modelo de mejor ajuste. Una vez que arribamos al modelo de mejor ajuste, evaluamos si la interacción fue o no significativa para cada uno de los componentes estudiados.

Para recomendar cultivares en sentido amplio se exploraron los valores *BLUPs* de los cultivares en un MLM que incluyó los factores cultivar, localidad y su interacción como aleatorios. Luego para cada cultivar se evaluó la estabilidad por medio de una AMMI. Es decir, se ajustó un modelo con efecto fijo cultivar y localidad para quedarnos con los residuos del modelo que contiene la interacción cultivar \times ambiente. Con los residuos se realizó un ACP y quedándonos con la CP1, variable sintética que explica el mayor porcentaje de variabilidad de la interacción, se realizó un índice de estabilidad según su desvío de la media. Posteriormente, para determinar la recomendación en sitio específico, se realizó un análisis de los *BLUPs* de la interacción entre ambiente y cultivar. Por otra parte, para recomendar cultivares en sentido específico se realizó el mismo procedimiento analítico que en el caso anterior, pero esta vez se calcularon los *BLUPs* de la interacción entre ambiente y cultivar.

Herramientas utilizadas para responder el objetivo 5

Para investigar qué factores climáticos explican la interacción entre el cultivar y el ambiente en relación al contenido de ácido linolénico, inicialmente redujimos la variabilidad de los datos para aislar la parte correspondiente a la interacción cultivar \times ambiente. Esto se logró ajustando un MLM que consideró los efectos aleatorios de ambiente, cultivar y repetición, eliminando así la variabilidad asociada a estos factores.

A continuación, exploramos dos enfoques de análisis, uno basado en estadística clásica y otro en aprendizaje automático. El enfoque basado en estadística clásica, constó del ajuste de una regresión de mínimos cuadrados parciales (PLS). Este enfoque se utilizó para comprender cómo las variables climáticas podrían explicar la variabilidad restante en los residuos, donde se encontraba presente la interacción cultivar \times ambiente.

En el enfoque de aprendizaje automático, se evaluaron múltiples algoritmos, incluidos *Random Forest*, *Light GBM* y *XGBoost Regression*. Se seleccionó aquel que demostró mejor ajuste en los datos y se optimizaron sus hiperparámetros utilizando la técnica *Leave-One-Out Cross-Validation* (LOOCV). Posteriormente, interpretamos la importancia de las variables utilizando el método de permutación. Este método mide cómo cambian los resultados del modelo luego de permutar los valores de una característica,

lo que permitió identificar las variables climáticas de mayor influencia al explicar la interacción cultivar × ambiente en el contenido de ácido linolénico.

Resultados, interpretaciones y recomendaciones

Resultados del objetivo 1

El cultivar número 2 se destacó por tener, en promedio, el contenido más elevado de ácido graso linolénico respecto a todos los cultivares, además esta diferencia fue estadísticamente significativa en comparación con el resto de los cultivares. Por otro lado, los cultivares 3 y 1 exhibieron el menor porcentaje promedio de ácido graso linolénico, y estas diferencias también fueron estadísticamente significativas en comparación con los demás cultivares. En contraste, el resto de los cultivares no mostraron diferencias estadísticamente significativas entre sí, lo que sugiere que comparten características similares en este aspecto (Figura 1).

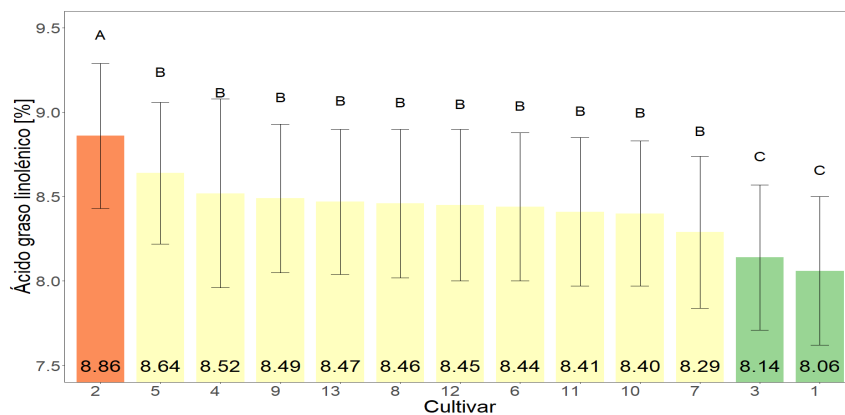


Figura 1: Contenido de ácido graso linolénico en los distintos cultivares de colza evaluados en el estudio y resultados de la prueba a posteriori de DGC.

Basándonos en estas observaciones, se establece que, en principio, si se busca recomendar un cultivar con una proporción de ácido graso linolénico superior al resto, el cultivar más recomendado sería el número 2. El MLM que minimizó los criterios de *AIC* y *BIC* para el contenido de ácido graso linolénico incluyó los siguientes componentes: efecto fijo de cultivar y repetición, efecto aleatorio de ambiente y una función de varianza constante *VarIdent* aplicada al cultivar.

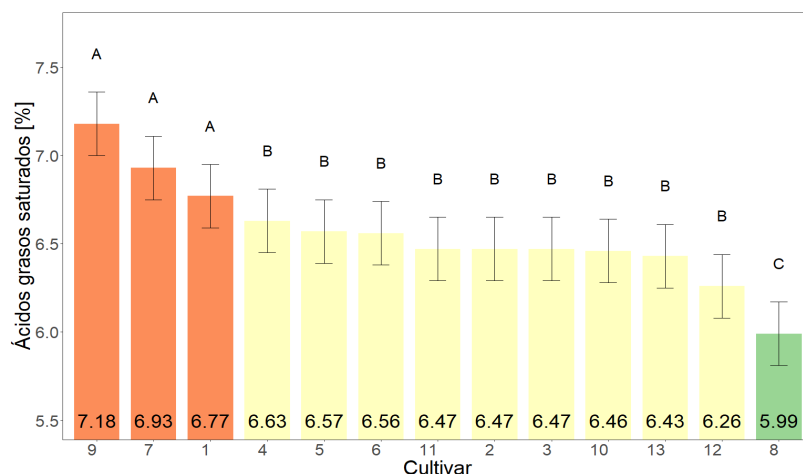


Figura 2: Contenido de ácidos grasos saturados en los distintos cultivares de colza evaluados en el estudio y resultados de la prueba a posteriori de DGC.

