

UNIVERSIDAD DE SALAMANCA
DEPARTAMENTO DE ESTADÍSTICA
DOCTORADO EN ESTADÍSTICA MULTIVARIANTE APLICADA



**Evaluación multivariante sobre
el rendimiento y la estabilidad
de variedades de caña de azúcar
en Ecuador**

TESIS DOCTORAL

LUIS HENRY TORRES ORDOÑEZ

Directora:

PhD. M^a Purificación Galindo Villardón

SALAMANCA, ESPAÑA 2024

Evaluación multivariante sobre el rendimiento y la estabilidad de variedades de caña de azúcar en Ecuador



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

DEPARTAMENTO DE ESTADÍSTICA

Memoria para optar al Grado de Doctor en
Estadística Multivariante Aplicada
por el Departamento de Estadística
de la Universidad de Salamanca,
presenta:

Luis Henry Torres Ordoñez

Salamanca 2024

Certificación



VNiVERSIDAD
D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

DRA. PURIFICACIÓN GALINDO VILLARDÓN

CATEDRÁTICA EMERITA DEL DEPARTAMENTO
DE ESTADÍSTICA DE LA UNIVERSIDAD
DE SALAMANCA, ESPAÑA.

CERTIFICA:

Que D. **LUIS HENRY TORRES ORDOÑEZ**, ha realizado en el Departamento de Estadística de la Universidad de Salamanca, bajo su dirección, el trabajo para optar al Grado de Doctor en Estadística Multivariante Aplicada, que presenta con el título **Evaluación multivariante sobre el rendimiento y la estabilidad de variedades de caña de azúcar en Ecuador**, autorizando expresamente su lectura y defensa, y para que conste, firman el presente certificado en Salamanca a 6 de Noviembre de 2024.

Dra. Ma. Purificación Galindo Villardón

Agradecimientos

Empiezo por agradecer a Dios, mi Señor, sin Él nada es posible.

Para la Dra. Purificación Galindo Villardón mi más sincero agradecimiento por su invaluable guía y apoyo en este proceso. Su dedicación y conocimiento han sido fundamentales para el desarrollo de mi investigación, y gracias a usted he aprendido a enfrentar los desafíos académicos con rigor y pasión. No solo he crecido como investigador, sino también como persona, aprendiendo a valorar el esfuerzo y la perseverancia. Aprecio profundamente cada consejo y el tiempo que ha dedicado para ayudarme a avanzar ya sea desde Salamanca o en Ecuador. Dra. Puri le estaré siempre agradecido por ser una mentora tan excepcional y por dejar una huella tan positiva en mi formación académica, Muchas gracias.

A la Dra. Elisa Frutos quien me apoyó en su debido momento y me supo direccionar en este tema de investigación cuando estaba en mis estancias doctorales en el departamento de Estadística en Salamanca. Al Dr. Diego Valenzuela, y su equipo de trabajo que desde el departamento de Estadística de la Universidad Estatal de Milagro me han apoyado en este trabajo de investigación. Estoy eternamente agradecido.

Cada objetivo alcanzado lleva impreso el amor, sacrificio y dedicación de mi familia. Mención especial merecen mis hijos, Henry David y Luis Francisco, la fe inquebrantable de ellos y su amor constante me han empujado a alcanzar alturas que jamás imaginé; mis padres, Efrén (+) y Alicia, ejemplos a seguir; mis

hermanos, mis cuñados y sobrinos, que es fuente inagotable de amor y apoyo. Sin ustedes, cada logro y meta alcanzada no tendría el mismo significado. Gracias por ser el motor detrás de mis sueños.

Y como no mencionar a ti, Argentina, no tengo palabras suficientes para agradecerte todo lo que has hecho por mí durante este proceso. Gracias por tu paciencia, tu apoyo incondicional y por creer en mí, incluso en los momentos en los que yo dudaba. Sin ti, no hubiera llegado hasta aquí. Te amo y este logro es tanto tuyo como mío.

A todos quienes han aportado al logro de este tan alto propósito, mi eterno agradecimiento.

Dedicatoria

*A mis hijos, Henry David y Luis
Francisco, a mi mamá Alicia y a mi
papá (+) Efrén que desde el cielo me ve
triunfar, esto va por ti papá.*

Resumen

Los métodos de interacción genotipo por genotipo+ambiente (GGE) son técnicas estadísticas empleadas en agronomía y mejoramiento genético para analizar y visualizar la variabilidad del rendimiento de diferentes genotipos en múltiples ambientes. Desde el punto de vista estadístico, el enfoque GGE se basa en descomponer la matriz de datos de rendimiento mediante análisis de componentes principales (PCA) o descomposición en valores singulares (SVD), enfocándose en los efectos combinados del genotipo y la interacción genotipo x ambiente ($G + G \times E$). Esto permite identificar patrones y relaciones entre genotipos y ambientes, facilitando la selección de genotipos superiores y la comprensión de cómo responden a distintas condiciones ambientales. Esta investigación es una aplicación del método GGE biplot, para analizar su eficiencia en la evaluación del rendimiento en múltiples ambientes, tomando a estos como ubicaciones y años. El método desarrollado se aplicó en la industria agronómica, específicamente en el sector cañicultor, para evaluar el rendimiento de caña de azúcar en toneladas métricas por hectáreas. Para una mejor comprensión de las observaciones se generaron análisis adicionales, como análisis de la varianza (ANOVA) y el test de Tukey HSD. Los resultados revelaron que emplear ANOVA es útil hasta cierto punto para capturar una pequeña parte de la varianza y determinar en específico la existencia de diferencias significativas, así mismo el test de Tukey demostró su capacidad para determinar los mejores genotipos, sectores y años, sin embargo, no considerando la interacción entre estas entradas. El modelo GGE biplot, demostró su gran capacidad para explorar los ensayos agronómicos, para determinar los mejores genotipos en distintos ambientes de prueba, ofreciendo bases teóricas y prácticas para la toma de decisiones en la selección de variedades específicas, además de ser beneficioso para los investigadores que utilizan técnicas de exploración de biplot.

Palabras claves: Métodos biplot, GGE biplot, Interacción, datos agronómicos

Índice General

CAPÍTULO I	14
1. INTRODUCCIÓN Y OBJETIVOS	14
1.1. Formulación del problema	17
1.1.1. Sistematización:	18
1.3. Objetivos de la investigación.....	18
1.3.1. Objetivo general.....	18
1.3.2. Objetivos específicos.....	18
1.4. Justificación de la investigación	19
1.5. Hipótesis del trabajo	20
CAPÍTULO II	22
2. MARCO TEÓRICO.....	22
2.1. Interacción Genotipo Ambiente (GEI)	22
2.2. Concepto del GGE	24
2.3. El modelo GGE biplot	25
2.4. El concepto de Biplot.....	27
CAPÍTULO III	32
3. CONTRIBUCIONES A LA PRODUCCIÓN AGRÍCOLA DE LA CAÑA DE AZÚCAR DESDE UNA PERSPECTIVA MULTIVARIANTE: ANÁLISIS DEL RENDIMIENTO DE VARIEDADES DE CAÑA DE AZUCAR Y SUS EFECTOS INTERACTIVOS GENOTIPO-AMBIENTE.	32
3.1. Diseño experimental	33
3.2. Material vegetal – Genotipos.....	34
3.3. Ambientes (sectores – años).....	35
3.4. Rendimiento (Toneladas/Hectáreas)	36
3.5. Conjunto de datos	37
3.6. Exploración de las fuentes de variación	38
3.7. Análisis de la varianza (ANOVA) y descomposición por sectores.....	41
3.7.1. Análisis de las fuentes de variación.....	45
3.8. Análisis de los patrones biplot	48
3.8.1. Representación GGE biplot	49
CAPÍTULO IV	60
4. CONTRIBUCIONES CIENTÍFICAS.....	60
CONCLUSIONES	69

Índice de Tablas

Tabla 1. Modelo de interacción Genotipo-Ambiente (GEI).....	22
Tabla 2. Características de los ambientes experimentales.	38
Tabla 3. Rendimiento medio del monocultivo en una tabla de tres entradas.....	40
Tabla 4. Análisis de Varianza para efectos principales e interactivos.	42
Tabla 5. Suma Cuadrática de diferentes ANOVAS.....	44
Tabla 6. Prueba Post Hoc de Tukey HSD	46

Índice de Figuras

Figura 1. Representación gráfica de los tipos de GEI	24
Figura 2. Diagrama comparativo entre el modelo completo y el modelo GGE.....	26
Figura 3. Diagrama de caja y bigote de diferentes genotipos.	39
Figura 4. Diagrama de caja y bigote de diferentes ambientes.	40
Figura 5. Ilustración GGE biplot básica.....	52
Figura 6. Patron gge biplot (Which-won-where).....	53
Figura 7. Patron gge biplot Mean vs. Stability.....	55
Figura 8. Patrones gge biplot Ranking y Discriminativeness vs. Representativeness.	57

CAPÍTULO I
INTRODUCCIÓN Y OBJETIVOS

CAPÍTULO I

1. INTRODUCCIÓN Y OBJETIVOS

1.1. Introducción

Los ensayos multi-ambientales (MET) se utilizan rutinariamente para evaluar variedades de plantas, comparando múltiples genotipos en diversos entornos (años y ubicaciones) y para varios rasgos. Estos ensayos generan datos estructurados en tres dimensiones: genotipo (G), entorno (E) y rasgo (F). A pesar de su valor potencial para la selección y recomendación de cultivares, el análisis de estos ensayos suele centrarse en un solo rasgo, generalmente el rendimiento, lo que deja otros rasgos sin explotar (Yan & Tinker, 2006).

Tradicionalmente, el análisis de varianza (ANOVA) ha sido uno de los métodos más utilizados para describir las respuestas fenotípicas de los genotipos en diferentes entornos. Sin embargo, el ANOVA presenta limitaciones importantes, ya que no captura adecuadamente las interacciones genotipo \times ambiente ($G \times E$). Este modelo asume que todos los genotipos responden de manera similar a los cambios en el entorno, lo cual no refleja la realidad, donde las respuestas de los genotipos suelen variar significativamente en función de las condiciones ambientales. Esta falta de capacidad para modelar la interacción $G \times E$ hace que el ANOVA sea inadecuado para estudios multi-ambientales más complejos. En respuesta a estas limitaciones, se han desarrollado métodos más avanzados, como los modelos de regresión sobre la media o los modelos lineales-bilineales, que permiten una caracterización más flexible y precisa de estas interacciones (Oladosu et al., 2017; Van Eeuwijk et al., 2005).

Para abordar esta problemática, los MET han impulsado el desarrollo de diversos modelos estadísticos que permiten no solo modelar los efectos principales, sino también las interacciones genotipo-ambiente ($G \times E$), al tiempo que manejan la heterogeneidad de las varianzas y las correlaciones genéticas entre ambientes, superando las limitaciones del ANOVA. Además, estos modelos integran tanto efectos fijos (como el genotipo, el entorno o las condiciones de manejo) como efectos aleatorios (como los bloques incompletos y las replicaciones), proporcionando un análisis más detallado de las interacciones $G \times E$ (Smith et al., 2001).

Algunos de estos modelos tienen la capacidad de estimar parámetros clave, como la heredabilidad en sentido amplio, que indica la proporción de la variación observada atribuible a factores genéticos. Asimismo, permiten calcular las correlaciones genéticas y fenotípicas entre ambientes y rasgos, lo que facilita la identificación de genotipos con buen desempeño en una variedad de entornos y condiciones de manejo (Alvarado et al., 2020). Entre los modelos que permiten evaluar la estabilidad y el rendimiento del genotipo en diferentes entornos destacan los modelos mixtos multiplicativos, el modelo GGE (Genotipo \times Genotipo-Ambiente) y AMMI (Efectos Principales Aditivos e Interacción Multiplicativa) (Liu et al., 2022; Smith et al., 2001; Yan et al., 2000, 2001).

Por un lado, el modelo AMMI (Efectos Principales Aditivos e Interacción Multiplicativa) combina el análisis de los efectos principales aditivos, que capturan las respuestas promedio de los genotipos y entornos, y la interacción multiplicativa, que descompone las interacciones $G \times E$ en componentes principales (Annicchiarico, 1997; Sa'diyah & Hadi, 2016). Tiene la capacidad para identificar y separar las respuestas reales de los genotipos a los ambientes (patrón) del ruido o variabilidad aleatoria y se logra mediante la adición del término multiplicativo $\sum_{k=1}^k \lambda_k \gamma_{ik} \delta_{jk}$ donde se contienen los valores singulares y los vectores de carga de los genotipos y entornos para el componente principal k (Annicchiarico, 1997). Por otro lado, El modelo GGE Biplot además de también utilizarse para analizar datos de ensayos multi-ambientales, facilita la evaluación gráfica de los efectos principales del genotipo (G) y la interacción genotipo-ambiente ($G \times E$). Este modelo es efectivo porque combina ambas fuentes de variación, que son clave para la selección de cultivares y la identificación de ambientes representativos. Este modelo es efectivo porque combina ambas fuentes de variación, que son clave para la selección de cultivares y la identificación de ambientes representativos (Yan et al., 2001; Yan & Kang, 2002).

Aunque existe un debate entre los defensores de los métodos biplot AMMI y GGE con respecto a cuál es más efectivo para evaluar el rendimiento y la estabilidad genética en los cultivos, ambos métodos son efectivos para este fin en estudios multi-ambientales, pero cada uno presenta ventajas específicas. (Maulana et al., 2023; Mehareb et al., 2022; Yadawad et al., 2023). Por su parte los modelos AMMI, combinan efectos principales aditivos y componentes multiplicativos que permiten descomponen las interacciones genotipo-ambiente ($G \times E$) en partes interpretables, en este sentido AMMI es

particularmente útil para identificar genotipos estables, aquellos que tienen un rendimiento consistente a través de diferentes entornos, y para seleccionar los ambientes que generan menos interacción $G \times E$ (Annicchiarico, 1997; Haider et al., 2017; Scavo et al., 2023). Por otro lado, los modelos GGE biplot, como ya se mencionó combina el efecto del genotipo y la interacción $G \times E$, lo que facilita la interpretación gráfica del rendimiento de los genotipos y su estabilidad, sin embargo, una ventaja clave del GGE es su capacidad para identificar mega-ambientes, además de su capacidad de identificar genotipos “ganadores” permitiendo que los mejoradores optimicen la selección de genotipos para diferentes regiones geográficas (Haider et al., 2017; Yan & Kang, 2002).

Una de las principales ventajas del modelo GGE biplot es su capacidad para generar diferentes patrones de representación que facilitan la interpretación de los datos. Entre los patrones más importantes se encuentra el de “Which-Won-Where” (cuál-ganó-dónde), el cual se basa en la correlación entre genotipos y ambientes (Yan & Kang, 2002). Este patrón genera un polígono con vértices que representan los genotipos estudiados. Cuando un genotipo se encuentra en el vértice de un sector donde caen varios ambientes, sugiere que ese genotipo ha proporcionado un mayor rendimiento y ha tenido un mejor desempeño en dichos ambientes. Por el contrario, si un genotipo ocupa un vértice donde no hay ningún marcador ambiental, indica que ese genotipo ha tenido un rendimiento deficiente en todos los ambientes evaluados, finalmente, los genotipos ubicados dentro del polígono tienden a ser menos sensibles al ambiente en comparación con aquellos en los vértices (Esan et al., 2023; Khan et al., 2021).

El segundo patrón relevante es el de “Media vs. Estabilidad”, que permite seleccionar y evaluar el genotipo ideal dentro de un mega-ambiente de referencia. Este patrón evalúa a los genotipos tanto por su rendimiento promedio como por su estabilidad en diferentes ambientes. El patrón muestra la coordinación ambiental promedio (AEC), representada por una línea con una flecha que actúa como el eje del ambiente promedio (AEA) (Yan & Tinker, 2006). La flecha apunta hacia el genotipo con el mayor rendimiento, mientras que la ordenada del AEC es perpendicular a la abscisa del AEC, indicando mayor variabilidad (y, por tanto, menor estabilidad) en cualquier dirección (Frutos et al., 2014; Khan et al., 2021).

Finalmente, el patrón de “Discriminación vs. Representatividad” es crucial para determinar el entorno de prueba más adecuado, un aspecto clave en la selección exitosa de genotipos superiores. La discriminatividad se refiere a la capacidad de un entorno para distinguir entre genotipos, mientras que la representatividad mide qué tan bien un entorno representa a los demás ambientes evaluados (Khan et al., 2021; Oladosu et al., 2017). Este patrón forma una matriz de visualización bidimensional en los ejes X e Y, donde los datos son centrados y descompuestos utilizando valores singulares (SV) para generar puntajes GE en los componentes principales (PC1 y PC2)(Yan & Tinker, 2006). El PC1 refleja el nivel de desempeño, mientras que el PC2 indica la estabilidad de los genotipos (Frutos et al., 2014).

La representación gráfica de estos patrones es posible gracias a las técnicas biplot, ya que el modelo GGE, por sí solo, genera una matriz de datos de doble entrada que incluye las puntuaciones de los genotipos y ambientes, así como los efectos multiplicativos que capturan la interacción genotipo \times ambiente. Sin embargo, para lograr una representación en un plano bidimensional que facilite la exploración de los datos agronómicos, se emplean técnicas biplot específicamente el HJ biplot, nombrado así por Galindo (Galindo, 1985; Galindo-Villardón, 1986), para demostrar la representación simultánea de filas y columnas en un mismo sistema de coordenadas, con una alta calidad de representación tanto para filas como columnas. En este sentido, el modelo GGE se beneficia de los biplots al combinar gráficamente los efectos principales del genotipo y la interacción $G \times E$ en un mismo plano. A través del análisis de componentes principales, se reduce la dimensionalidad de los datos, capturando la mayor parte de la variación en los primeros componentes. Esto facilita la visualización y comprensión de las complejas relaciones entre genotipos y ambientes, permitiendo identificar genotipos superiores y ambientes representativos. De este modo, los biplots se convierten en una herramienta esencial para interpretar y aprovechar la información derivada de los ensayos multi-ambientales en la selección y recomendación de variedades(Esan et al., 2023; Yan et al., 2000, 2001; Yan & Kang, 2002).

1.1. Formulación del problema

¿Es posible optimizar el sector productivo de la caña de azúcar en Ecuador mediante la selección de genotipos basada en un análisis multivariante GGE biplot, que

evalúa el rendimiento y la estabilidad de estas variedades según la zona y el año de producción?