El cultivar número 8 se destacó al presentar el menor contenido de ácidos grasos saturados de entre todos los cultivares, además esta diferencia fue estadísticamente significativa en comparación con los demás. Por otro lado, los cultivares 9, 7 y 1 exhibieron la mayor proporción de ácidos grasos saturados, y estas diferencias también fueron estadísticamente significativas en comparación con los otros cultivares. En contraste, los demás cultivares no demostraron diferencias estadísticamente significativas entre sí (Figura 2).

Basándonos en estas observaciones, podemos afirmar que, si se busca recomendar un cultivar con una proporción de ácidos grasos saturados inferior, el cultivar más recomendado sería el número 8. Esta información es relevante para contextos donde se prioriza la reducción de ácidos grasos saturados en la colza. El MLM que minimizó los criterios de *AIC* y *BIC* para el contenido de ácidos grasos saturados incluyó los siguientes componentes: efecto fijo de cultivar y repetición, y un efecto aleatorio de ambiente.

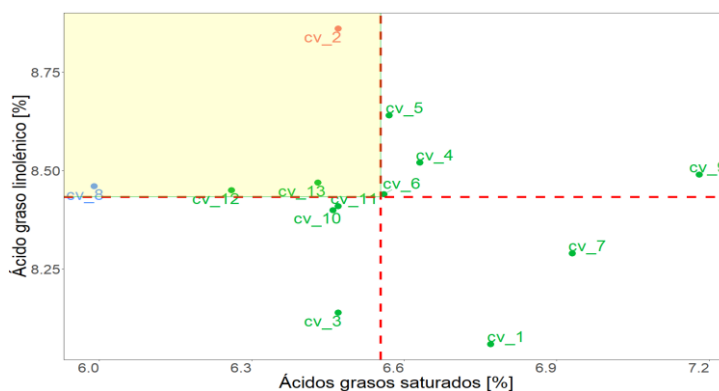


Figura 3: Análisis conjunto de las componentes de contenido de ácido graso linolénico y ácidos grasos saturados mediante un diagrama de dispersión. El área de resultados óptima se encuentra coloreada, además de los cultivares que exhibieron diferencias significativas en la prueba a posteriori que lo distingue como el óptimo.

En la Figura 3, las medias para cada una de las componentes se representan mediante líneas punteadas en color rojo. Se destaca un área resaltada en amarillo (cuadrante superior izquierdo), que

identifica la zona donde se ubican los cultivares considerados óptimos en función de maximizar el contenido de ácido graso linolénico y minimizar el contenido de ácidos grasos saturados. De esta manera, podemos observar que los cultivares 2, 8, 12 y 13 se ubican en el área óptima. A partir de estos resultados, emergen dos nuevos candidatos a óptimos (los cultivares 12 y 13). Sin embargo, es relevante mencionar que estos dos últimos cultivares no lograron diferenciarse estadísticamente del resto en ninguno de los criterios considerados en este estudio.

Por lo tanto, con base a los resultados presentados, recomendamos dos cultivares:

- Cultivar 2: Si lo que se busca es maximizar el contenido de ácido graso linolénico, este cultivar es una opción, ya que se encuentra en el área óptima, diferenciándose estadísticamente del resto en contenido de ácido graso linolénico y exhibiendo contenido de ácidos grasos saturados medio.
- Cultivar 8: Si lo que se busca es minimizar el contenido de ácidos grasos saturados, este cultivar es una elección recomendada, ya que también se encuentra en el área óptima, diferenciándose estadísticamente del resto y exhibiendo un alto contenido de ácido graso linolénico medio.

Resultados del objetivo 2

Para abordar el análisis conjunto de los perfiles de ácidos grasos, tocoferoles e índice de estabilidad en diferentes condiciones ambientales, se empleó un enfoque de agrupamiento para identificar patrones entre los cultivares. Inicialmente, se aplicó un algoritmo de conglomerado jerárquico después de estandarizar los datos. Este enfoque proporcionó una visión general de la variabilidad de los perfiles sin tener en cuenta los efectos del ambiente y la repetición.

Sin embargo, al profundizar en el análisis, se reconoció que la variabilidad relacionada con el ambiente y la repetición seguía siendo significativa. Para abordar esto, se realizó un análisis condicionado por ambiente, obteniendo el historial de conglomeración para cada ambiente. Este enfoque permitió identificar cómo se agrupaban los cultivares dentro de cada ambiente y considerar las repeticiones. A pesar de este refinamiento, se observó que aún existía variabilidad considerable entre las repeticiones para un mismo cultivar.

Para un análisis más integral, se calculó el promedio de cada uno de los componentes del perfil químico para cada cultivar, agrupando las repeticiones, y se procedió a realizar el análisis condicionado por ambiente. Los resultados en esta etapa mostraron patrones de agrupamiento comunes entre cultivares para las diferentes localidades. A pesar de que la localidad desempeña un papel importante en la formación de los grupos, surgieron tendencias similares al comparar los diferentes ambientes de cultivo. A lo través de las distintas localidades, se identificaron similitudes entre los siguientes pares de cultivares: cultivar 1 y cultivar 3; cultivar 10 y cultivar 2; cultivar 7 y cultivar 9; cultivar 11 y cultivar 5.

Dado que tanto la localidad como la repetición introducían variabilidad que no podía ser controlada, se aplicó un MLM para reducir su efecto en los datos. Este modelo incluyó como efectos aleatorios los factores cruzados de localidad y repetición. Los residuos resultantes de este modelo se utilizaron para analizar con mayor precisión los patrones de agrupamiento entre los cultivares, permitiendo una evaluación más enfocada en las relaciones intrínsecas entre los cultivares en términos de sus perfiles de ácidos grasos y otras características analizadas.

Para determinar el número adecuado de grupos que podrían formarse a partir de los datos, realizamos un análisis utilizando los índices de Calinski Harabasz y Silhouette. Al analizar su variación en función de la cantidad de grupos, determinamos que la cantidad óptima de grupos a formar es de 3.

Una vez estimados los residuos, procedimos a calcular los promedios para cada uno de los cultivares y a partir de estos datos, nuevamente aplicamos un análisis de conglomerados jerárquico. En la Figura 4 puede apreciarse el dendrograma resultante del análisis de conglomerados considerando los datos estandarizados, la distancia euclidiana y el método de enlazamiento completo.

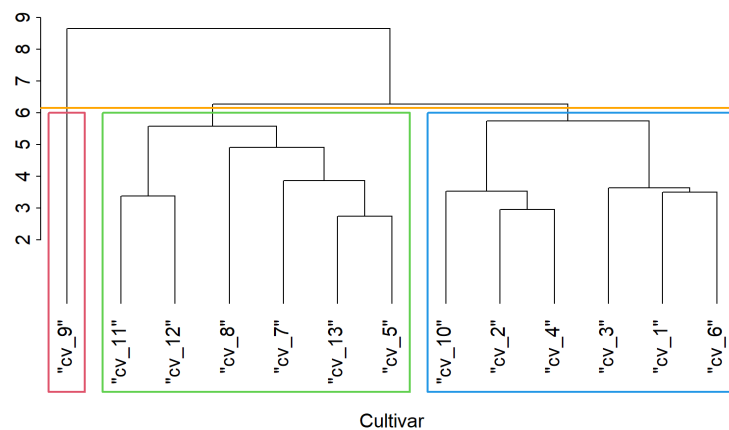


Figura 4: Dendrograma de agrupamiento jerárquico obtenido con la métrica euclídea y el algoritmo de agrupamiento completo. Los cuadrados indican los agrupamientos de los cultivares al considerar 3 grupos.

Los resultados muestran que el grupo 1 se encuentra conformado por el cultivar 9, el grupo 2 por los cultivares 11, 12, 8, 7, 13 y 5, y el grupo 3 por los cultivares 10, 2, 4, 3, 1 y 6.

Para corroborar los resultados obtenidos, procedimos a realizar un agrupamiento implementando el algoritmo de *K means clustering*. Los resultados obtenidos luego de especificar un total de 3 grupos exhibieron estrecha similitud con respecto a los del *clustering* jerárquico previamente obtenidos. El grupo 1 quedó conformado por los cultivares 7 y 9, el grupo 2 por los cultivares 11, 12, 8, 13 y 5, y el grupo 3 por los cultivares 10, 2, 4, 3, 1 y 6.

La Figura 5 muestra el *biplot* resultante del análisis de ACP, este resultado sostiene que la conformación de 3 grupos, similar a la recomendada por los índices de Calinski Harabasz y Silhouette, podría

ser una buena sugerencia. Además, este resultado sostiene y fundamenta los resultados obtenidos por los algoritmos previamente ejecutados. A pesar de esto, dadas las distancias observadas en el gráfico decidimos quedarnos con el enfoque del agrupamiento jerárquico. A partir del análisis conjunto de los resultados de los distintos enfoques detectamos que es más pertinente conservar el cultivar 7 conjuntamente con los cultivares 11, 12, 8, 13 y 5; a pesar de esto, es el cultivar que mayor similitud guarda con el número 9.

Si bien las dos primeras componentes explican solo el 55.5% de la variabilidad, se observa la agrupación similar a los clústeres. Además, nos permite tener una referencia de las características de cada grupo siendo el grupo 1 cultivares con ácidos grasos saturados. El grupo 2 con mayores concentraciones de ácido graso Oleico, Linolénico, baja concentración de ácido graso *d tocopherol* y ácido graso escénico. El grupo 3 se caracteriza por tener concentraciones superiores de ácido linoleico, tocoferoles, ácido graso palmítico. Por último, el grupo 3 es el de mayores concentraciones de tocoferoles en general con baja concentración de ácido graso oleico y contenido medio de ácidos grasos saturados (Figura 5).

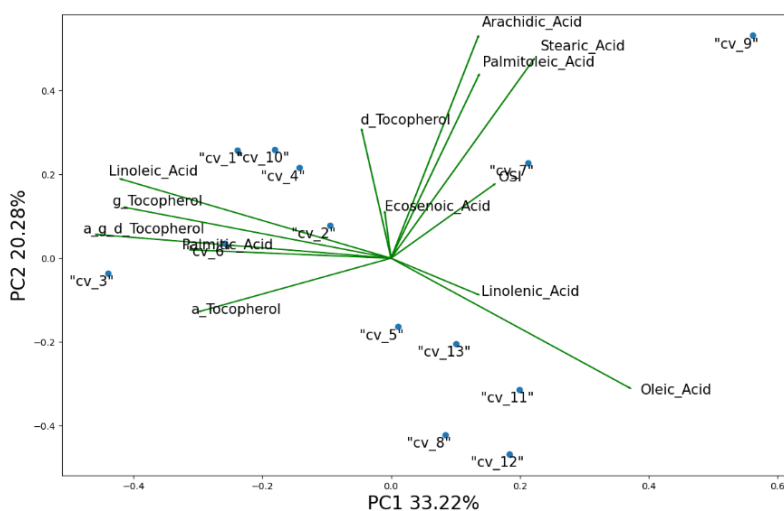


Figura 5: Gráfico biplot resultante de un Análisis de Coordenadas Principales a partir las componentes que definen el perfil químico de los cultivares de Colza evaluados.

Resultados del objetivo 3

Se observó una fuerte correlación entre las distintas variables climáticas, en especial entre las temperaturas, exhibiendo coeficientes siempre superiores a 0.93. Ante esta situación, se llevó a cabo un proceso de selección para identificar aquella temperatura que mostrase la mayor correlación con la variable respuesta. La temperatura máxima mostró el más alto coeficiente de correlación en valor absoluto (0.5). La alta correlación entre las diversas variables climáticas nos llevó a explorar un enfoque que permitiese aislar el efecto de las demás variables climáticas (evapotranspiración y precipitación) en la relación entre la temperatura y el contenido de ácido linolénico. Como respuesta a este desafío, se decidió explorar los residuos parciales del modelo, incorporando todas las variables ambientales seleccionadas previamente (precipitación, evapotranspiración y temperatura máxima). Esta estrategia nos permitió evaluar el impacto

individual de cada variable climática en la relación de interés, facilitando la identificación de los efectos específicos de la temperatura sobre el contenido de ácido linolénico.