1.1.1. Sistematización:

- ¿Cuál es la eficacia de las representaciones Biplot para explorar y visualizar datos agronómicos complejos?
- ¿Cómo facilita la integración de métodos GGE Biplot en el análisis de datos multi-ambientales la identificación de características productivas y adaptativas específicas de las variedades de caña de azúcar?
- ¿Es posible identificar los genotipos más prometedores de caña de azúcar a través de GGE Biplot, considerando su rendimiento en diversos ambientes y temporadas para su aplicación en el mejoramiento productivo?

1.3. Objetivos de la investigación

1.3.1. Objetivo general

Evaluar la capacidad del modelo GGE Biplot para identificar variedades de caña de azúcar óptimas, analizando su rendimiento y estabilidad en diversos sectores y años de producción, con el fin de optimizar la selección de genotipos en el sector agrícola ecuatoriano.

1.3.2. Objetivos específicos

- Realizar una profunda revisión bibliográfica sobre el estado del arte de la técnica multivariante GGE biplot.
- Explorar las características genotípicas, ambientales e interactivas que poseen las variedades de caña de azúcar que son de gran importancia para el sector productivo ecuatoriano.
- Analizar comparativamente los rendimientos de diversas variedades de caña de azúcar cultivadas en distintos sectores y temporadas, identificando aquellas que ofrezcan los mejores rendimientos bajo un amplio rango de ambientes.

- Proporcionar al sector agrícola recomendaciones estratégicas para la selección dirigida de variedades de caña de azúcar para optimizar la producción agrícola.

1.4. Justificación de la investigación

La falta de una buena selección de variedades de caña de azúcar afecta de manera importante al sector productivo, sobre todo en términos de rendimiento, calidad de la biomasa y sostenibilidad en la producción de bioenergía (Castillo & Silva Cifuentes, 2022). Por ejemplo, los agricultores podrían elegir variedades menos productivas, lo que resultaría en una menor cantidad de subproductos disponibles para otros usos, como la bioenergía. Además, una mala elección puede tener un impacto negativo en el medio ambiente, ya que la calidad del bagazo varía entre las variedades. Algunas variedades de caña producen un bagazo de baja calidad, con menos poder calorífico y bajos niveles de lignina, celulosa y hemicelulosa, o con altos niveles de nitrógeno y azufre. Esto puede reducir la eficiencia energética y aumentar las emisiones de contaminantes durante la combustión (Silva et al., 2020). Finalmente, no seleccionar las mejores variedades también puede reducir la competitividad del sector agrícola en el mercado de la bioenergía. Con el creciente interés en las energías renovables, los productores que no optimicen sus variedades de caña podrían quedarse atrás frente a sus competidores, afectando su rentabilidad y sostenibilidad a largo plazo (Akinwale et al., 2014; Castillo & Silva Cifuentes, 2022).

La caña de azúcar es uno de los cultivos más relevantes a nivel mundial, con un uso predominante en la producción de azúcar y biocombustibles (Cevallos-Molina et al., 2023). Sin embargo, en Ecuador, a pesar de su ubicación en una región tropical, la industria azucarera es relativamente pequeña en comparación con otros países de América Latina. En 2023, la producción total alcanzó los 6,25 millones de toneladas, distribuidas en 75 mil hectáreas de superficie cultivada. Las condiciones climáticas del país, caracterizadas por una radiación solar neta promedio de apenas dos horas diarias y una baja oscilación térmica anual de solo siete horas, no son las más favorables para el desarrollo óptimo de la caña de azúcar. Como consecuencia, se observa una limitada acumulación de sacarosa, lo que genera un alto tonelaje de caña cosechada, pero con un bajo rendimiento en la producción de azúcar (Castillo & Silva Cifuentes, 2022).

El modelo GGE biplot es una herramienta eficaz para analizar datos agronómicos, especialmente útil para visualizar las interacciones $G \times E$. No solo permite identificar a aquellos que superan el rendimiento promedio y mantienen estabilidad en diferentes entornos, sino que también destaca a quienes se desempeñan mejor en condiciones ambientales específicas (Yan, 2002). Ha sido aplicado con éxito en la selección de variedades modelos de girasol (Pourdad & Moghaddam, 2013), papa (Scavo et al., 2023), caña de azúcar (Mehareb et al., 2022) y remolacha azucarera (Dang et al., 2024). Además, su aplicación va más allá de responder a variaciones en ubicaciones o temporadas de siembra; Es igualmente eficaz en diversos estudios, como la respuesta al estrés hídrico (Natarajan et al., 2020), la resistencia a patógenos (Akan et al., 2023) y otros contextos en los que el rendimiento no es el principal indicador de respuesta, como en la activación de genes (Zhao et al., 2023), y el estudio de microorganismos de biocontrol (Valenzuela-Cobos et al., 2023).

1.5.Hipótesis del trabajo

Si los métodos estadísticos multivariantes como el modelo GGE biplot permiten revelar patrones subyacentemente imperceptibles, entonces será posible emplear estas técnicas para evaluar y seleccionar las mejores variedades para mejorar el rendimiento de producción de caña de azúcar ecuatoriana.

CAPÍTULO II
MARCO TEÓRICO

CAPÍTULO II

2. MARCO TEÓRICO

2.1. Interacción Genotipo Ambiente (GEI)

La comprensión de la interacción entre el rendimiento de los cultivos y las condiciones ambientales ha sido históricamente una temática fundamental para fitomejoradores y genetistas. El rendimiento agrícola, observable fenotípicamente, se deriva de la interacción del genotipo (variedad o cultivar) con el entorno y la interacción genotipo-ambiente (GEI, por sus siglas en inglés). Esta interacción se manifiesta cuando distintos cultivares o genotipos exhiben respuestas diferenciadas a variaciones ambientales. La relevancia de la GEI es reconocida por los investigadores principalmente cuando sus efectos son estadísticamente significativos, alterando el orden jerárquico de los genotipos en distintos entornos, es decir, diferentes genotipos prevalecen como superiores en condiciones ambientales variadas (Haldane, 1946). La detección de la GEI requiere, para su validación estadística, que al menos dos genotipos o cultivares sean evaluados en dos o más ambientes distintos. Por tanto, un modelo básico para incorporar la GEI sería:

$$P = G + E + GE$$

Este modelo puede representarse estadísticamente como $P = u + G_i + E_j + (GE)_{ij}$. En esta ecuación, u es la media global. Este enfoque plantea que un genotipo puede manifestar diversos fenotipos dependiendo del entorno y la interacción genotipo-ambiente (GEI).

Aplicar adecuadamente este modelo requiere al menos la presencia de dos genotipos y entornos distintos. Los diversos elementos que componen este modelo se pueden describir como (Simmonds, 1981):

Tabla 1. Modelo de interacción Genotipo-Ambiente (GEI)

Genotipo	Ambiente		Diferencia (efecto Ambiente)
	E1	E2	
G1	a	c	$\Delta 1 = c - a$
G2	b	d	$\Delta 2 = d - b$

Diferencia
(efecto Genotipo)

$$\Delta 3 = b - a$$

$$\Delta 4 = d - c$$

Entonces la interacción Genotipo-Ambiente es:

$$(\Delta 2 - \Delta 1) = (\Delta 4 - \Delta 3) \text{ o } (d - b) - (c - a) = (d - c) - (b - a)$$

o también

$$(\Delta 1 + \Delta 4) = (\Delta 2 + \Delta 3) \text{ o } (c - a) + (d - c) = (d - b) + (b - a)$$

El efecto del genotipo, denotado por $\Delta 3$, ilustra la variación entre los genotipos en el ambiente E1, mientras que $\Delta 4$ indica la variación entre los genotipos en el ambiente E2. Por otro lado, el efecto ambiental, representado por $\Delta 1$, refleja la modificación debido a los ambientes para el genotipo G1 y $\Delta 2$ corresponde a la modificación debido a los ambientes para el genotipo G2.

$$\text{Efecto total (T)} = G + E + GE = (d - a)$$

O también

$$GE = T - G - E$$

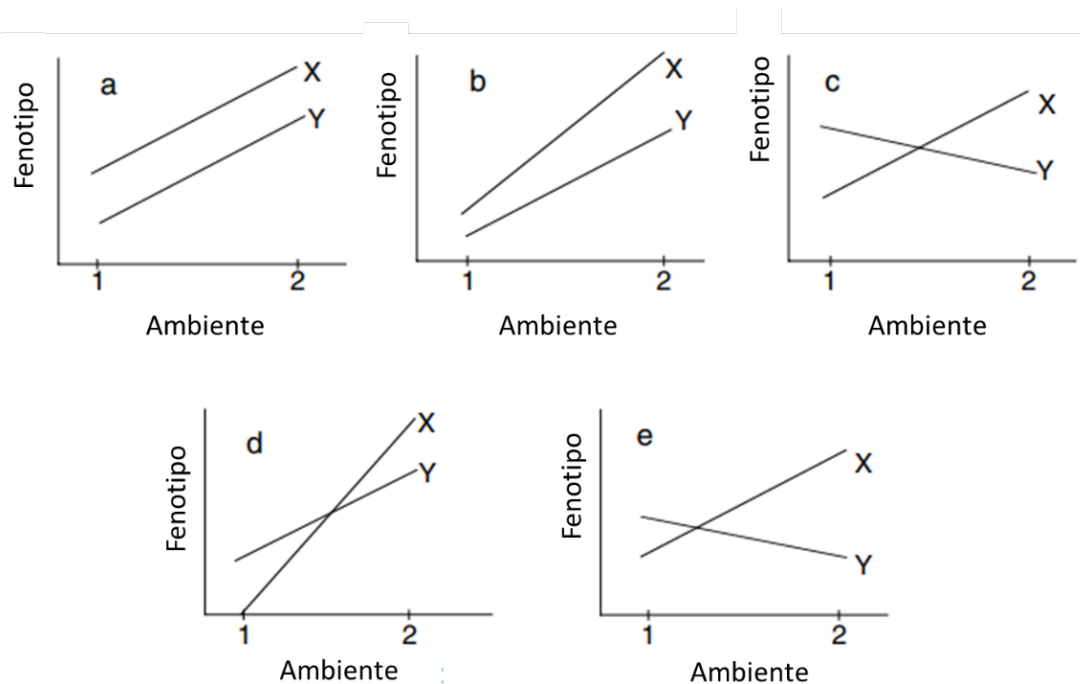


Figura 1. Representación gráfica de los tipos de GEI

Donde (a) una respuesta paralela X e Y en los dos entornos – no hay interacción; (b) crecimiento tanto para X e Y pero sus diferencias Inter-genotípica provocan respuestas desiguales en los dos entornos – interacción no cruzada; (c) Cambios genotípicos que ocurren en direcciones opuestas, manteniendo constante la disparidad entre genotipos – Interacción cruzada; (d) Aunque X e Y se incrementan, la disparidad Inter-genotípica varía – Interacción cruzada; (e) Inconsistencia en la diferencia Inter-genotípica entre los entornos, donde X incrementa e Y disminuye en el segundo entorno.

2.2. Concepto del GGE

Todos los estudios biológicos se encuentran familiarizado con el hecho de que la variación fenotípica (P) observada consiste en variaciones del ambiente (E), del genotipo (G) y su interacción genotipo-ambiente (GEI).

Un ambiente puede definirse como un año de cultivo (Y), una localización (L) o la interacción entre ambos (L×Y). En este contexto, las combinaciones de genotipo y ambiente se clasifican en genotipo-año (G×Y), genotipo-localización (G×L) y su interacción triple (G×Y×L). No obstante, en los ensayos de múltiples ambientes (MET), un solo año no resulta suficiente para realizar comparaciones precisas; por ello, el

concepto de ambiente se limita únicamente a la localización, y la interacción se reduce a $G \times L$.

Una forma más entendible de explicar estas interacciones GE, se realiza con el concepto de interacción cruzada o cambio de rango (Baker, 1990; Hühn, 1996). En este sentido, solo las GEI cruzada, que causen cambios de rango del cultivar en diferentes ambientes, pueden afectar la selección de un genotipo. Para determinar esta GEI cruzada de los cultivos en diferentes ambientes, se debe considerar tanto el G como el GE, esto implica que investigar la interacción GE es más significativa cuando se trabaja conjuntamente con el G, en este sentido se pueden señalar varios puntos:

Ambiente (E) es grande pero no relevante en la evaluación de cultivares, por lo que debe eliminarse de los datos.

Solo G y GE son relevantes para una evaluación significativa de cultivares y deben considerarse simultáneamente al decidir seleccionar un genotipo.

Este concepto se representa como:

$$P - E = G + GE$$

Donde el termino $G + GE$ es forma alargada de GGE.

2.3.El modelo GGE biplot

Las ecuaciones como $P = G + E + GE$ y $P - E = G + GE$, están representadas en términos de componentes de la varianza, sin embargo, cuando estas se presentan como efectos, que tienen la unidad de los valores medidos originalmente, se convierten en:

$$\hat{y}_{ij} = \mu + \alpha_i + \beta_j + \Phi_{ij}$$

O

$$\hat{y}_{ij} - \mu - \beta_j = \alpha_i + \Phi_{ij}$$

Donde:

\hat{y}_{ij} = es el rendimiento esperado del genotipo i en el entorno j ,

μ = es la media general de todas las observaciones,

α_i = el efecto principal del genotipo i ,

β_j = el efecto principal del entorno j ,

Φ_{ij} = la interacción entre el genotipo i y en el entorno j .

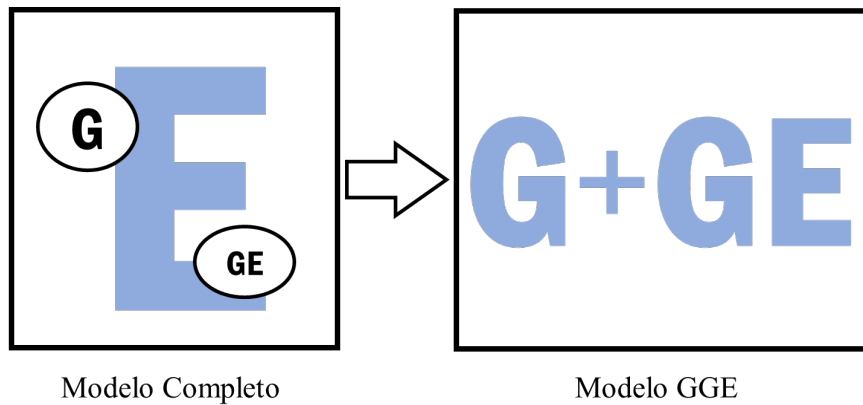


Figura 2. Diagrama comparativo entre el modelo completo y el modelo GGE.

En este sentido, en lugar de separar el efecto independiente G e interactivo GE , un modelo GGE mantiene juntos G y GE y particiona esta mezcla GGE en dos términos multiplicativos.

$$\hat{y}_{ij} - \mu - \beta_j = g_{i1}e_{1j} + g_{i2}e_{2j} + \varepsilon_{ij}$$

Donde:

g_{i1} y e_{1j} = Son las puntuaciones primarias para el genotipo i y el ambiente j , respectivamente.

g_{i2} y e_{2j} = Son las puntuaciones secundarias para el genotipo i y el ambiente j , respectivamente.

ε_{ij} = Son los residuos no explicados por los efectos primarios y secundarios.

Sin embargo, para la representación de un modelo GGE, se puede graficar las puntuaciones g_{i1} contra g_{i2} y e_{1j} contra e_{2j} en un solo diagrama de dispersión, introduciéndonos a las representaciones biplot.

2.4.El concepto de Biplot

Cualquier matriz Y de tamaño $n \times m$ y rango r puede descomponerse en el producto de una matriz G de tamaño $n \times r$ y una matriz R de tamaño $m \times r$, ambas necesariamente de rango r . Esto se expresa como:

$$Y = GH^T.$$

Esta factorización puede escribirse de forma escalar como:

$$y_{ij} = g_i^T h_j,$$

donde y_{ij} es el elemento en la fila i y columna j de Y ; g_i es la fila i -ésima de G y h_j es la fila j -ésima de H . De esta manera, se asigna un vector g_1, \dots, g_n (efectos de fila) a cada una de las n filas de Y , y un vector h_1, \dots, h_m (efectos de columna) a cada columna de Y . Por lo tanto, esta expresión proporciona una representación de Y mediante el producto interno de los vectores correspondientes de efectos de fila y columna.

En el caso de una matriz de rango igual a dos, estos $n + m$ vectores pueden representarse en el plano, ofreciendo una representación gráfica de los elementos de Y a través del producto interno de los vectores de efectos de fila y columna. Esta representación se conoce como biplot, ya que permite visualizar conjuntamente los efectos de filas y columnas (Gabriel, 1971).

Para matrices con rangos superiores a dos, no es posible una representación exacta mediante un biplot. Sin embargo, si una matriz Y puede aproximarse por una matriz de rango dos, denotada como $Y^{(2)}$, los productos internos de los efectos de fila y columna pueden aproximar los elementos de Y .

Para aproximar cualquier matriz Y de tamaño $n \times m$ y rango r por una matriz de rango menor, se puede utilizar la Descomposición en Valores Singulares (SVD). La SVD permite expresar Y como:

$$Y = UKV^T = \sum_{k=1}^r k_k u_k v_k^T$$

donde:

$U = [u_1, \dots, u_r]$ es una matriz $n \times r$ cuyos vectores columna u_k son los vectores singulares izquierdos.

$V = [v_1, \dots, v_r]$ es una matriz $m \times r$ cuyos vectores columna v_k son los vectores singulares izquierdos.

$K = \text{diag}(k_1, \dots, k_r)$ es una matriz $r \times r$ cuyos vectores columna k_k son los vectores singulares izquierdos.

Los valores singulares y vectores singulares se eligen para satisfacer las siguientes condiciones para cada $K = 1, \dots, r$:

$$U^T Y = K V^T,$$

$$Y V = U K,$$

$$Y Y^T u_k = k_k^2 u_k,$$

$$Y^T Y v_k = k_k^2 v_k$$

con $k_1 \geq k_2 \geq \dots \geq k_r$, y donde $u_k^T u_l = v_k^T v_l = \sigma_{kl}$, siendo σ_{kl} la delta de Kronecker.

Utilizando el método de mínimos cuadrados, la expresión anterior se puede aproximar considerando solo los s primeros valores singulares, obteniendo así la mejor aproximación de rango s de Y :

$$Y_{(s)} = U_{(s)} K_{(s)} V_{(s)}^T = \sum_{k=1}^s k_k u_k v_k^T,$$

donde $U_{(s)} K_{(s)} V_{(s)}$ son las matrices formadas por los s primeros vectores y valores singulares. Esta aproximación minimiza el error en el sentido de los mínimos cuadrados y es conocida como la mejor aproximación de rango s de Y (Householder & Young, 1938).