Los resultados de la regresión lineal múltiple considerando las variables climáticas temperatura máxima, precipitación y evapotranspiración mostró un R^2 ajustado de 0.68, un AIC de 120.51 y un BIC de 133.38. Por otra parte, los coeficientes resultaron ser todos estadísticamente significativos, excepto el respectivo a la evapotranspiración. La interpretación de los residuos parciales es una valiosa técnica para examinar cómo una variable independiente específica se relaciona con la variable dependiente, teniendo en cuenta el control de todas las demás variables ambientales. En el caso de la temperatura, se observó que podría existir tanto una relación lineal como cuadrática con el contenido de ácido linolénico (Figura 6), mientras que, para la precipitación, la relación pareciera ser meramente lineal (Figura 7). Por otro lado, los resultados indicaron que la evapotranspiración no mostró significancia en esta relación.

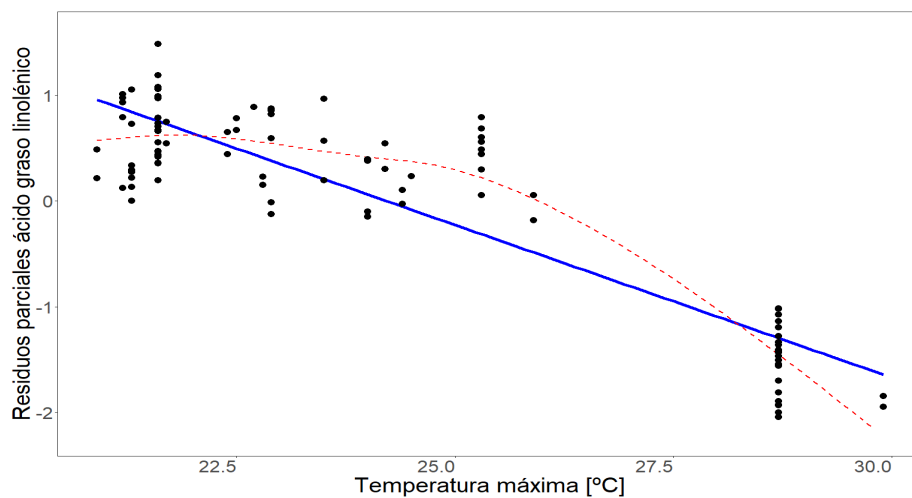


Figura 6: Gráfico de residuos parciales de ácido graso linolénico contra temperatura máxima.

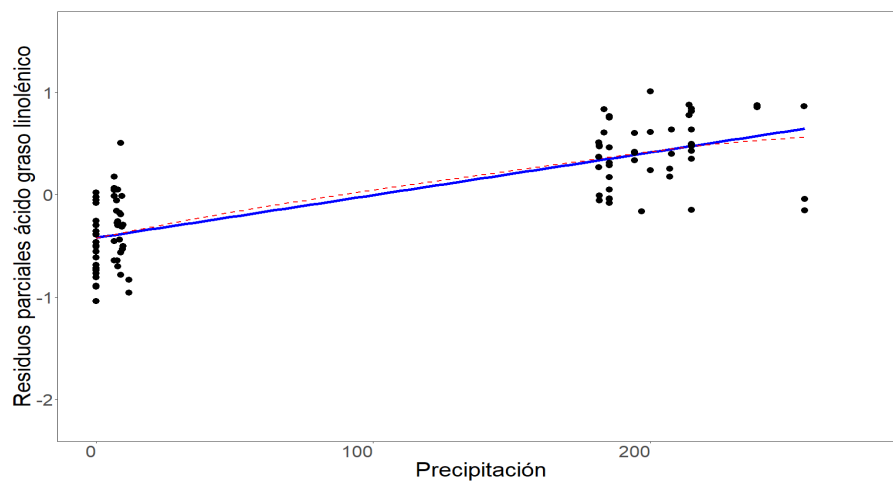


Figura 7: Gráfico de residuos parciales de ácido graso linolénico contra precipitación.

Con base en estas observaciones, procedimos a realizar un ajuste iterativo de modelos de regresión lineal múltiple hasta determinar el modelo que mejor se ajustaba a los datos. Durante este proceso, verificamos los supuestos necesarios y eliminamos tres puntos de datos con residuos inusuales que afectaban la validez de los supuestos del modelo.

La regresión lineal múltiple de mejor ajuste incluyó un término lineal y cuadrático para la temperatura máxima, y uno lineal para la precipitación. Este modelo exhibió un R^2 ajustado de 0.82, un AIC de 64.01 y un BIC de 76.88. Por otra parte, los coeficientes resultaron ser todos estadísticamente significativos, indicando que la temperatura máxima se relaciona de forma cuadrática con el contenido de ácido graso linolénico y la precipitación de forma lineal.

Una vez obtenidos estos resultados y para inferir, respecto del efecto de las variables ambientales que fueron más relevantes para explicar el contenido de ácido graso linolénico, se procedió a realizar un análisis de superficies y contornos, a partir del modelo de mejor ajuste.

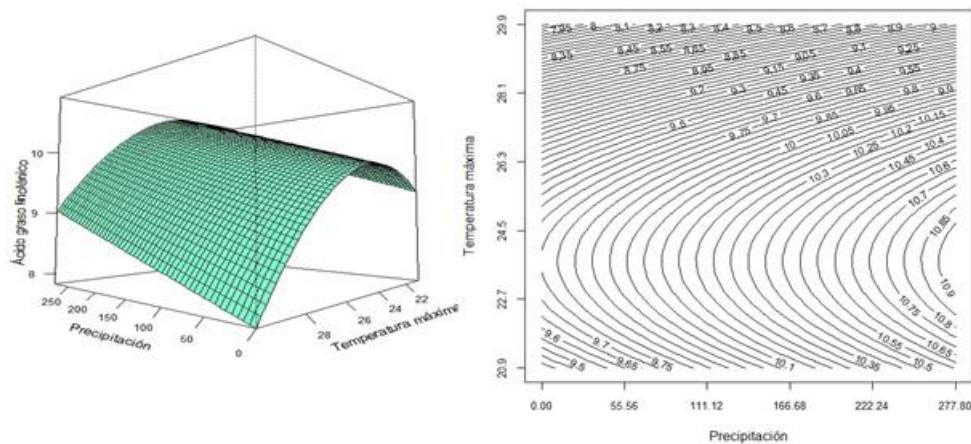


Figura 8: Gráfico de superficie y contorno para la relación de vinculación entre la temperatura máxima, la precipitación y el contenido de ácido graso linolénico. Además, se incluye su vista en planta con isolíneas de rendimiento.

A partir de los resultados exhibidos en la Figura 8, puede concluir que las condiciones climáticas óptimas para alcanzar el contenido máximo de ácido graso linolénico corresponden a precipitaciones más elevadas, específicamente registradas en la base de datos con un valor de 277.8, y temperaturas máximas que son iguales o inferiores a 23.5. En estas condiciones, se logra un contenido de ácido graso linolénico que puede alcanzar hasta 10.9. Estos hallazgos indican que la interacción entre estas variables climáticas específicas puede ser determinante para maximizar el contenido deseado de ácido graso linolénico.

Resultados del objetivo 4

A partir del modelo lineal mixto ajustado, se extrajeron los *BLUPs* para los cultivares y para la interacción $G \times E$ (en el gráfico simbolizado $G \times A$), en función de estos se recomendaron cultivares en sentido amplio y específico respectivamente. La recomendación se realizó a través de un criterio múltiple que tuvo en cuenta las variables total tocoferoles, concentración de ácido graso linolénico, ácido graso oleico,

contenido de aceite, ácidos grasos insaturados y OSI. Este criterio se formuló a partir de la literatura consultada referida a aceite de colza para consumo humano.

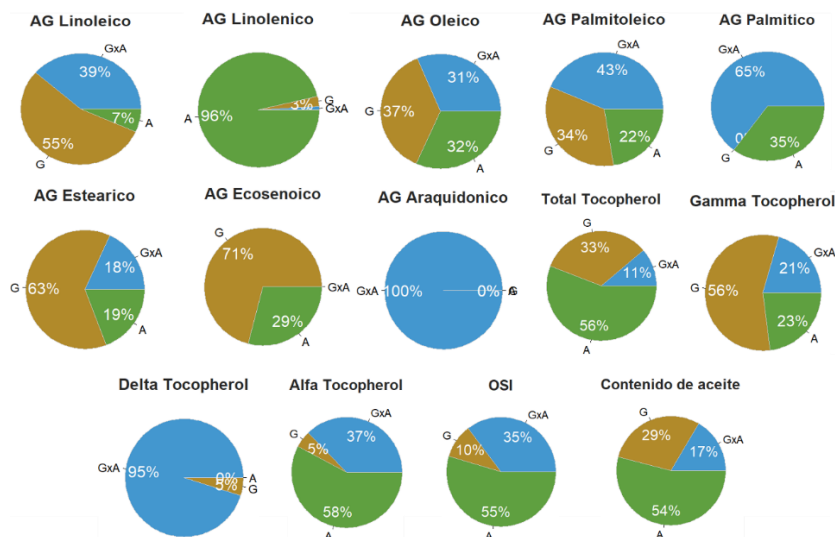


Figura 9: Gráfico de sectores de la proporción de varianza para cada componente del perfil explicada por G (marrón), A (verde) y GxA (azul).

A partir de la Figura 9, puede observarse que el porcentaje de variabilidad explicado por cada uno de los factores difiere para cada una de las componentes del perfil químico. Existen múltiples casos en los que la interacción GxA explicó el mayor porcentaje de la variabilidad en los componentes del perfil químico, estos fueron: el tocoferol delta (95%), el ácido graso palmítico (65%) y el ácido graso palmitoleico (43%). Por otro lado, los componentes del perfil químico más en los que el ambiente explica el mayor porcentaje de la variabilidad fueron el ácido graso linolénico (96%), alfa tocoferol (58%), tocoferoles totales (56%) y por último el índice de estabilidad oxidativa (55%). Los compuestos en los cuales el mayor porcentaje de variabilidad fue explicado por el genotipo fueron: son el ácido graso eicosenoico (77%), el ácido graso esteárico (63%) y el gamma tocoferol (56%). Finalmente, el ácido graso oleico es influenciado de forma similar por los distintos factores.