Esta aproximación corresponde a la mejor aproximación de Y con rango igual a s (Householder & Young, 1938). Es decir, la expresión mencionada proporciona una matriz $n \times m$ \times $m \times m$, denotada como $M = (m_{ij})$, de rango s que minimiza la norma cuadrática:

$$\|Y - M\|^2 = \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - m_{ij})^2$$

Dado que $\|Y\|^2 = k_1^2 + \dots + k_r^2$, donde son los valores singulares de Y , la bondad de ajuste del modelo a los datos puede medirse mediante el porcentaje:

$$\text{Bondad de ajuste (\%)} = \left(\frac{\sum_{k=1}^s k_k^2}{\sum_{k=1}^r k_k^2} \right) \times 100$$

Para obtener las matrices de marcadores G y H , es posible utilizar regresiones alternadas. Específicamente:

Si fijamos los marcadores de fila G , los marcadores de columna se pueden calcular mediante regresión lineal como:

$$H^T = (G^T G)^{-1} G^T Y.$$

De manera similar, si fijamos los marcadores de columna H , los marcadores de fila se obtienen como:

$$G^T = (H^T H)^{-1} H^T Y^T.$$

Al alternar iterativamente estos pasos (1) y (2), los productos resultantes convergen hacia la Descomposición en Valores Singulares (SVD). Para más detalles sobre este procedimiento (Ukkelberg & Borgen, 1993).

Para la construcción del biplot, al elegir los factores G y H de $Y^{(2)}$ como en la Ecuación (1), se utiliza la factorización proporcionada por la SVD:

$$G = UK^c \text{ y } H = VK^{1-c}$$

donde:

U y V son las matrices de vectores singulares izquierdo y derecho, respectivamente.

K es la matriz diagonal de valores singulares k_k .

El escalar c es un valor entre 0 y 1 que controla la partición de los valores singulares.

Dependiendo del valor de c , se obtienen diferentes interpretaciones del biplot:

Cuando $c = 1$: Los valores singulares se asignan completamente a los vectores de las filas ($G = UK$), resultando en el biplot que preserva la métrica de las filas o biplot JK (Gabriel, 1971). En este caso, se cumple que $YY^T = GG^T$, lo que permite recuperar las distancias euclidianas entre las filas. Es adecuado para visualizar similitudes y diferencias entre los factores de fila.

Cuando $c = 0$: Los valores singulares se asignan completamente a los vectores de las columnas ($H = VK$), obteniendo el biplot que preserva la métrica de las columnas o biplot GH (Gabriel, 1971). Aquí, ($Y^TY = HH^T$), que corresponde a la matriz de sumas de cuadrados y productos cruzados de Y . Si Y está centrada en las columnas, entonces Y^TY es proporcional a la matriz de covarianza, lo que facilita el estudio de las relaciones entre los factores de columna.

Dos reglas importantes para esta partición son:

Correlación y ángulos: Si los datos se centran en las columnas antes de aplicar la SVD, la correlación entre dos columnas se aproxima mediante el coseno del ángulo entre sus vectores en el biplot.

Longitud de los vectores: La longitud del vector de una columna en el biplot es proporcional a $\sqrt{m - 1}$ veces la desviación estándar de ese factor de columna a lo largo de las filas.

Cuando $c = 0.5$: Se asignan roles simétricos a filas y columnas, obteniendo el biplot simétrico o biplot SQRT (Cárdenas y Galindo, 2003). Aquí, los valores singulares se reparten equitativamente entre G y H , lo que permite analizar simultáneamente las relaciones tanto entre filas como entre columnas.

CAPÍTULO III

**CONTRIBUCIONES A LA PRODUCCIÓN
AGRÍCOLA DE LA CAÑA DE AZÚCAR
DESDE UNA PERSPECTIVA
MULTIVARIANTE: ANÁLISIS DEL
RENDIMIENTO DE VARIEDADES DE CAÑA
DE AZUCAR Y SUS EFECTOS
INTERACTIVOS GENOTIPO-AMBIENTE.**

CAPÍTULO III

3. APLICACIONES EN LA PRODUCCIÓN AGRÍCOLA DE LA CAÑA DE AZÚCAR DESDE UNA PERSPECTIVA MULTIVARIANTE: ANÁLISIS DEL RENDIMIENTO DE VARIEDADES DE CAÑA DE AZÚCAR Y SUS EFECTOS INTERACTIVOS GENOTIPO-AMBIENTE.

Desde el siglo XVI en América Latina la caña de azúcar ha tenido un rol importante en su economía, impactando de forma significativa en el empleo formal y salarios, principalmente en municipios pequeños. Esta importancia radica en los múltiples usos que tiene esta materia prima, incluida la producción de azúcar, etanol, alimentos para el ganado y demás subproductos (Defante et al., 2020; Hall, 2008). Proporciona ganancias relativamente altas, mercados seguros e insumos subsidiados para los agricultores, esto debido a su adaptabilidad a diversos entornos. La caña de azúcar se cultiva en más de 100 países, lo que mejora potencialmente los ingresos y la calidad de vida en las regiones en desarrollo. La tecnificación de esta producción puede incrementar la productividad, reducir costos y mejorar las condiciones de trabajo, así como las operaciones como la labranza y la siembra puede aumentar la productividad general en un 10-15% (Abnave, 2020; Heinrichs et al., 2017).

La caña de azúcar promete con un alto potencial de producción de biomasa para múltiples aplicaciones que sugieren que podría ser una fuente clave de ingresos y mejorar la calidad de vida en muchas regiones, siempre que cuente con las condiciones socioeconómicas adecuadas para su explotación, sin embargo, este proceso presenta desafíos para potencializar la producción como la necesidad de una organización política y gubernamental adecuada para aprovechar al máximo el potencial de la caña de azúcar y en un enfoque más técnico la selección adecuada de variedades (Defante et al., 2020; Heinrichs et al., 2017).

La selección inadecuada de variedades de caña de azúcar presenta desafíos significativos para la industria azucarera, ya que uno de los problemas centrales es lograr un equilibrio óptimo entre el contenido de sacarosa y el rendimiento de la caña; existe una tendencia a favorecer variedades con mayor contenido de sacarosa, lo que puede conducir a una disminución en la productividad general debido a posibles compromisos en otros atributos agronómicos esenciales (Butterfield & Thomas, 1996). Además, la susceptibilidad a enfermedades es un factor crítico que no puede ser ignorado, pues se estima que hasta

el 40% de las variedades prometedoras son rechazadas debido a su vulnerabilidad a enfermedades, lo que representa un costo oculto considerable para la industria.

El proceso de selección es complejo y multifacético, involucrando múltiples etapas y criterios de evaluación, donde características morfológicas como el número de tallos, el diámetro y la erectilidad pueden influir significativamente en el potencial de rendimiento y, por ende, en los resultados de la selección (Lavín-Castañeda et al., 2020). Asimismo, se ha observado que la variación genética en el contenido de azúcar tiende a disminuir a medida que aumentan las medias familiares, lo que plantea la necesidad de implementar estrategias de mejoramiento genético cuidadosas para mantener la diversidad genética y asegurar la sostenibilidad a largo plazo, ya que la reducción en la variabilidad genética puede limitar la capacidad de adaptación de las variedades a diferentes condiciones ambientales y a nuevos patógenos emergentes (Butterfield & Thomas, 1996; Castillo & Silva Cifuentes, 2022). Los estudios han revelado que los criterios de selección de una variedad de caña de azúcar deben ser riguroso y preferirse indicadores como el contenido de azúcar en comparación con el rendimiento de la caña, además que en su defecto los mejoradores consideren múltiples rasgos de forma simultánea (Abnave, 2020; Zhou et al., 2011).

Los datos utilizados en esta investigación fueron recopilados, de dos grandes productores de caña de azúcar ecuatorianos ubicados en la zona de guayas, durante los años del 2018 a 2020.

3.1. Diseño experimental

El diseño experimental de este estudio fue longitudinal, abarcando ensayos de campo reales realizados durante los periodos productivos comprendidos entre 2018 y 2020. Se realizaron mediciones del rendimiento de caña de azúcar anualmente durante la temporada seca, comprendida entre los meses de junio a diciembre, para asegurar condiciones climáticas consistentes. Todas las variedades recibieron los mismos tratamientos agronómicos, incluyendo riego, fertilización y manejo fitosanitario, con el objetivo de minimizar la variación debida a prácticas de cultivo, esto se aseguró tomando como base los métodos de producción establecidos por únicamente dos importantes productores de caña de azúcar en Guayas, Ecuador. Esto permitió que la ubicación geográfica y las condiciones climáticas fueran las principales fuentes de variación en el

rendimiento, eliminando en su medida de lo posible cualquier factor extrínseco que pueda afectar las observaciones (Mehdi et al., 2024).

Al controlar los factores agronómicos y enfocarse en las variables ambientales, se buscó determinar el impacto específico del clima y la ubicación en el rendimiento de cada variedad (Mehdi et al., 2024). Este enfoque facilitó una comparación más precisa entre las variedades, resaltando sus características genéticas y su adaptabilidad a diferentes entornos.

3.2. Material vegetal – Genotipos

En los programas de mejoramiento genético como los de caña de azúcar, es importante la diversidad genética en los monocultivos, esto con el objetivo de desarrollar variedades de alto rendimiento adaptadas a diversos entornos. Esta diversidad puede ser intrapoblacional, referida a la variación genética que existe dentro de una misma población o línea de caña de azúcar y por otro la diversidad interpoblacional referido a la variación genética que existe entre diferentes poblaciones o líneas de caña de azúcar. Algunos estudios han demostrado que existe una variación genética significativa mayor de diversidad intrapoblacional que interpoblacional en las variedades de caña de azúcar (You et al., 2013).

La caña de azúcar es un cultivo genéticamente complejo, ya que es una especie poliploide con un alto nivel de heterocigosidad. Esto significa que posee múltiples copias de cada cromosoma y una gran diversidad genética natural, por ello, gestionar adecuadamente esta diversidad genética es esencial para desarrollar variedades mejoradas que sean sostenibles a largo plazo y optimizar los recursos (de Setta et al., 2014). En este estudio, se utilizó una colección de 10 variedades de caña de azúcar como material vegetal, evaluando su rendimiento en toneladas de material cosechado por hectárea. La colección incluyó dos variedades colombianas (CC-8592 y CC93-4181), una variedad dominicana (CR-74250) y una mezcla de diferentes variedades provenientes de un programa de mejoramiento denominado “CLONES”. Además, se incorporaron seis variedades ecuatorianas: cuatro variedades estándar (ECU-01, EC-02, EC-03 y EC-05) liberadas entre 2007 y 2013, y dos variedades mejoradas (EC-07 y EC-08) liberadas en 2016.

3.3. Ambientes (sectores – años)

La selección de ubicaciones en un estudio multiambiente es fundamental para evaluar la interacción genotipo por ambiente y determinar la adaptabilidad y estabilidad de las variedades en diferentes condiciones. En general, se eligen sitios que representen una variedad de condiciones agroclimáticas y edáficas dentro de la región de interés, lo que permite capturar la variabilidad ambiental y su efecto en el rendimiento de los cultivos.

En el presente estudio, se seleccionaron cinco sectores (01, 02, 03, 04, 05) dentro de la provincia de Guayas para el cultivo de caña de azúcar. Aunque todas las ubicaciones se encuentran dentro de la misma provincia, cada sector presenta diferencias en factores como el tipo de suelo, la altitud, la temperatura y el régimen de precipitaciones. La elección de estos sectores se realizó con el objetivo de representar las principales zonas cañeras de la provincia y capturar la variabilidad ambiental existente dentro de ella.

Al seleccionar estos sectores, se consideraron los siguientes criterios (Podlich et al., 1999):

- **Representatividad agroecológica:** Los sectores fueron elegidos para abarcar diferentes microclimas y tipos de suelo presentes en la provincia de Guayas, asegurando que las condiciones ambientales fueran lo suficientemente diversas para evaluar la interacción genotipo-ambiente.
- **Accesibilidad y logística:** Se seleccionaron ubicaciones que permitieran una gestión eficiente del experimento, facilitando el monitoreo y mantenimiento de las parcelas, así como la recolección de datos.
- **Historial agrícola:** Se tomó en cuenta el historial de cultivo y manejo agrícola de cada sector para minimizar la influencia de factores no controlados, como la presencia de plagas o enfermedades endémicas.

En este apartado también se consideraron a los años como ambientes, esto debido a la definición propuesta por Yan & Hunt, (1998), quienes mencionan al “ambiente” como una combinación única de factores que pueden incluir tanto la ubicación como el año de realización del ensayo. Esto permite que rendimientos obtenidos en diferentes años para el mismo sector se consideren como ambientes distintos dentro del análisis estadístico. Este enfoque facilita la evaluación de la interacción genotipo × ambiente considerando

variaciones temporales, además de espaciales, lo que es fundamental para una interpretación más completa de la estabilidad y adaptabilidad de los genotipos evaluados. En este sentido, las combinaciones entre sectores y años ofrecen un conjunto de 15 ambientes a evaluarse (ver tabla 2).

3.4. Rendimiento (Toneladas/Hectáreas)

La estimación precisa del rendimiento de la caña de azúcar, expresado en toneladas por hectárea (ton/ha), es esencial para una gestión eficiente y sostenible de los cultivos. A lo largo de los años, las metodologías para calcular este rendimiento han experimentado importantes avances. Inicialmente basadas en métodos visuales tradicionales, estas técnicas han evolucionado hacia el uso de teledetección, imágenes satelitales y algoritmos de aprendizaje automático (Huang et al., 2024). Por ejemplo, la integración de imágenes satelitales como Landsat y ASTER con índices de vegetación ha permitido predecir rendimientos con márgenes de error tan bajos como 1-5,65%. Además, el empleo de vehículos aéreos no tripulados (UAV) y variables alométricas ha mostrado resultados prometedores, y los algoritmos de aprendizaje automático han mejorado aún más la precisión de las predicciones (Almeida et al., 2006; Huang et al., 2024). Estas contribuciones no solo ofrecen pronósticos más precisos y oportunos, sino que también respaldan una mejor planificación y gestión de la cosecha, revolucionando la industria de la caña de azúcar (Luciano et al., 2021).

En este estudio para calcular el rendimiento de caña de azúcar en toneladas por hectárea (ton/ha), se siguió un procedimiento que implica medir la cantidad de caña cosechada en un área conocida y luego extrapolar ese valor a una hectárea.

3.4.1. Medición del área de muestreo

Se selecciona una parcela de tamaño conocido dentro del campo. Las parcelas de muestreo suelen ser de dimensiones estándar para facilitar los cálculos. Un tamaño común es de 10 metros de largo por 10 metros de ancho, lo que equivale a 100 metros cuadrados (0.01 hectáreas).

3.4.2. Cosecha de la caña en el área seleccionada

El proceso de cosecha comenzó seleccionando el mejor momento según la madurez y el contenido de azúcar; posteriormente se realiza una quema previa, (práctica

que aún se lleva a cabo hoy en día), para eliminar hojas secas y ahuyentar plagas, facilitando así el corte; posteriormente los agricultores cortan la caña a ras del suelo, luego la deshojan y agrupan en manojos; para ser transportados rápidamente a los ingenios azucareros para su pesaje y procesamiento inmediato, evitando la pérdida de sacarosa por fermentación y combinando así prácticas ancestrales con tecnologías modernas (Amador-Sacoto & Helfgott-Lerner, 2023; Elwakeel et al., 2022).

3.4.3. Pesaje del Material Cosechado

El pesaje se realiza al llegar el camión con carga, primero se pesa en una báscula especializada para obtener el peso bruto (camión+carga), luego descarga la caña y se pesa nuevamente vacío para obtener el peso tara; al restar el peso tara del peso bruto se obtiene el peso neto de la caña entregada, que se registra y documenta para asegurar una medición precisa y transparente, fundamental para el pago al proveedor y el control de producción en el ingenio(Asobanca [Asociación de Bancos Privados del Ecuador], 2022).

3.4.4. Cálculo del rendimiento

El rendimiento se calculó usando la siguiente ecuación:

$$Rendimiento = \frac{\text{Peso de la caña cosechada (Toneladas)}}{\text{Área de la parcela (ha)}}$$

3.5. Conjunto de datos

El conjunto de datos utilizado en este estudio comprende un total de 2,555 registros, los cuales fueron obtenidos a través de la recolección sistemática de información sobre los 10 genotipos seleccionados para los ensayos agrícolas. Estos genotipos fueron distribuidos de manera equitativa en cada uno de los 5 sectores productivos, asegurando una representación balanceada y homogénea a lo largo de toda la zona de estudio. Los sectores productivos abarcan distintas áreas dentro de la región de Guayas, cada una con características agroclimáticas particulares.

La recopilación de datos se llevó a cabo durante tres periodos consecutivos, comprendiendo los años 2018, 2019 y 2020. Esta continuidad temporal permitió evaluar el comportamiento de los genotipos bajo condiciones ambientales variables y detectar posibles tendencias o anomalías en el rendimiento a lo largo del tiempo. Cada sector productivo fue monitoreado de manera intensiva, registrando principalmente el rendimiento en toneladas por hectárea (ton/ha).

Debido a la estructura del diseño experimental, que incluye 5 sectores (ubicaciones) y 3 años consecutivos de recolección de datos, se generan un total de 15 ambientes distintos para la evaluación. Cada combinación sector- año representa un ambiente único que permite analizar la interacción entre los genotipos y las condiciones específicas de cada ambiente. La siguiente tabla ilustra la distribución de los ambientes evaluados, detallando la combinación de sectores productivos y años de recolección de datos:

Tabla 2. Características de los ambientes experimentales.

Sectores	Nombre del Cañicultor	Año	Hectareas (Ha)	Material total cosechado	Rendimiento medio	Codigo del Ambiente
01	Ingenio	2018	15,88	1141,15	71,85	A
		2019	16,54	1300,73	78,62	B
		2020	16,15	1268,12	78,53	C
02	Ingenio and Isabel maria	2018	15,23	1130,96	73,43	D
		2019	16,39	1407,28	85,93	E
		2020	16,42	1398,54	85,22	F
03	Ingenio	2018	17,36	1312,59	75,6	G
		2019	17,33	1666,7	96,19	H
		2020	18,18	1773,76	97,58	I
04	Ingenio	2018	17,08	1059,5	62,03	J
		2019	16,76	1246,57	74,39	K
		2020	16,48	1244,68	75,55	L
05	Ingenio	2018	16,25	1048,75	64,55	M
		2019	16,17	1265,4	78,25	N
		2020	16,63	1309,51	78,77	O

3.6. Exploración de las fuentes de variación

A continuación, se presentando dos diagramas de caja y bigotes, para identificar de forma visual las tendencias centrales y de dispersión de las fuentes variación, genotipos (ver figura 3) y ambientes (ver figura 4).