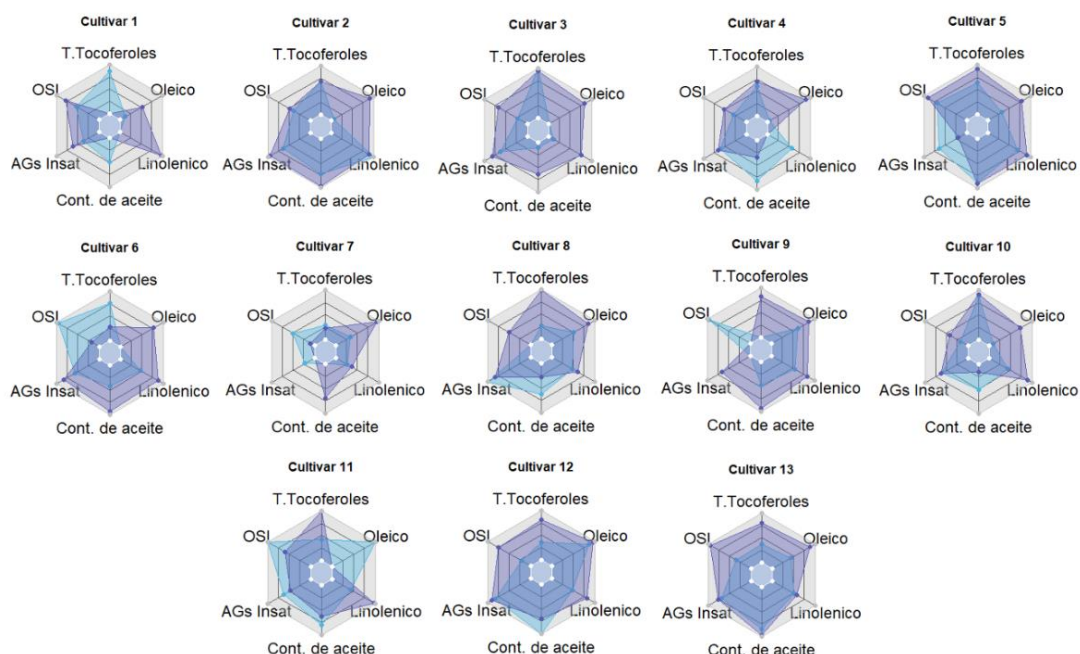


Figura 10: Gráfico de radar para los BLUPs extraídos del factor cultivar, en se indica en violeta el índice de estabilidad y en azul el rendimiento para total tocoferoles, concentración de ácido graso linolénico, ácido graso oleico, contenido de aceite, ácidos grasos insaturados y OSI.

En el proceso de selección de cultivares con fines de recomendación para su uso en la cocina, el objetivo primordial es la estabilidad de los cinco componentes del perfil químico previamente mencionados. Los componentes seleccionados incluyen el contenido total de tocoferoles, que indican su aporte vitamínico; los ácidos grasos oleico y linolénico, indicativos de su uso como fritura; así como el índice de estabilidad oxidativa (OSI), que desempeña un papel vital en su resistencia a la oxidación. Además, se evalúa el contenido global de ácidos grasos insaturados y la cantidad total de aceite, aspectos fundamentales para medir su rendimiento.

No obstante, además de tener un perfil químico adecuado, el rendimiento del cultivo es un factor determinante para garantizar su rentabilidad económica. Considerando estos criterios en conjunto, surge como recomendación principal el cultivar 2. Este cultivar se caracteriza por sus rendimientos que oscilan entre moderados y altos en la mayoría de las características evaluadas. También se destaca por su notoria estabilidad en estas propiedades, acompañada de una baja producción de ácido graso oleico. Como resultado, se perfila como una opción particularmente idónea para producir aceite de mesa de alta calidad (Figura 10).

En segundo lugar, recomendamos al cultivar 5, el cual comparte múltiples similitudes con el cultivar 2, pero se distingue por ofrecer un mayor rendimiento en términos de cantidad de aceite. Asimismo, presenta un índice de estabilidad oxidativa más elevado, lo que lo convierte en una elección más apropiada para usos relacionados con la fritura y otras técnicas culinarias similares, sin comprometer la calidad. A pesar de esto,

posee menor concentración de ácido linolénico, lo cual lo coloca en una posición desfavorable frente al cultivar 2 (Figura 10).

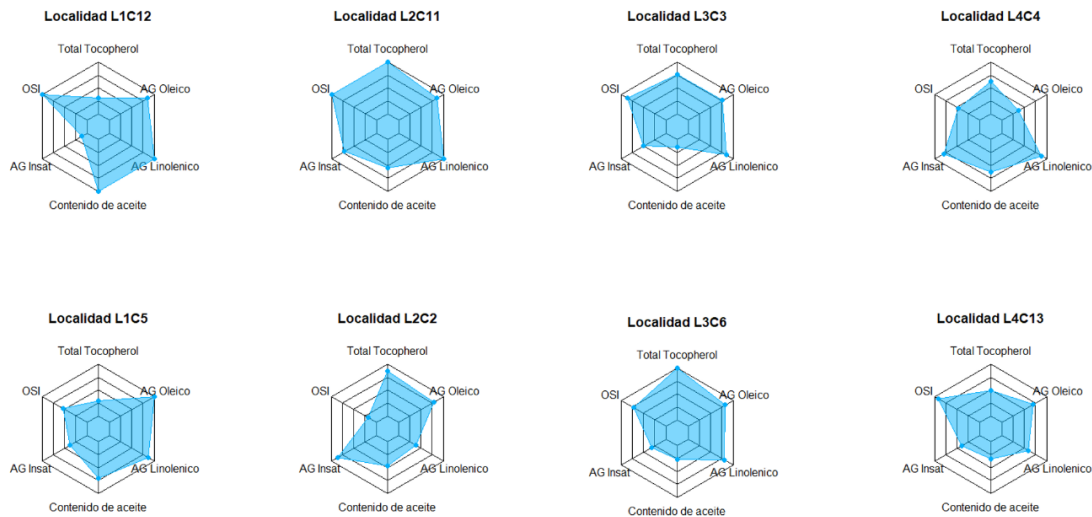


Figura 11: Gráfico de radar para los BLUPs extraídos del factor interacción entre cultivar y localidad, en se indica en violeta el índice de estabilidad y en azul el rendimiento para total tocoferoles, concentración de ácido graso linolénico, ácido graso oleico, contenido de aceite, ácidos grasos insaturados y OSI.

En el proceso de selección para sitio específico se evaluó en función del mismo criterio implementado para la recomendación en sentido amplio de cultivares. Se observó de manera consistente, en todas las localidades estudiadas, que el cultivar 5 se mantuvo en las posiciones de liderazgo. Dentro de este marco, en la Localidad 1 se optó por el cultivar 12 debido a su destacado desempeño. Este cultivar registró rendimientos notables en términos de contenido de aceite, ácido linolénico y oleico, además de presentar un índice de estabilidad oxidativa (OSI) elevado. En la Localidad 2, se tomó la decisión de seleccionar el cultivar 11. A pesar de tener un rendimiento ligeramente inferior en términos de contenido de aceite, este cultivar demostró un desempeño satisfactorio en la mayoría de los compuestos evaluados, consolidándose como una opción sólida. En cuanto a la Localidad 3, se identificó un comportamiento notablemente similar entre los cultivares 6 y 3, lo que llevó a la recomendación de ambos como opciones viables para esta ubicación. Finalmente, para la Localidad 4, se llegó a la conclusión de que el cultivar 4 es la elección más adecuada, considerando su rendimiento sobresaliente en comparación con las alternativas restantes (Figura 11).

Resultados del objetivo 5

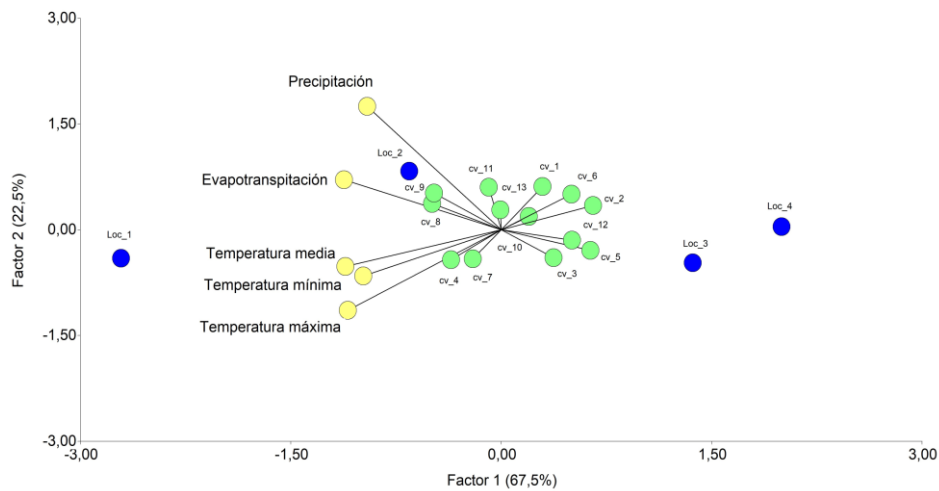


Figura 12: Biplot a partir de los resultados del análisis de PLS que demuestra los pesos de las variables climáticas para explicar la interacción entre cultivar y ambiente.

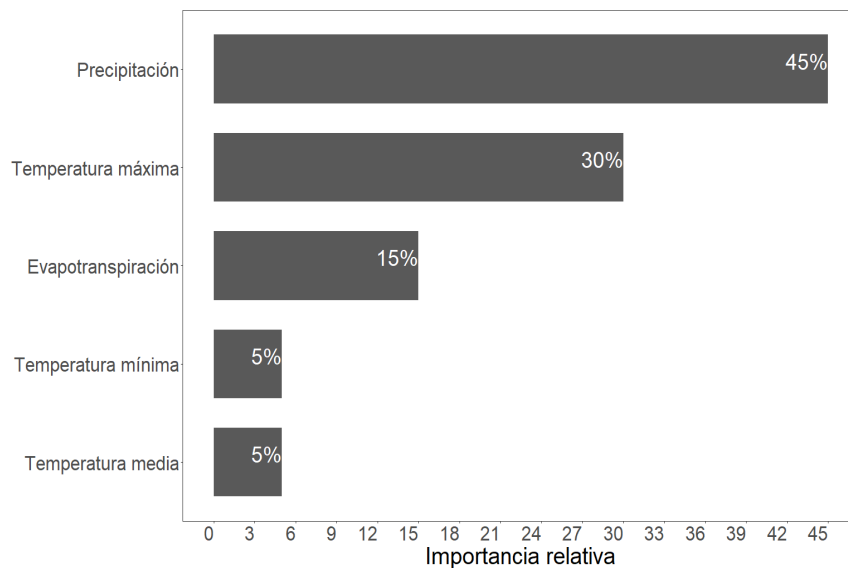


Figura 13: Importancia de las variables climáticas a partir del enfoque de aprendizaje automático para explicar la interacción entre cultivar y ambiente.

Los resultados del análisis de *PLS* (Figura 12) resaltan que las variables de mayor relevancia para explicar la interacción entre el cultivar y la localidad son la precipitación y la temperatura máxima. En este enfoque, la temperatura máxima demostró ser más influyente que la mínima y la media, dado a que su distancia con respecto al origen fue superior. A su vez, su distancia con respecto a la de la precipitación fue también superior, colocándola en el segundo lugar de importancia a la hora de explicar la interacción entre cultivar y localidad.

Por otro lado, el enfoque de aprendizaje automático reveló que las variables más significativas para explicar esta interacción fueron la precipitación en primer lugar, y en segundo la temperatura máxima (Figura 13).

En resumen, podemos concluir que las variables más cruciales para explicar la interacción entre el cultivar y la localidad son la precipitación y la temperatura máxima, en función de ambos enfoques de análisis.

Conclusión

En el contexto de este estudio, se han cumplido los objetivos planteados, iluminando aspectos esenciales en la producción de colza y sus efectos en la calidad de los aceites resultantes. Una contribución fundamental ha sido la identificación de cultivares de colza con perfiles de ácidos grasos altamente deseables. El Cultivar 2 ha sobresalido por su alto contenido de ácido linolénico, lo que lo convierte en una opción prometedora para obtener aceites enriquecidos con este componente beneficioso. De manera similar, el Cultivar 8 ha demostrado un contenido reducido de ácidos grasos saturados, presentándose como una elección óptima para minimizar estos componentes indeseados en los aceites.

La utilización de técnicas como el análisis de componentes principales (ACP), el agrupamiento *Kmeans* y el agrupamiento jerárquico han permitido clasificar los cultivares en grupos con perfiles químicos y de estabilidad similares. Aunque los resultados de los enfoques fueron notoriamente similares, se ha optado por mantener los grupos formados por el algoritmo jerárquico. En este sentido, el cultivar 7 ha sido incluido junto a los cultivares 11, 12, 8, 13 y 5, mientras que el cultivar número 9 ha quedado como único integrante del tercer grupo. A pesar de esto, el cultivar número 7 se destaca como el de mayor parentesco al 9 en caso de necesitar reemplazarlo.

Un análisis detenido del impacto climático en el contenido de ácido linolénico ha subrayado la influencia crítica de la temperatura y la precipitación. Los resultados de la regresión lineal múltiple indican una relación cuadrática con la temperatura máxima y una relación lineal con la precipitación, subrayando así la necesidad de considerar las condiciones climáticas para optimizar los perfiles de ácidos grasos durante el ciclo de desarrollo de la colza.