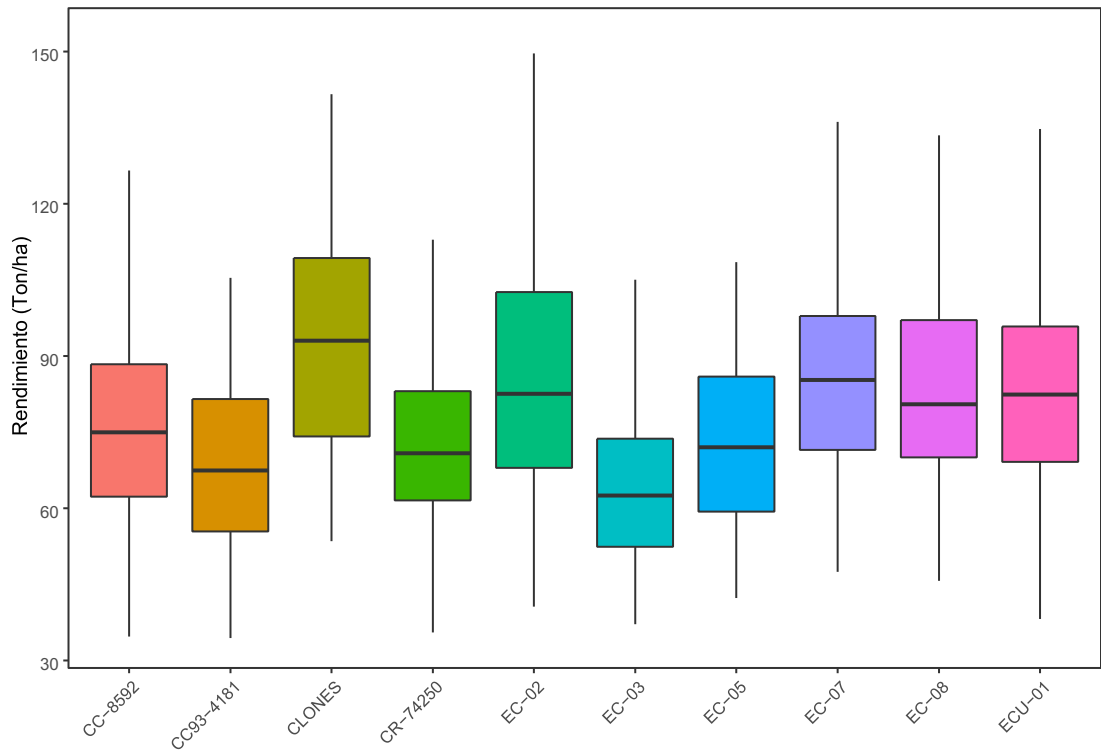


Figura 3. Diagrama de caja y bigote de diferentes genotipos.

El gráfico de caja y bigotes nos informa que, entre todas las variedades de caña de azúcar, destaca CLONES, con un mayor rendimiento (más de 90 Ton/Ha), seguido de las variedades EC-07 y ECU-01 respectivamente. Por otro lado, el diagrama muestra a la variedad EC-02 como la de mayor dispersión, posiblemente por los ambientes involucrados en el rendimiento.

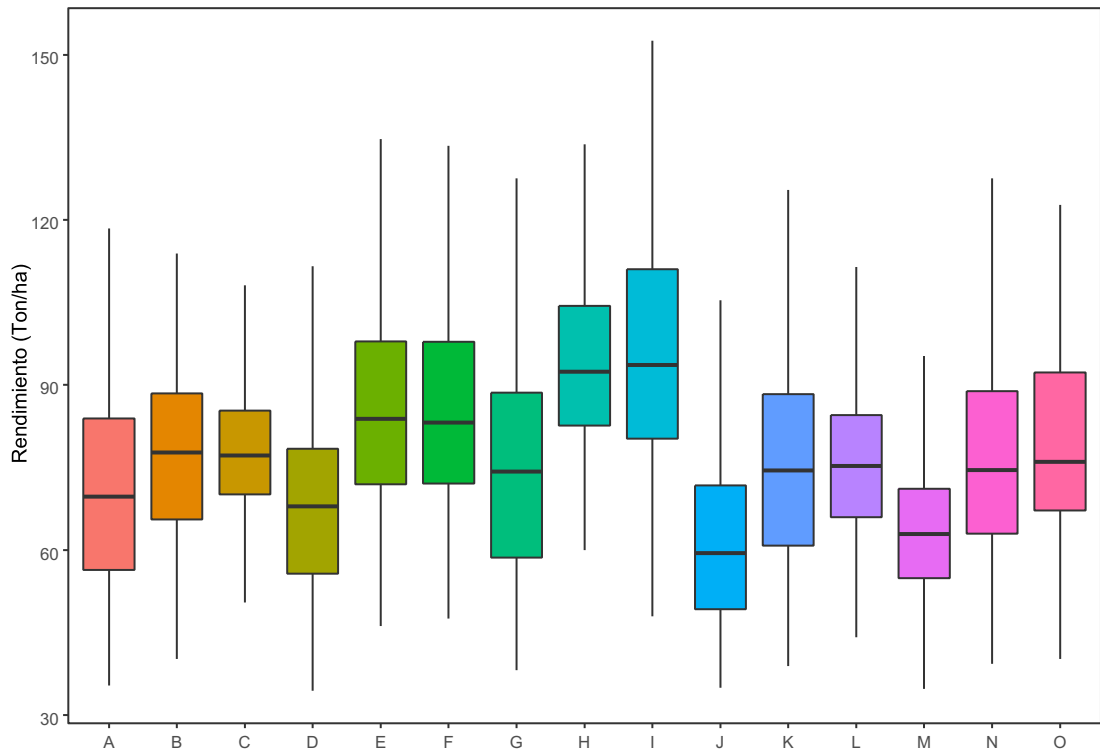


Figura 4. Diagrama de caja y bigote de diferentes ambientes.

Por otro lado, el gráfico de caja y bigote basado en los ambientes (sectores × años) muestran que el ambiente “I” destaca con el mayor rendimiento (más de 90 Ton/Ha), sin embargo, también destaca por tener la mayor dispersión de los datos. Este ambiente corresponde al sector 03 durante la cosecha del año 2020. Por otro lado, el ambiente “J” destaca con el menor rendimiento obtenido (menos de 60 Ton/Ha), este ambiente corresponde al sector 04 durante la cosecha del 2018.

Teniendo en cuenta que existen relaciones interactivas complejas entre los genotipos y ambientes (sectores × años) de forma sinérgica, se espera que empleando las representaciones GGE biplot se revelen patrones subyacentes, tal y como muestra la tabla x. donde se revelan que la combinación de los tres años de estudio con los cinco sectores seleccionados resultó en rendimientos diferentes para cada una de las 10 variedades

Tabla 3. Rendimiento medio del monocultivo en una tabla de tres entradas.

Variedades	Sector 01			Sector 02			Sector 03			Sector 04			Sector 05		
	2018	2019	2020	2018	2019	2020	2018	2019	2020	2018	2019	2020	2018	2019	2020
CC-8592	70,45	69,22	72,71	57,85	63,34	77,1	83,68	92,09	72,42	79,14	79,52	84,64	96,63	75,83	78,51
CC93-4181	66,97	60,62	66,69	71,91	59,03	65,25	78,71	91,52	73,66	54,41	64,28	69,98	78,74	67,14	72,01
CLONES	63,69	88,24	111,68	77,85	102,28	72,99	97	107,61	72,95	88,94	59,08	101,45	116,37	72,81	75,53

CR-74250	70,03	64,76	74,85	65,75	61,79	72,06	81,56	95,09	78,37	72,76	64,98	74,68	88,36	76,24	70,86
EC-02	70,95	67,81	91,03	53,39	60,01	86,95	88,09	103,06	81,79	86,95	81,75	91,55	106,18	82,44	85,62
EC-03	58,14	56,02	64,13	60,57	62,72	59,09	71,96	87,97	59,94	67,88	67,9	69,28	89,98	48,56	64,67
EC-05	86,66	60,73	74,19	57,68	61,76	72,56	77,64	83,56	81,93	71,47	72,17	70,92	84,35	49,74	50,85
EC-07	87,77	47,94	97,28	75,77	72,03	83,32	99,32	109,72	81,68	82,09	83,82	93,21	106,31	74,91	85,74
EC-08	77,63	78,32	84,42	75,72	75,14	78,92	82,3	102,06	78	77,05	85,9	90,51	99,36	73,17	79,82
ECU-01	77,42	70,06	73,36	67,94	71,13	84,28	92,24	95,3	84,83	85,79	75,58	87,62	91,37	80,87	82,89

3.7. Análisis de la varianza (ANOVA) y descomposición por sectores

Antes de emplear directamente el modelo GGE biplot, es necesario verificar que las fuentes de variación de nuestros datos agronómicos realmente presenten diferencias estadísticamente significativas entre grupos de variedades, años y sectores. Para ello se empleará el análisis de varianza (ANOVA) de dos factores, ajustando un modelo interactivo para cada una de las fuentes de variaciones, todo esto apoyado del entorno de trabajo R Studio junto con el paquete “agricolae” (“Agricolae: Statistical Procedures for Agricultural Research,” 2006).

Empleando el software Rstudio, se cargarán los datos. Para importar los datos en formato Excel se usó la librería **openxlsx** usando la función **read.xlsx**, por otro lado, para el análisis de varianza se usó la librería **agricolae** y la función **aov** para establecer un modelo que tome como referencia cuantitativa el rendimiento de caña de azúcar, tal que:

```
anova<-aov(Yield ~ Genotype * Environment, data=df)
```

Donde **Yield** corresponde al rendimiento reportado en toneladas por hectáreas, **Genotype** son las variedades tomadas como factores y finalmente **Environment** son los 15 ambientes (A hasta la O) tomados igualmente como factores y el símbolo “*” representa la interacción entre ambas fuentes de variación.

- Con la función **anova** obtenemos un resumen de los principales resultados del análisis de varianza, arrojando los siguientes resultados:
- **Df** (Grados de Libertad): Representa el número de valores independientes que pueden variar en el cálculo de una estadística. Indica la complejidad del modelo para cada fuente de variabilidad.

- **Sum Sq** (Suma de Cuadrados): Mide la variabilidad total atribuible a cada fuente (genotipo, año_sector, interacción) y a los Residuals (errores). Infiere que cuanto mayor es la suma de cuadrados, mayor es la contribución de esa fuente a la variabilidad total.
- **Mean Sq** (Cuadrados Medios): Es la suma de cuadrados dividida por sus respectivos grados de libertad. Este valor proporciona una medida de la variabilidad promedio por grado de libertad.
- **F value** (Valor F): Es la razón entre el cuadrado medio del factor y el cuadrado medio de los residuals. Un valor F alto sugiere que la fuente de variabilidad tiene un efecto significativo en el rendimiento.
- **Pr(>F)** (Valor p): Es la probabilidad de obtener un valor F igual o más extremo que el observado, bajo la hipótesis nula. Indica la significancia estadística de cada fuente de variabilidad. Un valor p bajo (usualmente < 0.05) sugiere que el efecto es significativo.

Una vez obtenido los resultados del ANOVA, podemos verificar los grados de libertad (df), la proporción de varianza explicada por cada fuente de variación (SC%) y la significancia de este factor por su efecto independiente e interactivo.

Tabla 4. Análisis de Varianza para efectos principales e interactivos.

Fuente de Variación	Grados de libertad	Suma de cuadrados	Media de cuadrados	SC%	Valor F
Genotipos	9	82664	9185	7.96	30.65***
Sectores	4	77022	19256	7.42	64.26***
Años	2	102203	51101	9.84	170.53***
G × Sector	36	19543	543	1.88	1.81**
G × Años	18	10569	587	1.02	1.96**
Sector × Year	8	11363	1420	1.10	4.74***
G × Sector × Year	72	14212	197	1.37	0.66(ns)
Residuales	2405	720674	300	69.41	-

ns, no significante; *, $p < .05$; **, $p < .01$; ***, $p < .001$

Efecto principal de los genotipos: La fuente de variación correspondiente a los genotipos presenta 9 grados de libertad (Df) y una suma de cuadrados (Sum Sq) de 82,664,

lo que resulta en una media de cuadrados (Mean Sq) de 9,185. Esto representa el 7.96% de la variabilidad total (SC%). El valor F obtenido es de 30.65, altamente significativo (***) , con un valor p menor a 0.001. Este resultado indica que existen diferencias significativas en el rendimiento entre los distintos genotipos evaluados. En otras palabras, la elección del genotipo tiene un impacto notable en la productividad de la caña de azúcar, sugiriendo que algunos genotipos son superiores a otros en términos de rendimiento.

Efecto principal de los Sectores: La variación atribuida a los sectores muestra 4 grados de libertad, una suma de cuadrados de 77,022 y una media de cuadrados de 19,256, representando el 7.42% de la variabilidad total. El valor F es de 64.26, también altamente significativo (***) , con un valor p menor a 0.001. Este hallazgo revela que las diferencias entre los distintos sectores productivos influyen significativamente en el rendimiento de la caña de azúcar. Las variaciones en las condiciones agroclimáticas, prácticas de manejo y otras características específicas de cada sector contribuyen de manera considerable a la variabilidad observada en los rendimientos.

Efecto principal de los Años: La fuente de variación correspondiente a los años cuenta con un valor F es de 170.53, extremadamente significativo (***) , es decir un valor p menor a 0.001. Este resultado indica que los tres años en que se realizaron las evaluaciones tiene un impacto muy significativo en el rendimiento de la caña de azúcar. Las variaciones climáticas anuales, como precipitaciones, temperaturas y otros factores ambientales, afectan de manera sustancial la productividad de los cultivos.

Efectos interactivos: El efecto interactivo **Genotipo**×**Sector** es significativo con un valor $p < 0.01$, sugiriendo que el efecto de los genotipos sobre el rendimiento varía dependiendo del sector productivo. Es decir, ciertos genotipos pueden desempeñarse mejor en algunos sectores que en otros (de momento no se puede saber esto), lo que refleja una interacción significativa entre estos dos factores, aunque el efecto es relativamente pequeño.

Por otro lado, la interacción **Sector**×**Año**, es altamente significativa, con un valor $p < 0.001$, indicando que la combinación de sector y año influye significativamente en el rendimiento de la caña de azúcar. Las condiciones específicas que resultan de la interacción entre el sector productivo y el año, como microclimas o prácticas de manejo particularizadas, afectan de manera notable la productividad.

La triple interacción **Genotipo**×**Sector**×**Año**, es no significativa con un valor p de 0.66, esto indica que no hay evidencia suficiente para afirmar que la interacción conjunta de genotipo, sector y año tiene un efecto significativo en el rendimiento de la caña de azúcar. En otras palabras, no se detecta una variación consistente en el rendimiento que dependa simultáneamente de estos tres factores. No obstante, para evaluar a mayor detalle los efectos triples, se puede optar por la descomposición por sectores, tal y como muestra la tabla x.

Residuales: Finalmente la variación residual presenta 2,405 grados de libertad, una suma de cuadrados de 720,674 y una media de cuadrados de 300, representando el 69.41% de la variabilidad total. Esta parte de la variación corresponde a la variabilidad no explicada por el modelo, es decir, por los factores y sus interacciones considerados en el análisis. La alta proporción de variabilidad residual sugiere que existen otros factores no incluidos en el modelo que están influyendo en el rendimiento de la caña de azúcar, o que hay una considerable variabilidad inherente al proceso que no puede ser capturada por los factores estudiados.

Tabla 5. Suma Cuadrática de diferentes ANOVAS

Ubicación	Genotipos	G SS%	Años	A SS%	G × Años	G × A SS%	Residuos	R SS%
Sector 01	12562***	9,42	3225**	2,41	4102(ns)	3,08	113475	85,08
Sector 02	28260***	10,17	39638***	14,27	4910(ns)	1,77	205058	73,8
Sector 03	26614***	10,75	36735***	14,84	4046(ns)	1,63	180116	72,77
Sector 04	10265***	7,29	16017***	11,38	7614*	5,41	106797	75,91
Sector 05	14688***	9,84	14670***	9,83	4610(ns)	3,08	115227	77,23

Nota. Los valores indican la suma cuadrática, significancia y porcentaje de varianza; ns, no significante; *, p < .05; **, p < .01; ***, p < .001

Posterior al análisis de varianza (ANOVA) global, se realizó este mismo análisis para cada uno de los cinco sectores productivos (Sector 01 a Sector 05) de la zona Guayas; donde se desglosaron las fuentes de variación en **Genotipos (G)**, **Años (A)**, **Interacción**

Genotipo × Años (G × A) y proporción de los Residuos%. Al examinar los resultados, se observó que únicamente en el Sector 04 la interacción entre genotipo y año resultó ser significativa ($p < 0.05$), mientras que en los demás sectores esta interacción no mostró significancia estadística (ns). El objetivo de descomponer el ANOVA por sector es comprender cómo los genotipos y los años afectan el rendimiento de la caña de azúcar en cada sector específico. Además, se busca identificar si la relación entre genotipo y rendimiento varía a lo largo de los diferentes años dentro de cada sector, evaluándose implícitamente una triple interacción Genotipo × Sector × Años (**G × S × A**).

En este sentido, en los sectores 01, 02, 03 y 05 la interacción **G × A** no fue significativa (ns), esto indica que el efecto de los genotipos sobre el rendimiento es estable a lo largo de los diferentes años evaluados, a priori los genotipos que muestran un mejor rendimiento tienden a mantener su superioridad independientemente del año, lo que sugiere una baja variabilidad en la respuesta de los genotipos frente a las condiciones anuales. En contraste, en el Sector 04, la interacción Genotipo × Año (**G × A**) resultó ser significativa ($p < 0.05$). Este hallazgo sugiere que el efecto de los genotipos sobre el rendimiento varía significativamente entre los años evaluados. Es decir, ciertos genotipos pueden rendir mejor en un año específico y no en otro, indicando una variabilidad en la respuesta de los genotipos frente a las condiciones anuales particulares de este sector. Desde un punto de vista estadístico, esta interacción significativa refleja una interacción Genotipo × Ambiente (GEI) fuerte (Yan & Kang, 2002).

3.7.1. Análisis de las fuentes de variación

Ahora calculamos mediante la prueba post hoc de Tukey HSD, para conocer que tan diferentes son los grupos de genotipos y ambientes.

Tras determinar con el ANOVA que si existen diferencias significativas entre grupos y una leve pero significativa interacción Genotipo × Ambiente, evaluamos las diferencias significativas entre los grupos, para ello se emplea la prueba post hoc Tukey Honestly Significant Difference (HSD) con un nivel de significancia $\alpha = 0.05$. Este método permite realizar comparaciones múltiples entre los pares de medias de los genotipos, controlando la tasa de error familiar (family-wise error rate) para evitar la inflación del error Tipo I. El cálculo de la diferencia significativa de Tukey HSD se basa en la siguiente fórmula:

$$HSD = q_{\alpha} \times \sqrt{\frac{MS_{Residuales}}{n}}$$

donde q_{α} es el valor crítico de la distribución de rango estudiantil para el nivel de significancia elegido, $MS_{Residuales}$ es la media de cuadrados residual obtenida del ANOVA, y n es el número de observaciones por grupo. En este sentido, esta prueba post hoc, compara todas las posibles parejas de medias de genotipos, y aquellas diferencias que superan el valor calculado de HSD se consideran estadísticamente significativas. Este enfoque permite identificar específicamente qué genotipos, sector o año difieren entre sí en términos de rendimiento, proporcionando una comprensión detallada de las variaciones observadas en el estudio (Tukey, 1949).

Los resultados encontrados de la prueba son los siguientes:

Se calcula las diferencias estadísticas por medio de la función **HSD.test**, usando de base el estadístico de prueba “anova” planteado anteriormente tal que:

```
HSD.test(anova, "referencia", group = T, console = TRUE)
```

Donde “referencia” es cualquier factor de entrada (categoría), en este caso se plantearon tres referencias, las variedades de caña, los sectores y los años, obteniendo la siguiente tabla:

Tabla 6. Prueba Post Hoc de Tukey HSD

Fuente de variación	Rendimiento±Error	Desviación Estandar	Rankin basado la media total	Rank by GGE
Genotipos				
CC-8592	76,54 ± 0,511 ^c	19,63	6	4
CC93-4181	68,67 ± 1,845 ^{d,e}	17,58	9	6
CLONES	92,64 ± 2,334 ^a	22,4	1	5
CR-74250	72,52 ± 1,065 ^d	16,13	7	5
EC-02	85,27 ± 1,301 ^{a,b}	23,83	2	2
EC-03	63,98 ± 1,636 ^e	15,06	10	7
EC-05	72,63 ± 2,026 ^{c,d}	16,13	8	6

EC-07	86,1 ± 1,912 ^{a,b}	20,18	3	1
EC-08	83,29 ± 1,697 ^b	19,47	4	3
ECU-01	82,95 ± 0,813 ^b	19,72	5	3
Sectores				
sector 01	75,17 ± 0,813 ^c	17,17	3	3
sector 02	79,84 ± 0,647 ^b	19,71	2	2
sector 03	88,44 ± 0,799 ^a	22,99	1	1
sector 04	70,77 ± 0,81 ^{c,d}	17,56	5	3
sector 05	73,68 ± 0,807 ^d	18,03	4	2
Años				
2018	68,45 ± 0,58 ^b	18,42	3	3
2019	82,31 ± 0,58 ^a	18,95	2	2
2020	83,61 ± 0,624 ^a	19,52	1	1

Note. Valores indican el rendimiento medio ± Error estándar, Letras distintas dentro de una columna indican diferencias significativas según la prueba HSD de Tukey ($p < 0,05$).

La tabla proporciona información para identificar a los genotipos, sectores y años “ganadores” con el rendimiento más alto. En este sentido, se revela que el genotipo **EC-03**, de origen ecuatoriano, presentó el rendimiento medio más bajo (63,98 ton/ha) y fue significativamente inferior a todas las demás variedades, lo que indica una menor adaptabilidad o productividad en comparación con los otros genotipos. Le sigue el genotipo dominicano **CC93-4181** con un rendimiento medio de 68,67 ton/ha, también significativamente menor que la mayoría de las variedades. En contraste, el grupo denominado **CLONES** mostró el rendimiento más alto (92,64 ton/ha), superando significativamente a todas las demás variedades evaluadas. Las variedades **EC-07** (86,1 ton/ha) y **EC-02** (85,27 ton/ha) también exhibieron rendimientos elevados y no presentaron diferencias significativas entre sí, aunque sí diferían significativamente de las variedades **EC-08** (83,29 ton/ha) y **ECU-01** (82,95 ton/ha). Estas diferencias subrayan la importancia de la selección de genotipos específicos para optimizar la productividad de la caña de azúcar.

En cuanto a los sectores de estudio, se observaron diferencias significativas en el rendimiento de la caña de azúcar. El **Sector 03** destacó con el mayor rendimiento medio

(88,44 ton/ha), siendo significativamente superior a los demás sectores. Los Sectores **02** (79,84 ton/ha) y **01** (75,17 ton/ha) presentaron rendimientos moderados, aunque el Sector 02 fue notablemente superior al Sector 01, con un rendimiento aproximadamente un 6% mayor. Por otro lado, los Sectores **04** (70,77 ton/ha) y **05** (73,68 ton/ha) registraron los rendimientos más bajos, con hasta un 8,3% menos rendimiento que el Sector 03.

Finalmente, al analizar los periodos de producción, se identificaron diferencias significativas en el rendimiento entre los años evaluados. El año 2018 registró el rendimiento más bajo (68,45 ton/ha), significativamente inferior a los años 2019 (82,31 ton/ha) y 2020 (83,61 ton/ha). Sin embargo, no se observaron diferencias significativas entre los rendimientos de 2019 y 2020, sugiriendo una mejora en la productividad a lo largo del tiempo que se estabilizó entre estos dos años.

3.8. Análisis de los patrones biplot

Antes de obtener los patrones gge biplot, se debe plantear el modelo GGE (Genotipo + Genotipo × Ambiente). En este estudio se empleó la librería `metan` en R (Olivoto & Lúcio, 2020), se inicia asegurando que el conjunto de datos contenga las columnas necesarias: genotipos (variedades de caña de azúcar), ambientes (sectores × años) y al menos una variable de respuesta, en este caso el rendimiento de caña en toneladas métricas por hectárea. Es fundamental que estos datos estén correctamente estructurados, donde cada fila represente una observación única de un genotipo en un ambiente específico. Una vez preparado el dataset, se procede a instalar y cargar la librería `metan` mediante los comandos `install.packages("metan")` y `library(metan)`.

A continuación, se utiliza la función `gge()` para ajustar el modelo GGE. Esta función requiere especificar el dataset, así como los nombres de las columnas que corresponden a los ambientes (`env`), los genotipos (`gen`) y la variable de respuesta (`resp`). Además, aunque menos estricto se deben definir parámetros adicionales como el método de centrado (`centering`), el método de escalado (`scaling`) y la partición de valores singulares (`svp`). En este sentido, el modelo GGE empleado en este estudio fue el siguiente:

```
Gge_mod <- (.data = df, env = Environment, gen = Genotype,
  resp = Yield, centering = "environment", svp = "environment")
```

Donde:

- **Df** Son los ensayos agronómicos reales
- **Environment** Es la categoría de los ambientes organizada en una combinación sectores \times años que van alfabéticamente de la A a la O.
- **Genotype** es la categoría dentro del dataset que contiene a las 10 variedades de caña de azúcar estudiadas
- **Yield** es el rendimiento de las cañas de azúcar medido en toneladas métricas por hectárea.
- **"environment"** son los valores que pueden tomar tanto el centrado con el método svp:
 - En el caso del centrado, debe ser uno de los siguientes: 'none | 0', para ningún centrado; 'global | 1', para centrado global (E+G+GE); 'environment | 2' (por defecto), para centrado en el entorno (G+GE); o 'double | 3', para doble centrado (GE). No se puede producir un biplot con modelos producidos sin centrado (Olivoto & Lúcio, 2020).
 - Para la partición del valor singular. Debe ser uno de los siguientes: 'genotype | 1', (El valor singular se particiona completamente en los vectores propios del genotipo, también llamado preservación de la métrica de filas); 'environment | 2', por defecto, (El valor singular se particiona completamente en los vectores propios del entorno, también llamado preservación de la métrica de columnas); o 'simétrico | 3' (El valor singular se divide simétricamente en los vectores propios del genotipo y del entorno. Este SVP se utiliza con mayor frecuencia en el análisis AMMI y otros análisis de biplots, pero no es ideal para visualizar la relación entre genotipos o entre entornos) (Olivoto & Lúcio, 2020).

3.8.1. Representación GGE biplot

Una vez ajustado el modelo GGE, se procede a generar los diagramas GGE biplots utilizando la función `plot()` de la librería `metan` en R. Para este propósito, se emplea el siguiente código:

```
plot(gge_model, repel = TRUE, col.gen = "blue", col.env = "#8B6508", size.text.env = 3.5, size.text.gen = 4, size.line = 0.6, plot_theme = theme_metan(grid = "both"), type = 1, col.alpha = 1, col.circle = "#6B8E23", col.alpha.circle = 1.5, axis_expand = 1.2, title = TRUE)
```

En este análisis, **repel = TRUE** se utiliza para evitar la superposición de etiquetas en el gráfico, mejorando la legibilidad de los nombres de genotipos y ambientes. Los argumentos **col.gen = "blue"** y **col.env = "#8B6508"** asignan colores específicos a los genotipos y ambientes, respectivamente, facilitando la diferenciación visual entre ambos elementos en el biplot. Las opciones **size.text.env = 3.5** y **size.text.gen = 4** ajustan el tamaño del texto de las etiquetas de ambientes y genotipos, asegurando una presentación clara y proporcional. El parámetro **size.line = 0.6** controla el grosor de las líneas que representan las relaciones entre genotipos y ambientes, mientras que **plot_theme = theme_metan(grid = "both")** aplica un tema gráfico personalizado que incluye una cuadrícula completa, lo que ayuda a interpretar mejor las posiciones relativas en el biplot.

Finalmente, el argumento más importante es **type =**, ya que arroja los resultados para cada patrón gge biplot, tomando valores del 1 al 10, siendo estos (Olivoto & Lúcio, 2020; Yan & Kang, 2002):

1. Biplot básico.
2. Rendimiento medio frente a estabilidad (gge biplots) o la vista de coordinación del probador medio para biplots genotipo-rasgo y genotipo-rendimiento*rasgo.
3. Cuál-gana-dónde.
4. Discriminatividad frente a representatividad.
5. Examinar un entorno (o combinación de rasgo/rendimiento*rasgo).
6. Clasificación de entornos (o combinación de rasgo/rendimiento*rasgo).
7. Examinar un genotipo.
8. Clasificación de genotipos.

9. Comparación de dos genotipos.

10. Relación entre entornos (o combinación rasgo/rendimiento*rasgo).

3.8.1.1. Representación GGE biplot

El análisis GGE biplot presentado en la Figura 1 empleó el método de descomposición en valores singulares (SVD) con escala simétrica basada en la desviación estándar, logrando explicar el 74,77% de la varianza total. Este modelo evalúa la interacción entre genotipo, sector y año ($G \times S \times A$), proporcionando una visión detallada de cómo estos factores influyen conjuntamente en el rendimiento de la caña de azúcar (Mehareb et al., 2022).

En el biplot, la proximidad de los genotipos dentro de un ambiente específico indica que dichas variedades pueden responder de manera más favorable en esos ambientes particulares (Akan et al., 2023). Observamos que los sectores F, G e I se encuentran más alejados del origen, lo que sugiere que estos ambientes poseen una mayor capacidad discriminadora, permitiendo distinguir con mayor claridad las diferencias entre genotipos. Sin embargo, esta alta discriminación viene acompañada de una menor representatividad, lo que implica inestabilidad en su desempeño. Por otro lado, los sectores J y K se localizan más cerca del origen, lo que indica que son ambientes más representativos y estables, aunque con una capacidad discriminadora relativamente débil.

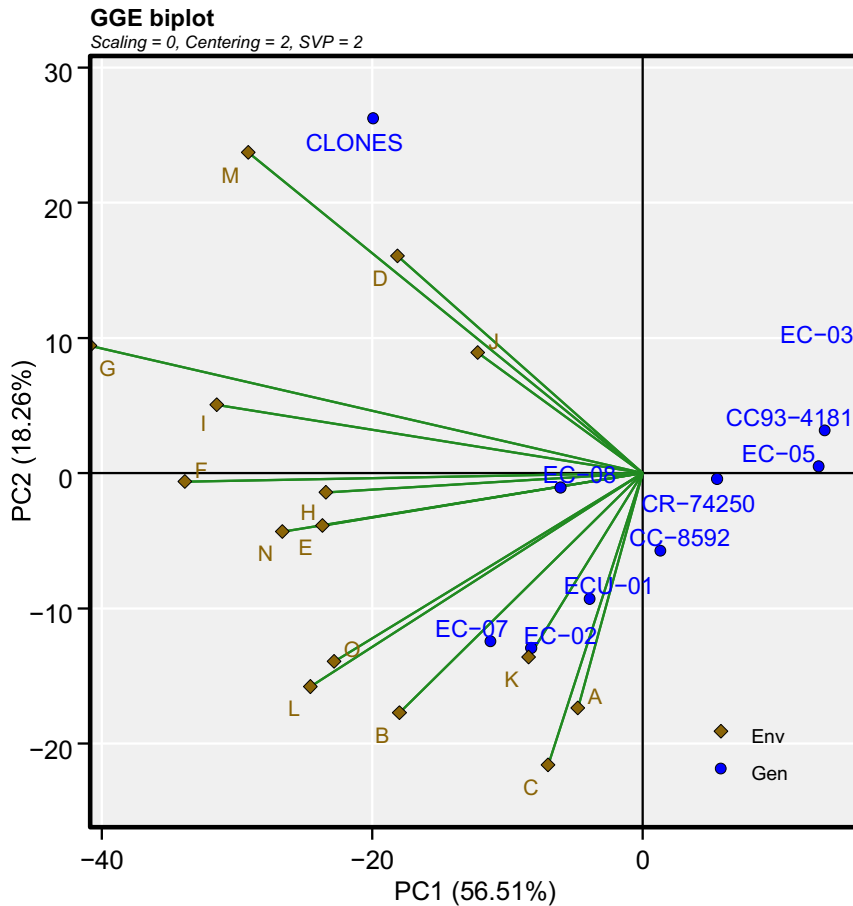


Figura 5. Ilustración GGE biplot básica.

Este estudio evidencia variaciones genotípicas significativas en los rendimientos de la caña de azúcar evaluados, destacando notables diferencias en el material cosechado por hectárea entre las distintas variedades. La interacción más pronunciada entre genotipo y ambiente ($G \times A$) se origina principalmente por las diferencias entre sectores. En Ecuador, diversas empresas líderes en la producción de azúcar gestionan extensas áreas de cultivo, lo que facilita el acceso a una amplia variedad de ambientes para realizar pruebas. Esta diversidad ambiental se refleja en la considerable variación observada, atribuida tanto a los efectos independientes de genotipo y ambiente como a sus interacciones $G \times A$. Dichas interacciones capturan el patrón y la magnitud de las respuestas genotípicas en distintos ambientes de producción dentro del país, asegurando que el presente estudio se realice en condiciones óptimas para evaluar adecuadamente estas dinámicas.

3.8.1.2. Representación Quién-Gano-Donde (Which-Won-Where)

Este patrón “cuál-ganó-dónde”, se encuentra basado en la correlación entre genotipos y ambientes (Yan & Hunt, 1998). Tal como se presenta en la figura, se muestra el patrón de los datos de prueba en múltiples ambientes, donde el eje x PC1 y el eje y PC2 explican el 56,51% y el 18,26%, respectivamente, de la variación total (74,77%). Los genotipos más alejados del origen en varias direcciones formaron un polígono que abarcó todos los demás genotipos, lo que indica que eran las mejores variedades en sus respectivos sectores. El polígono se dividió en cuatro sectores cuando se trazó una línea perpendicular desde el origen a cada lado del polígono, dividiendo el biplot en sectores por año. Los 15 ambientes se agruparon en diferentes clústeres (ambiente A y C como un grupo, ambientes B, K, L como otro grupo, ambientes E, H, I como un tercer grupo y, finalmente, ambientes F, G y D, J como grupos independientes), y luego se identificaron los genotipos adecuados para cada grupo.

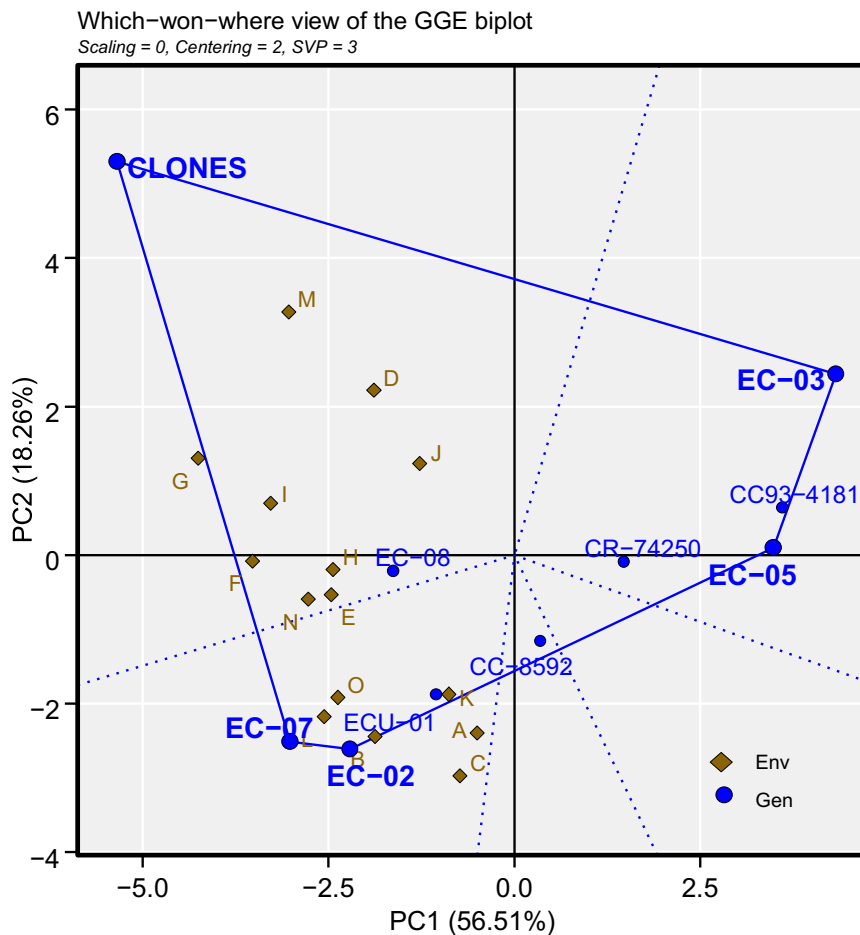


Figura 6. Patrón gge biplot (Which-won-where)

El genotipo ganador, que tiene el mayor rendimiento en todos los ambientes que dividen el sector, se ubica en el vértice. Según (Yan & Tinker, 2006) los genotipos ubicados en los vértices del polígono han tenido el mejor o peor desempeño en uno o más ambientes. Existen cinco sectores con los genotipos “CLONES”, “EC-02”, “EC-03”, “EC-05” y “EC-07” como genotipos de esquina o vértice. El análisis mostró que el genotipo CLONES tuvo el mayor rendimiento en los ambientes J y D, mientras que el genotipo EC-07 tuvo el mayor rendimiento en el ambiente B. Por otro lado, el genotipo EC-08 mostró mayores rendimientos en ambientes más diversos como E, H, I y en menor medida, ambientes L, K, F. Este patrón sugiere que el ambiente objetivo puede consistir en dos megaambientes diferentes y que se deben seleccionar y desplegar diferentes cultivares para cada uno.