Las recomendaciones que surgen del estudio son fundamentales tanto para la elección de cultivares en un sentido amplio como para contextos específicos. El cultivar 2 se erige como la opción principal para la producción de aceite de alta calidad debido a su estabilidad y rendimientos equilibrados. Por otro lado, el cultivar 5 se destaca para aplicaciones culinarias, como la fritura. En el contexto de ubicaciones específicas, el cultivar 12 ha demostrado un desempeño destacado en la localidad 1, mientras que el cultivar 11 se presenta como una opción sólida en la localidad 2, los cultivares 3 y 6 se distinguen en la localidad 3, y finalmente el cultivar 4 se recomienda en la localidad 4.

Una conclusión vital es que la interacción entre los factores cultivar y ambiente está principalmente influenciada por la temperatura máxima y la precipitación. Estos factores climáticos se han revelado como los principales impulsores de la variación de respuesta de los cultivares en distintas ubicaciones. Este

hallazgo recalca la importancia de ajustar la selección de cultivares a las condiciones locales para maximizar la calidad y el rendimiento.

En síntesis, este estudio ha proporcionado una comprensión holística de la producción de colza, desde la identificación de cultivares óptimos hasta la influencia del clima en los perfiles de ácidos grasos. Las recomendaciones específicas para distintos usos y contextos resaltan la aplicabilidad práctica de los resultados. Habiendo delineado claramente la interacción genotipo \times ambiente, este trabajo se presenta como un recurso invaluable para la toma de decisiones en la industria de la colza y sus derivados.

Bibliografía

- Baux, A., Colbach, N., Allirand, J. M., Jullien, A., Ney, B., & Pellet, D. (2013). Insights into temperature effects on the fatty acid composition of oilseed rape varieties. *European Journal of Agronomy*, 49, 12–19. <https://doi.org/10.1016/J.EJA.2013.03.001>
- Canvin, D. T. (2011). THE EFFECT OF TEMPERATURE ON THE OIL CONTENT AND FATTY ACID COMPOSITION OF THE OILS FROM SEVERAL OIL SEED CROPS. *https://doi.org/10.1139/B65-008*, 43(1), 63–69. <https://doi.org/10.1139/B65-008>
- Casanoves, F., Macchiavelli, R., & Balzarini, M. (2005). Error Variation in Multienvironment Peanut Trials: Within-Trial Spatial Correlation and Between-Trial Heterogeneity. *Crop Science*, 45(5), 1927–1933. <https://doi.org/10.2135/CROPSCI2004.0547>
- Deng, X., & Scarth, R. (1998). Temperature effects on fatty acid composition during development of low-linolenic oilseed rape (*Brassica napus* L.). *Journal of the American Oil Chemists' Society*, 75(7), 759–766. <https://doi.org/10.1007/S11746-998-0223-4>
- Denis, J. B., Piepho, H. P., & Van Eeuwijk, F. A. (1997). Modelling expectation and variance for genotype by environment data. *Heredity* 1997 79:2, 79(2), 162–171. <https://doi.org/10.1038/hdy.1997.139>
- Friedt, W., Tu, J., & Fu, T. (2018). *Academic and Economic Importance of Brassica napus Rapeseed*. 1–20. https://doi.org/10.1007/978-3-319-43694-4_1
- Gauthier, M., Pellet, D., Monney, C., Herrera, J. M., Rougier, M., & Baux, A. (2017). Fatty acids composition of oilseed rape genotypes as affected by solar radiation and temperature. *Field Crops Research*, 212, 165–174. <https://doi.org/10.1016/J.FCR.2017.07.013>
- Jahreis, G., & Schäfer, U. (2011). Rapeseed (*Brassica napus*) Oil and its Benefits for Human Health. *Nuts and Seeds in Health and Disease Prevention*, 967–974. <https://doi.org/10.1016/B978-0-12-375688-6.10114-8>
- Piepho, H. P. (1994). Best Linear Unbiased Prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. *Theoretical and Applied Genetics*, 89(5), 647–654. <https://doi.org/10.1007/BF00222462/METRICS>
- Piepho, H. P. (1998). Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. *Theoretical and Applied Genetics*, 97(1–2), 195–201. <https://doi.org/10.1007/S001220050885/METRICS>

- Piepho, H. P., Möhring, J., Melchinger, A. E., & Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*, 161(1–2), 209–228. <https://doi.org/10.1007/S10681-007-9449-8/FIGURES/4>
- Quantitative Analysis of Data from Participatory Methods in Plant Breeding* - Google Libros. (n.d.). Retrieved August 16, 2023, from https://books.google.com.ar/books?hl=es&lr=&id=iLGFJr4zz-EC&oi=fnd&pg=PA44&ots=HdolXZrDSI&sig=k8z7mBBFI1_o4qXYHCUBNOsmanY&redir_esc=y#v=onepage&q&f=false
- Raboanatahiry, N., Li, H., Yu, L., & Li, M. (2021). Rapeseed (*Brassica napus*): Processing, Utilization, and Genetic Improvement. *Agronomy* 2021, Vol. 11, Page 1776, 11(9), 1776. <https://doi.org/10.3390/AGRONOMY11091776>
- Smith, A., Cullis, B., & Gilmour, A. (2001). Applications: The Analysis of Crop Variety Evaluation Data in Australia. *Australian & New Zealand Journal of Statistics*, 43(2), 129–145. <https://doi.org/10.1111/1467-842X.00163>
- Trémolières, H., Trémolières, A., & Mazliak, P. (1978). Effects of light and temperature on fatty acid desaturation during the maturation of rapeseed. *Phytochemistry*, 17(4), 685–687. [https://doi.org/10.1016/S0031-9422\(00\)94209-2](https://doi.org/10.1016/S0031-9422(00)94209-2)
- Windhausen, V. S., Wagener, S., Magorokosho, C., Makumbi, D., Vivek, B., Piepho, H. P., Melchinger, A. E., & Atlin, G. N. (2012). Strategies to Subdivide a Target Population of Environments: Results from the CIMMYT-Led Maize Hybrid Testing Programs in Africa. *Crop Science*, 52(5), 2143–2152. <https://doi.org/10.2135/CROPSCI2012.02.0125>
- Wittkop, B., Snowdon, R. J., & Friedt, W. (2009). Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. *Euphytica*, 170(1), 131–140. <https://doi.org/10.1007/S10681-009-9940-5/TABLES/3>