3.8.1.3. Representación Media vs Estabilidad (Mean vs. Stability)

El patrón “media vs. estabilidad”, se encuentra medido según el genotipo (Khan et al., 2021; Yan & Hunt, 1998). Este patrón selecciona y evalúa el genotipo ideal dentro de un único megaambiente. En este contexto, los genotipos deben evaluarse tanto por su rendimiento promedio como por su estabilidad en diferentes ambientes. Tal como muestra la Figura, esta representa la coordinación ambiental promedio (AEC) del biplot GGE (Yan et al., 2000, 2001; Yan & Tinker, 2006).

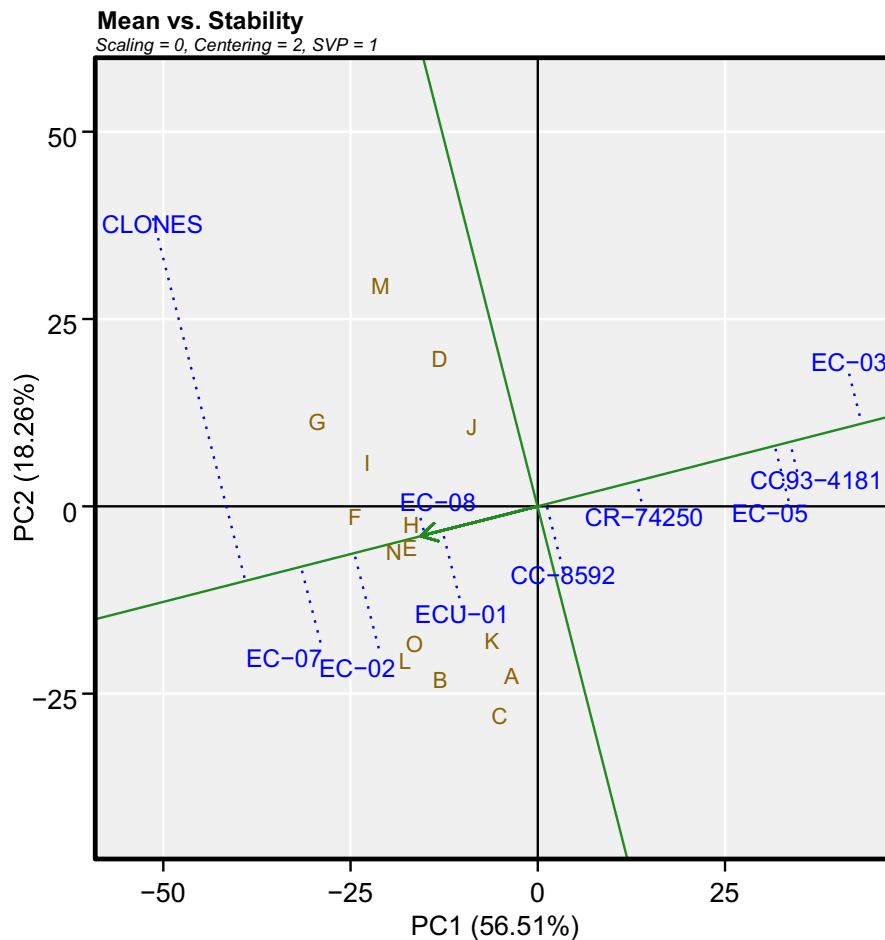


Figura 7. Patron gge biplot Mean vs. Stability

El gráfico muestra una línea con una sola flecha, la cual sirve como abscisa del AEC y se conoce como eje del ambiente promedio (AEA) que apunta al genotipo con el mayor rendimiento. En este sentido, el genotipo “CLONES” muestra tener el rendimiento promedio más alto, seguido de “EC-07”, “EC-02”, etc., mientras que “EC-03” tuvo el rendimiento promedio más bajo. No obstante, otra parte de la representación es la ordenada AEC que pasa por el origen del gráfico y es perpendicular a la abscisa AEC, lo que indica una mayor variabilidad (peor estabilidad) en cualquier dirección. Por lo tanto, “CLONES” fue muy inestable, mientras que “CR-74250” junto con “EC-08” fueron muy estables.

3.8.1.4. Representación del Ranking y Discriminación vs. Representatividad

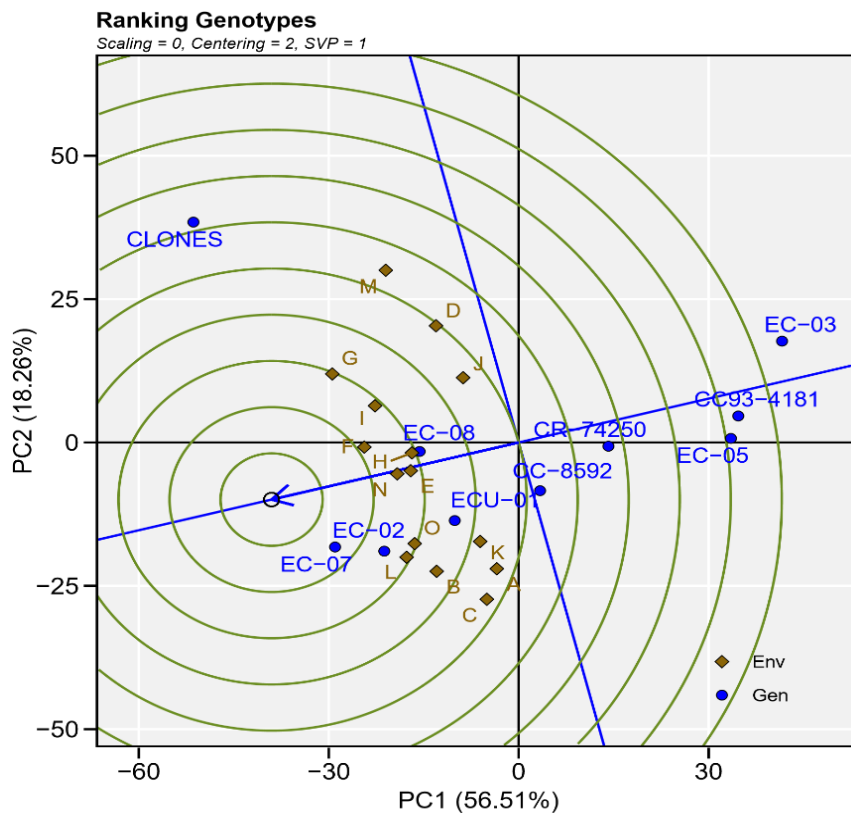
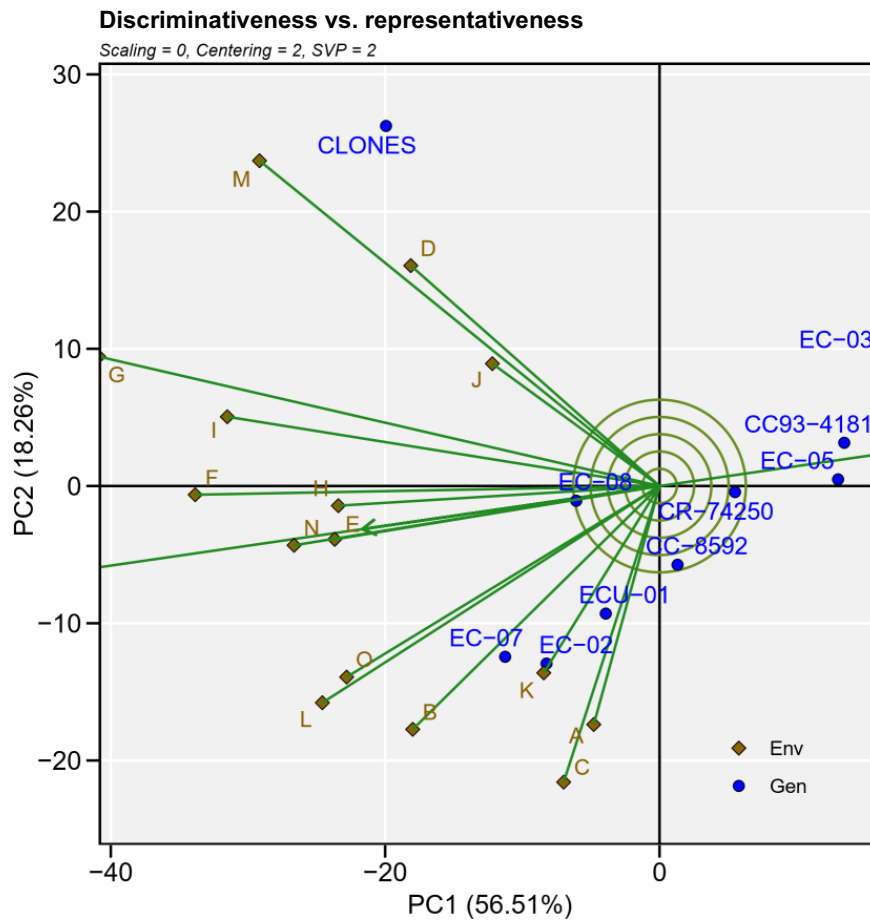


Figura 8. Patrones gge biplot Ranking y Discriminativeness vs. Representativeness.

Determinar el ambiente de prueba más adecuado, es decir, el ambiente ideal, es fundamental para el éxito en los programas de mejoramiento genético orientados a la selección de genotipos superiores. Este proceso es esencial porque permite identificar con precisión cuáles genotipos presentan el mejor desempeño bajo ciertas condiciones ambientales, lo que a su vez conduce a avances significativos en la productividad y calidad de los cultivos (Khan et al., 2021).

Para ello, el patrón biplot “Discriminatividad vs. Representatividad” definen y grafican esta idealidad de los ambientes de prueba. Por un lado, la Discriminatividad se refiere a la capacidad de un ambiente para distinguir entre diferentes genotipos, es decir, un Determinar el ambiente de prueba más adecuado, es decir, el ambiente ideal, es fundamental para el éxito en los programas de mejoramiento genético orientados a la selección de genotipos superiores. Este proceso es esencial porque permite identificar con precisión cuáles genotipos presentan el mejor desempeño bajo ciertas condiciones ambientales, lo que a su vez conduce a avances significativos en la productividad y calidad de los cultivos. Por otro lado, la representatividad es la capacidad de un ambiente para reflejar o representar el comportamiento promedio de todos los demás ambientes evaluados. Un ambiente representativo proporciona información que es aplicable a una amplia gama de condiciones ambientales, lo que es especialmente importante cuando el objetivo es desarrollar genotipos que sean exitosos en múltiples entornos o en condiciones variables (Esan et al., 2023; Khan et al., 2021).

En el contexto del patrón de discriminatividad vs. representatividad en el GGE biplot, los círculos concéntricos juegan un papel importante al facilitar la visualización de las longitudes de los vectores ambientales. Estas longitudes son proporcionales a la desviación estándar dentro de cada ambiente y sirven como una medida del poder discriminativo de los mismos (Yan & Tinker, 2006). Un vector ambiental más largo indica que el ambiente tiene una mayor capacidad para distinguir entre los genotipos evaluados debido a una mayor variabilidad en los resultados. Analizando los quince ambientes estudiados, se observó que el ambiente G, correspondiente al sector 03 en el año 2018, presentó la mayor longitud de vector. Esto significa que fue el ambiente más discriminativo o informativo, ya que pudo diferenciar con mayor claridad entre los distintos genotipos. Así mismo, el ambiente promedio, representado por una flecha, tiene

las coordenadas medias de todos los ambientes de prueba. En este caso, la AEA es la línea que conecta el ambiente promedio con el origen del biplot. Un ambiente de prueba con un ángulo más pequeño respecto a la AEA es más representativo de los demás ambientes de prueba (Frutos et al., 2014); por lo tanto, los ambientes N, E y H, correspondientes a los sectores 02, 03 y 05 durante el periodo 2019, son los más representativos, mientras que los sectores M y D son los menos representativos.

Mediante el uso del biplot de ranking de genotipos (ver figura), podemos identificar un genotipo ideal en comparación con los otros genotipos evaluados. Por lo general, un genotipo ideal se sitúa en el círculo más interno y está relativamente cerca de la punta de la flecha en el centro del anillo circular. El genotipo ubicado dentro de este círculo interno es altamente deseable en comparación con los genotipos que se encuentran en los círculos externos (Khan et al., 2021). Aunque en algunos casos, puede no superponerse dentro del círculo interno; por lo que se deben considerar aquellas variedades cercanas a este círculo. Un anillo en la punta de la flecha en el eje horizontal del AEC generalmente representa a un genotipo ideal, y la idealidad de un genotipo se indica con un pequeño círculo en la línea de abscisas del AEC (Khan et al., 2021; Yan & Tinker, 2006). En este estudio en particular, dentro del Ranking GGE, el genotipo EC-07 se posiciona como el más ideal, seguido por EC-02 y EC-08.

CAPÍTULO IV
CONTRIBUCIONES CIENTIFICAS

CAPÍTULO IV

4. CONTRIBUCIONES CIENTÍFICAS

1.1.Relevancia de los efectos genotípicos, ambientales y genotipos-ambientales

Un análisis de varianza (ANOVA) de dos vías se empleó para evaluar el rendimiento de 10 variedades de caña de azúcar cultivadas durante tres períodos en cinco sectores diferentes. Los resultados demostraron que existe variación genotípica en respuesta al ambiente (tabla 3), con una alta significancia tanto en los efectos independientes G, S y Y ($p < .001$), como con los efectos interactivos GxS ($p < .01$), GxY ($p < .01$) y SxY ($p < .001$). Es decir, el rendimiento de la caña de azúcar se veía afectado según el periodo de cultivo, el sector y la variedad de caña empleada. Sin embargo, la triple interacción entre genotipo, sector y año (G x S x Y) no fue estadísticamente significativa, sugiriendo que una parte de las interacciones no influye de manera decisiva en el rendimiento bajo las condiciones de estudio.

Tabla 3 Two-way anova on the yield of 10 sugarcane varieties grown in 3 periods in 5 different sectors.

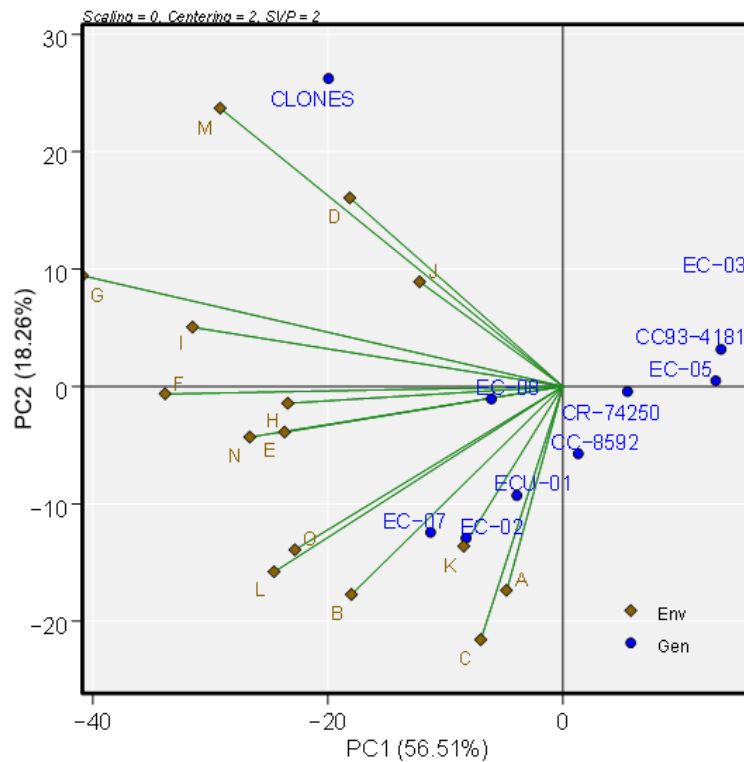
Source	Df	Sum of square	Mean of Square	SS%	F value
G	9	82664	9185	7.96	30.65***
Sector	4	77022	19256	7.42	64.26***
Year	2	102203	51101	9.84	170.53***
G x Sector	36	19543	543	1.88	1.81**
G x Year	18	10569	587	1.02	1.96**
Sector x Year	8	11363	1420	1.10	4.74***
G x Sector x Year	72	14212	197	1.37	0.66(ns)
Residuals	2405	720674	300	69.41	-

*ns, not significant; *, $p < .05$; **, $p < .01$; ***, $p < .001$*

El rendimiento de caña demuestra que contiene en su mayoría patrones de variación influenciado por factores que no están presentes en el modelo o debido a la variabilidad natural que se encuentra inherente en el cultivo de caña de azúcar (69,41%). Por otro lado, la parte explicada por el modelo (30,59%) posee que todas sus interacciones son significativas, donde la interacción (G x S) que plantea las diferencias de rendimiento entre variedades y sectores es la que posee el mayor poder explicativo entre los grupos.

El análisis GGE biplot mostrado en la figura 1, empleó el método de descomposición de valores singulares (SVD) con escalonado simétrico (desviación estándar) y tiene un poder explicativo del 74,77% de la varianza total que evalúa la interacción G x S x Y (Mehareb et al., 2022).

Figura 1 GGe biplot



Nota. **A:** Sector 1-Año 1; **B:** Sector 1-Año 2; **C:** Sector 1-Año 3; **D:** Sector 2-Año 1; **E:** Sector 2-Año 2; **F:** Sector 2-Año 3; **G:** Sector 3-Año 1; **H:** Sector 3-Año 2; **I:** Sector 3-Año 3; **J:** Sector 4-Año 1; **K:** Sector 4-Año 2; **L:** Sector 4-Año 3; **M:** Sector 5-Año 1; **N:** Sector 5-Año 2; **O:** Sector 5-Año 3.

La proximidad de los genotipos en un entorno específico indica que las variedades de caña seleccionadas podrían tener mejor respuesta en un ambiente correspondiente (Akan et al., 2023). Los sectores F, G, I están más alejados del origen, lo que sugiere que los cuatro entornos tenían mejor capacidad de discriminación, pero no eran representativos (inestables). Por el contrario, los sectores J, K estaban más cerca del origen, lo que sugiere que eran los más representativos (estable) pero con una capacidad discriminativa relativamente débil.

El presente estudio revela que existen variaciones genotípicas significativas entre los rendimientos de caña de azúcar evaluados, destacando diferencias en la materia cosechada por hectárea. La interacción más evidente entre genotipo y ambiente (G x E) proviene del efecto de los sectores. En Ecuador, varias empresas líderes en la producción azucarera administran extensas áreas de cultivo, lo que facilita el acceso a una diversidad de ambientes para pruebas, esto se refleja en la variación considerable observada debido a efectos independientes y en las interacciones GxE. Estas últimas probablemente capturan el patrón y la magnitud de las interacciones GxE en diferentes entornos de producción en el país, por lo que este estudio se lleva a cabo bajo condiciones óptimas para evaluar adecuadamente estas dinámicas.

1.2. Efecto por localización en los componentes del rendimiento

La capacidad de explicación del rendimiento de los cultivos se determinó mediante la combinación de sus tres componentes, tomando como referencia las localidades y los 5 sectores seleccionados juntos sus 10 genotipos como columnas. La tabla 4 presenta la descomposición de las sumas de cuadrados del rendimiento de la caña de azúcar, desglosada por componentes de Genotipos (G), Año (Y), y su interacción (G x Y) en los cinco sectores de cultivo.

La descomposición de los sectores aumentó los niveles de varianza para cada uno de los efectos incluyendo la interacción, sin embargo, solo los efectos independientes de los genotipos y años tuvieron varianzas estadísticamente significativas con valores entre 7 a 10% y 2 a 14% respectivamente, lo que reafirma la dependencia de una respuesta sensible a los diferentes genotipos y ambientes en los cultivos de caña.

Tabla 4 sum of squares of cane yield by sector and year

Location	Genotypes	G SS%	Year	Y SS%	G x Year	GxY SS%	Residuals	R SS%
Sector 01	12562***	9,42	3225**	2,41	4102(ns)	3,08	113475	85,08
Sector 02	28260***	10,17	39638***	14,27	4910(ns)	1,77	205058	73,8
Sector 03	26614***	10,75	36735***	14,84	4046(ns)	1,63	180116	72,77
Sector 04	10265***	7,29	16017***	11,38	7614*	5,41	106797	75,91
Sector 05	14688***	9,84	14670***	9,83	4610(ns)	3,08	115227	77,23

*ns, not significant; *, $p < .05$; **, $p < .01$; ***, $p < .001$*

En el sector 01 de cultivo posee su menor poder explicativo en el año de cultivo, demostrando que las variaciones anuales en este sector tienen un menor impacto (pero significativo) en el rendimiento que otros factores, sugiriendo que las condiciones climáticas, prácticas de cultivo o factores económicos que varían de un año a otro no cambian sustancialmente el rendimiento en esta localidad. Por otro lado, el sector 03 es la localidad que mejor explica la variabilidad de los datos, abarcando en promedio un 5,24% más de varianza en sus componentes, concentrándose esta en el año de cultivo, infiriendo que las variaciones anuales por el contrario tienen un impacto mayor en el rendimiento y reflejando la dependencia o sensibilidad a factores externos que varían con el tiempo, como precipitaciones, temperaturas o disponibilidad de recursos.

Finalmente, los sectores 01, 02, 03, y 05 no tuvieron interacciones G x Y significativas. En contraste, el sector 04 la interacción G x Y mostró una significancia moderada ($p < .05$) y se le atribuye gran proporción de la varianza observada el rendimiento de caña para este factor. Esto indica que el rendimiento de los distintos genotipos varió significativamente a lo largo de los diferentes periodos en esta localidad, lo que demuestra la importancia de considerar la selección de genotipos en función de las características específicas de cada localidad y del periodo de cultivo.

1.3.Comparación de las fuentes de variación

La tabla 5 analiza y compara las fuentes de variación sobre el rendimiento en caña de azúcar, esta muestra los valores medios \pm SE del rendimiento agrupados mediante la prueba de Tukey HSD a, b ($\alpha = 0,05$). Entre los genotipos, la variedad ecuatoriana EC-03 fue significativamente menor a las otras variedades con el valor medio más bajo en rendimiento, seguido de la variedad dominicana CC93-4181. Por el contrario, el grupo “CLONES” presento el valor más alto, con un rendimiento de 92,64, superando significativamente a las demás variedades, seguido de las variedades EC-07 y EC-02 las cuales no presentaron diferencias significativas entre sí, pero si con las variedades EC-08 y ECU-01.

En cuanto a los sectores de estudio, se observo que el sector 03 destacó significativamente con el mayor rendimiento, seguido de los sectores 01 y 02 con rendimientos moderados, sin embargo, el sector 02 fue notablemente superior al 01, con un 6% mas de rendimiento. Finalmente, los sectores 04 y 05 registraron los rendimientos más bajos con hasta un 8,3% menos de rendimiento que el sector 03. Finalmente, al comparar los períodos de producción, se observa que el año 2018 registró el rendimiento más bajo, con menos de 70 toneladas por hectárea, seguido del año 2020 que registra rendimiento más alto; no obstante, este no difirió significativamente a los valores observados en 2019.

Los ambientes con mayor rendimiento promedio

Tabla 5 Yield and Classification of Genotypes, Sectors and Years by Tukey's Test and GGE Model.

Source of Variation	Yield	SD	Rank by grand mean	Rank by GGE
Genotypes				
CC-8592	76,54 \pm 0,511 ^c	19,63	6	4
CC93-4181	68,67 \pm 1,845 ^{d,e}	17,58	9	6
CLONES	92,64 \pm 2,334 ^a	22,4	1	5
CR-74250	72,52 \pm 1,065 ^d	16,13	7	5
EC-02	85,27 \pm 1,301 ^{a,b}	23,83	2	2
EC-03	63,98 \pm 1,636 ^e	15,06	10	7
EC-05	72,63 \pm 2,026 ^{c,d}	16,13	8	6
EC-07	86,1 \pm 1,912 ^{a,b}	20,18	3	1
EC-08	83,29 \pm 1,697 ^b	19,47	4	3
ECU-01	82,95 \pm 0,813 ^b	19,72	5	3
Sector				
sector 01	75,17 \pm 0,813 ^c	17,17	3	3
sector 02	79,84 \pm 0,647 ^b	19,71	2	2
sector 03	88,44 \pm 0,799 ^a	22,99	1	1
sector 04	70,77 \pm 0,81 ^{c,d}	17,56	5	3
sector 05	73,68 \pm 0,807 ^d	18,03	4	2
Years				
2018	68,45 \pm 0,58 ^b	18,42	3	3
2019	82,31 \pm 0,58 ^a	18,95	2	2

2020	$83,61 \pm 0,624^a$	19,52	1	1
-------------	---------------------	-------	---	---

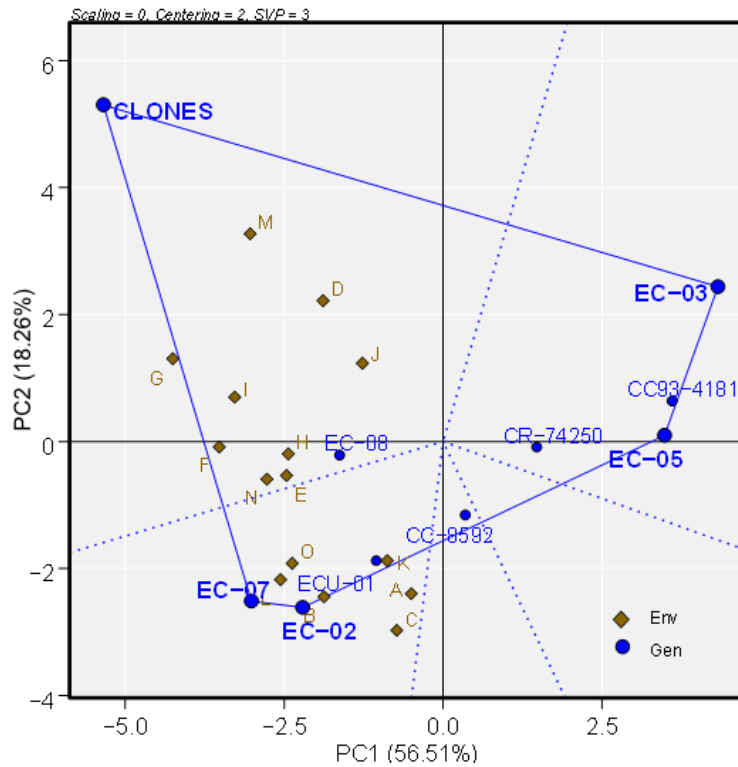
Note. Values indicate the yield mean \pm SE (standard error); Different letters within a column indicate significant differences according to the Tukey HSD test ($p < 0.05$).

1.4. Patrones biplot para análisis multivariante

El GGe biplot (“Which-won-where”) de la figura 2, muestra el patrón de los datos de prueba en múltiples ambientes, donde el eje x PC1 y eje y PC2 explica el 57.99% y 18.86%, respectivamente, de la variación total (76.85%). Los genotipos más alejados del origen en varias direcciones formaron un polígono que abarcaba a todos los demás genotipos, mostrando que fueron las mejores variedades en sus propios sectores. El polígono se dividió en 4 sectores cuando se dibujó una perpendicular desde el origen a cada lado del polígono, dividiendo al biplot en sectores por año. Los 15 ambientes se dividieron en diferentes grupos (ambiente A y C como un grupo, los ambientes B, K, L, como otro grupo, los ambientes E, H, I como un tercer grupo, finalmente los ambientes F, G y D, J como grupos independientes), luego se identificaron los genotipos adecuados para cada grupo.

El genotipo ganador el cual tiene el mayor rendimiento en todos los ambientes que dividen al sector es el que se encuentra en el vértice. Según (Yan & Tinker, 2006) dice que los genotipos ubicados en los vértices del polígono han tenido el mejor o el peor rendimiento en uno o más ambientes. Hay cinco sectores con los genotipos “CLONES”, “EC-02”, “EC-03”, “EC-05” y “EC-07” como genotipos de esquina o vértice. El análisis mostró que el genotipo CLONES tenía el mayor rendimiento en los entornos J y D mientras el genotipo EC-07 tiene un mayor rendimiento en el entorno B. Por otro lado, el genotipo EC-08 mostró un mayor rendimiento en más ambientes diferentes como el E, H, I y en menor medida los ambientes L, K, F. Este patrón sugiere que el entorno objetivo puede consistir en dos mega-entornos diferentes y que deberían seleccionarse y desplegarse diferentes cultivares para cada uno de ellos.

Figura 2 “Which won where”



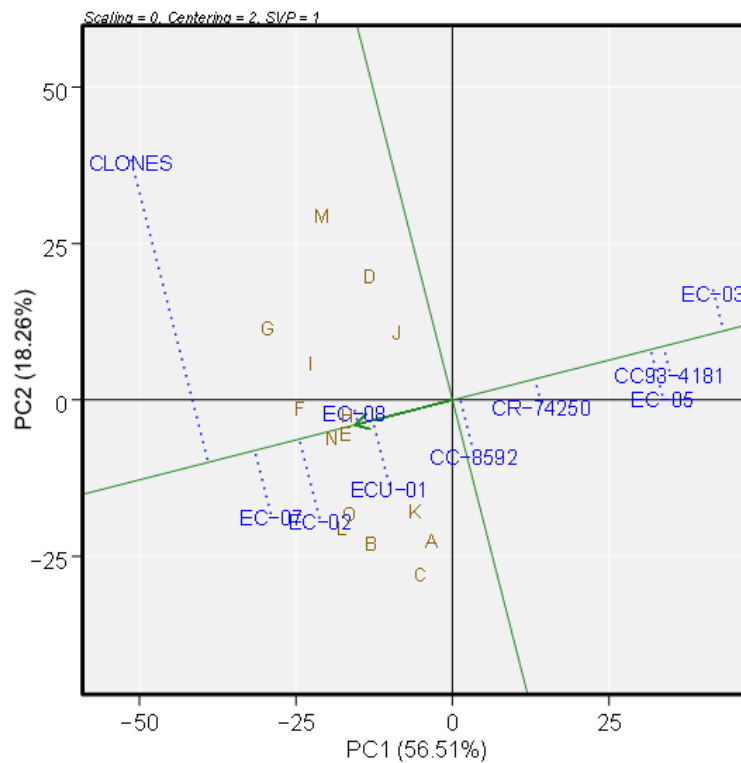
Nota. A: Sector 1-Año 1; **B:** Sector 1-Año 2; **C:** Sector 1-Año 3; **D:** Sector 2-Año 1; **E:** Sector 2-Año 2; **F:** Sector 2-Año 3; **G:** Sector 3-Año 1; **H:** Sector 3-Año 2; **I:** Sector 3-Año 3; **J:** Sector 4-Año 1; **K:** Sector 4-Año 2; **L:** Sector 4-Año 3; **M:** Sector 5-Año 1; **N:** Sector 5-Año 2; **O:** Sector 5-Año 3.

Por otro lado, el patrón GGE biplot (“Media frente Estabilidad”), selecciona y evalúa el genotipo ideal dentro de un mismo mega ambiente. En este sentido, los genotipos deben ser evaluados tanto por su rendimiento medio como por su estabilidad en los distintos entornos, La Figura 3 muestra la coordinación del entorno promedio (AEC) del GGE Biplot que tiene la siguiente interpretación (Yan & Tinker, 2006):

La línea de una sola flecha es la abscisa AEC (o AEA) y apunta a un rendimiento medio más alto en todos los entornos, por lo tanto, “CLONES” tuvo el rendimiento medio más alto, seguido de “EC-07”, EC-02”, etc., mientras que “EC-03” tuvo el rendimiento medio más bajo.

La ordenada AEC pasa por el origen del gráfico y es perpendicular a la abscisa AEC y apunta a una mayor variabilidad (peor estabilidad) en cualquier dirección, por lo tanto, “CLONES” era muy inestable, mientras que “CC-8592” junto con “EC-03” era muy estables.

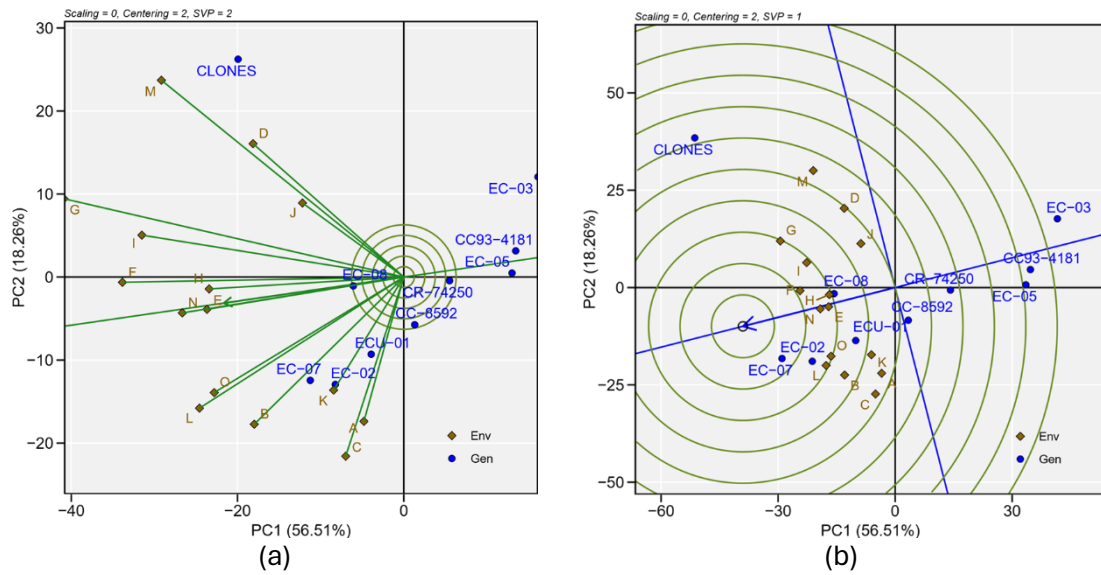
Figura 3 Mean vs Stability



Nota. **A:** Sector 1-Año 1; **B:** Sector 1-Año 2; **C:** Sector 1-Año 3; **D:** Sector 2-Año 1; **E:** Sector 2-Año 2; **F:** Sector 2-Año 3; **G:** Sector 3-Año 1; **H:** Sector 3-Año 2; **I:** Sector 3-Año 3; **J:** Sector 4-Año 1; **K:** Sector 4-Año 2; **L:** Sector 4-Año 3; **M:** Sector 5-Año 1; **N:** Sector 5-Año 2; **O:** Sector 5-Año 3.

Finalmente, el patrón GGE biplot (“Discriminación frente a representatividad” y “Ranking by GGE”) muestra el poder discriminante y representatividad de los ambientes de prueba, tal como se muestra en la figura 4, donde los círculos concéntricos ayudan a visualizar la longitud de los vectores ambientales, que es proporcional a la desviación estándar dentro de los respectivos entornos y es una medida del poder de discriminación de los entornos (Yan & Tinker, n.d.). Por lo tanto, entre los cinco ambientes estudiados, sector 03 fue el más discriminatorio (informativo) y sector 04 fue el menos discriminante. El entorno promedio (representado por una flecha) tiene las coordenadas promedio de todos los entornos de prueba. AEA es la línea que pasa por el medio ambiente y el origen del biplot. Un entorno de prueba que un ángulo más pequeño con el AEA es más representativo de otros entornos de prueba (Frutos et al., 2014b), por lo tanto, el sector 03 es el más representativo, mientras el sector 01 es el menos representativo.

Figura 4 Discriminativeness vs. Representativeness and Ranking Genotypes



Nota. A: Sector 1-Año 1; **B:** Sector 1-Año 2; **C:** Sector 1-Año 3; **D:** Sector 2-Año 1; **E:** Sector 2-Año 2; **F:** Sector 2-Año 3; **G:** Sector 3-Año 1; **H:** Sector 3-Año 2; **I:** Sector 3-Año 3; **J:** Sector 4-Año 1; **K:** Sector 4-Año 2; **L:** Sector 4-Año 3; **M:** Sector 5-Año 1; **N:** Sector 5-Año 2; **O:** Sector 5-Año 3.

CONCLUSIONES

CONCLUSIONES

Los resultados obtenidos en esta tesis permiten concluir lo siguiente:

1. Los modelos de ANOVA utilizados para evaluar datos agronómicos demuestran eficiencia, aunque presentan importantes limitaciones, ya que capturan una proporción muy pequeña de la varianza en los ensayos. No obstante, logran encontrar evidencias estadísticamente significativas para detectar diferencias entre las fuentes de variación en el rendimiento de la caña de azúcar.
2. Las pruebas post hoc basadas en el estadístico ANOVA son útiles para determinar, dentro de un mismo grupo, cuál tiene el mejor rendimiento basado en la media general o total. Sin embargo, al igual que el ANOVA, estas pruebas establecen diferencias y rankings basándose en una única entrada por prueba, ignorando las relaciones subyacentes que pueden existir en una matriz de dos o tres dimensiones.
3. Las técnicas biplot, como el GH y el HJ biplot, que permiten reducir la dimensionalidad de una matriz de múltiples datos y trabajan con densidades gaussianas para definir nuevas variables y ubicar individuos próximos en el hiperespacio, son la base de nuevas técnicas propuestas. Estas técnicas pueden capturar un mayor porcentaje de la variabilidad no explicada en modelos como el ANOVA.
4. La implementación de técnicas como el GGE biplot en el análisis de datos agronómicos, específicamente en el rendimiento de grandes extensiones de monocultivos, combina la capacidad de diferenciar estadísticamente a los grupos involucrados con la construcción de visualizaciones en 2D (o 3D) basadas en la reducción de dimensiones, mediante una selección manual de la descomposición en valores singulares (SVD), escalado y centrado.
5. Se ha desarrollado código en R que permite la aplicación de estas técnicas en diferentes cultivos, extendiendo su uso más allá de la caña de azúcar.

BIBLIOGRAFÍA

Bibliografía

- Abnave, V. B. (2020). Economic viability of sugarcane cultivation: A comparative analysis. *Journal of Sugarcane Research*, 10(2), 158. <https://doi.org/10.37580/JSR.2020.2.10.158-173>
- agricolae: Statistical Procedures for Agricultural Research. (2006). In *CRAN: Contributed Packages*. <https://doi.org/10.32614/CRAN.package.agricolae>
- Akan, K., Cat, A., Hocaoglu, O., & Tekin, M. (2023). Evaluating Scald Reactions of Some Turkish Barley (*Hordeum vulgare* L.) Varieties Using GGE Biplot Analysis. *Agronomy*, 13(12). <https://doi.org/10.3390/agronomy13122975>
- Akinwale, R. O., Fakorede, M. A. B., Badu-Apraku, B., & Oluwaranti, A. (2014). Assessing the usefulness of GGE biplot as a statistical tool for plant breeders and agronomists. *Cereal Research Communications*, 42(3), 534–546. <https://doi.org/10.1556/CRC.42.2014.3.16/METRICS>
- Almeida, T. I. R., De Souza Filho, C. R., & Rossetto, R. (2006). ASTER and Landsat ETM+ images applied to sugarcane yield forecast. *International Journal of Remote Sensing*, 27(19), 4057–4069. <https://doi.org/10.1080/01431160600857451>
- Amador-Sacoto, C., & Helfgott-Lerner, S. (2023). Sustainability of Sugarcane Farms in the Milagro Canton, Ecuador. *International Journal on Advanced Science, Engineering and Information Technology*, 13(3), 837–843. <https://doi.org/10.18517/ijaseit.13.3.18654>
- Annicchiarico, P. (1997). Additive main effects and multiplicative interaction (AMMI) analysis of genotype-location interaction in variety trials repeated over years. *Theoretical and Applied Genetics*, 94(8), 1072–1077. <https://doi.org/10.1007/S001220050517/METRICS>
- Asobanca [Asociación de Bancos Privados del Ecuador]. (2022). *GUÍA DE PRODUCCIÓN DE AZÚCAR*. <https://asobanca.org.ec/wp-content/uploads/2022/12/8.-Guia-Produccion-de-azucar.pdf>
- Butterfield, M., & Thomas, D. (1996). Sucrose, yield and disease resistance characteristics of sugarcane varieties under test in the SASEX selection programme. *In Proc S Afr Sug Technol Ass*, 70, 103–105.
- Castillo, R. O., & Silva Cifuentes, E. (2022). Sugarcane Breeding and Supporting Genetics Research in Ecuador. *Sugar Tech*, 24(1), 222–231. <https://doi.org/10.1007/S12355-021-01057-4/TABLES/4>
- Cevallos-Molina, E. R., Vélez-Vélez, T. V., Baquerizo-Crespo, R. J., & Gómez-Salcedo, Y. (2023). Anaerobic treatment of sugarcane bagasse. An opportunity for sustainability in rural environments? *Environmental Advances*, 13, 100427. <https://doi.org/10.1016/J.ENVADV.2023.100427>

- Dang, X., Hu, X., Ma, Y., Li, Y., Kan, W., & Dong, X. (2024). AMMI and GGE biplot analysis for genotype \times environment interactions affecting the yield and quality characteristics of sugar beet. *PeerJ*, *12*. <https://doi.org/10.7717/peerj.16882>
- de Setta, N., Monteiro-Vitorello, C. B., Metcalfe, C. J., Cruz, G. M. Q., Del Bem, L. E., Vicentini, R., Nogueira, F. T. S., Campos, R. A., Nunes, S. L., Turrini, P. C. G., Vieira, A. P., Ochoa Cruz, E. A., Corrêa, T. C. S., Hotta, C. T., de Mello Varani, A., Vautrin, S., da Trindade, A. S., de Mendonça Vilela, M., Lembke, C. G., ... Van Sluys, M. A. (2014). Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics*, *15*(1), 1–18. <https://doi.org/10.1186/1471-2164-15-540/FIGURES/5>
- Defante, L. R., Vilpoux, O. F., & Sauer, L. (2020). Importance of the sugarcane industry in the formal employment in the state of Mato Grosso do Sul during the period of 2008 to 2014. *Revista de Economia e Sociologia Rural*, *58*(4), e193496. <https://doi.org/10.1590/1806-9479.2020.193496>
- Elwakeel, A. E., Ahmed, S. F., Zein Eldin, A. M., & Hanafy, W. M. (2022). A review on sugarcane harvesting technology. *Al-Azhar Journal of Agricultural Engineering*, *2*(1), 54–63. <https://doi.org/10.21608/AZENG.2022.240434>
- Esan, V. I., Oke, G. O., Ogunbode, T. O., & Obisesan, I. A. (2023). AMMI and GGE biplot analyses of Bambara groundnut [*Vigna subterranea* (L.) Verdc.] for agronomic performances under three environmental conditions. *Frontiers in Plant Science*, *13*, 997429. <https://doi.org/10.3389/FPLS.2022.997429/BIBTEX>
- Frutos, E., Galindo, M. P., & Leiva, V. (2014). An interactive biplot implementation in R for modeling genotype-by-environment interaction. *Stochastic Environmental Research and Risk Assessment*, *28*(7), 1629–1641. <https://doi.org/10.1007/S00477-013-0821-Z/FIGURES/6>
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, *58*(3), 453–467. <https://doi.org/10.1093/BIOMET/58.3.453>
- Galindo, M. P. (1985). *Contribuciones a la representación simultánea de datos multidimensionales*. Universidad de Salamanca.
- Galindo-Villardón, P. (1986). Una Alternativa de Representación Simultánea: HJ-Biplot. *Qüestió: Quaderns d'estadística i Investigació Operativa*, *10*(1), 13–23.
- Haider, Z., Akhter, M., Mahmood, A., Ahsan, R., Khan, R., Kaku, S., & Lahore, P. (2017). Comparison of GGE biplot and AMMI analysis of multi-environment trial (MET) data to assess adaptability and stability of rice genotypes. *African Journal of Agricultural Research*, *12*(51), 3542–3548. <https://doi.org/10.5897/AJAR2017.12528>
- Hall, M. R. (2008). El azúcar en América Latina y el Caribe: Cambio tecnológico, trabajo, mercado mundial y economía azucarera: Perspectiva histórica y problemas actuales.

Hispanic American Historical Review, 88(2), 294–295.
<https://doi.org/10.1215/00182168-2007-125>

- Heinrichs, R., Otto, R., Magalhães, A., & Meirelles, G. C. (2017). Importance of Sugarcane in Brazilian and World Bioeconomy. *Economic Complexity and Evolution*, 205–217. https://doi.org/10.1007/978-3-319-58374-7_11
- Householder, A. S., & Young, G. (1938). Matrix Approximation and Latent Roots. *The American Mathematical Monthly*, 45(3), 165–171. <https://doi.org/10.1080/00029890.1938.11990787>
- Huang, Q., Feng, J., Gao, M., Lai, S., Han, G., Qin, Z., Fan, J., & Huang, Y. (2024). Precise Estimation of Sugarcane Yield at Field Scale with Allometric Variables Retrieved from UAV Phantom 4 RTK Images. *Agronomy* 2024, Vol. 14, Page 476, 14(3), 476. <https://doi.org/10.3390/AGRONOMY14030476>
- Khan, M. M. H., Rafii, M. Y., Ramlee, S. I., Jusoh, M., & Al Mamun, M. (2021). AMMI and GGE biplot analysis for yield performance and stability assessment of selected Bambara groundnut (*Vigna subterranea* L. Verdc.) genotypes under the multi-environmental trials (METs). *Scientific Reports* 2021 11:1, 11(1), 1–17. <https://doi.org/10.1038/s41598-021-01411-2>
- Lavín-Castañeda, J., Senties-Herrera, H. E., Trejo-Téllez, L. I., Bello-Bello, J.-J., Hidalgo-Contreras, J. V., & Gómez-Merino, F. C. (2020). Advances in the selection program of sugarcane (*Saccharum* spp.) varieties in the Colegio de Postgraduados. *Agro Productividad*, 13(11), 123–129. <https://doi.org/10.32854/AGROP.V13I11.1776>
- Liu, Q., Huang, L., Fu, C., Zhang, T., Ding, W., & Yang, C. (2022). Genotype–environment interaction of crocin in *Gardenia jasminoides* by AMMI and GGE biplot analysis. *Food Science and Nutrition*. <https://doi.org/10.1002/fsn3.3003>
- Maulana, H., Solihin, E., Trimio, L., Hidayat, S., Wijaya, A. A., Hariadi, H., Amien, S., Ruswandi, D., & Karuniawan, A. (2023). Genotype-by-environment interactions (GEIs) and evaluate superior sweet potato (*Ipomoea batatas* [L.] Lam) using combined analysis and GGE biplot. *Heliyon*, 9(9), e20203. <https://doi.org/10.1016/j.heliyon.2023.e20203>
- Mehareb, E. M., Osman, M. A. M., Attia, A. E., Bekheet, M. A., & Fouz, A. E. F. M. (2022). Stability assessment for selection of elite sugarcane clones across multi-environment based on AMMI and GGE-biplot models. *Euphytica*, 218(7). <https://doi.org/10.1007/s10681-022-03025-9>
- Mehdi, F., Cao, Z., Zhang, S., Gan, Y., Cai, W., Peng, L., Wu, Y., Wang, W., & Yang, B. (2024). Factors affecting the production of sugarcane yield and sucrose accumulation: suggested potential biological solutions. *Frontiers in Plant Science*, 15, 1374228. <https://doi.org/10.3389/FPLS.2024.1374228/BIBTEX>

- Natarajan, S., Basnayake, J., Lakshmanan, P., & Fukai, S. (2020). Limited contribution of water availability in genotype-by-environment interaction in sugarcane yield and yield components. *Journal of Agronomy and Crop Science*, 206(6), 665–678. <https://doi.org/10.1111/jac.12407>
- Oladosu, Y., Rafii, M. Y., Abdullah, N., Magaji, U., Miah, G., Hussin, G., & Ramli, A. (2017). Genotype × Environment interaction and stability analyses of yield and yield components of established and mutant rice genotypes tested in multiple locations in Malaysia*. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, 67(7), 590–606. <https://doi.org/10.1080/09064710.2017.1321138>
- Olivoto, T., & Lúcio, A. D. C. (2020). metan: An R package for multi-environment trial analysis. *Methods in Ecology and Evolution*, 11(6), 783–789. <https://doi.org/10.1111/2041-210X.13384>
- Podlich, D. W., Cooper, M., & Basford, K. E. (1999). Computer simulation of a selection strategy to accommodate genotype environment interactions in a wheat recurrent selection programme. *Plant Breeding*, 118(1), 17–28. <https://doi.org/10.1046/J.1439-0523.1999.118001017.X>
- Pourdad, S. S., & Moghaddam, M. J. (2013). Study on seed yield stability of sunflower inbred lines through GGE biplot. *Helia*, 36(58), 19–28. <https://doi.org/10.2298/HEL1358019P>
- Sa'diyah, H., & Hadi, A. F. (2016). AMMI Model for Yield Estimation in Multi-Environment Trials: A Comparison to BLUP. *Agriculture and Agricultural Science Procedia*, 9, 163–169. <https://doi.org/10.1016/j.aaspro.2016.02.113>
- Scavo, A., Mauromicale, G., & Ierna, A. (2023). Genotype × environment interactions of potato tuber quality characteristics by AMMI and GGE biplot analysis. *Scientia Horticulturae*, 310, 111750. <https://doi.org/10.1016/j.scienta.2022.111750>
- Silva, C., Pacheco, R., Arcentales, D., & Santos, F. (2020). Sustainability of sugarcane for energy purposes. In *Sugarcane Biorefinery, Technology and Perspectives* (pp. 89–102). Elsevier. <https://doi.org/10.1016/B978-0-12-814236-3.00005-6>
- Smith, A., Cullis, B., & Thompson, R. (2001). Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. *Biometrics*, 57(4), 1138–1147. <https://doi.org/10.1111/J.0006-341X.2001.01138.X>
- Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics*, 5(2), 99. <https://doi.org/10.2307/3001913>
- Ukkelberg, Å., & Borgen, O. S. (1993). Outlier detection by robust alternating regression. *Analytica Chimica Acta*, 277(2), 489–494. [https://doi.org/10.1016/0003-2670\(93\)80460-3](https://doi.org/10.1016/0003-2670(93)80460-3)
- Valenzuela-Cobos, J. D., Guevara-Viejó, F., Vicente-Galindo, P., & Galindo-Villardón, P. (2023). Eco-Friendly Biocontrol of Moniliasis in Ecuadorian Cocoa Using Biplot

Techniques. *Sustainability* 2023, Vol. 15, Page 4223, 15(5), 4223. <https://doi.org/10.3390/SU15054223>

- Van Eeuwijk, F. A., Malosetti, M., Yin, X., Struik, P. C., & Stam, P. (2005). Statistical models for genotype by environment data: from conventional ANOVA models to eco-physiological QTL models. *Australian Journal of Agricultural Research*, 56(9), 883–894. <https://doi.org/10.1071/AR05153>
- Yadawad, A., Patil, S. B., Kongawad, B. Y., Kadlag, A. D., & Hemaprabha, G. (2023). Multi environmental evaluation for selection of stable and high yielding sugarcane (*Saccharum officinarum* L.) clones based on AMMI and GGE biplot models. *Indian Journal of Genetics and Plant Breeding*, 83(3), 389–397. <https://doi.org/10.31742/ISGPB.83.3.11>
- Yan, W. (2002). Singular-Value Partitioning in Biplot Analysis of Multienvironment Trial Data. *Agronomy Journal*, 94(5), 990–996. <https://doi.org/10.2134/AGRONJ2002.9900>
- Yan, W., Cornelius, P. L., Crossa, J., Hunt, L. A., & Yan, W. (2001). Two Types of GGE Biplots for Analyzing Multi-Environment Trial Data. *CROP SCIENCE*, 41, 656–663.
- Yan, W., & Hunt, L. A. (1998). Genotype by Environment Interaction and Crop Yield. *Plant Breeding Reviews*, 135–178. <https://doi.org/10.1002/9780470650110.CH4>
- Yan, W., Hunt, L. A., Sheng, Q., & Szlavnic, Z. (2000). Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. *CROP SCIENCE*, 40(3), 597–605.
- Yan, W., & Kang, M. S. (2002). GGE Biplot Analysis : A Graphical Tool for Breeders, Geneticists, and Agronomists. *GGE Biplot Analysis*. <https://doi.org/10.1201/9781420040371>
- Yan, W., & Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *Canadian Journal of Plant Science*, 86(3), 623–645. <https://doi.org/10.4141/P05-169>
- You, Q., Xu, L., Zheng, Y., Que, Y., Bakhsh, A., Khawar, K. M., Onarici, S., Ozel, C. A., & Rao, A. Q. (2013). Genetic Diversity Analysis of Sugarcane Parents in Chinese Breeding Programmes Using gSSR Markers. *The Scientific World Journal*, 2013(1), 613062. <https://doi.org/10.1155/2013/613062>
- Zhao, F., Li, Y., Cui, T., & Bai, J. (2023). GGE Biplot-Based Transcriptional Analysis of 7 Genes Involved in Steroidal Glycoalkaloid Biosynthesis in Potato (*Solanum tuberosum* L.). *Agronomy*, 13(8). <https://doi.org/10.3390/agronomy13082127>
- Zhou, M. M., Kimbeng, C. A., Tew, T. L., Gravois, K. A., & Pontif, M. J. (2011). Artificial Neural Network Models as a Decision Support Tool for Selection in

Sugarcane: A Case Study Using Seedling Populations. *Crop Science*, 51(1), 21–31.
<https://doi.org/10.2135/CROPSCI2010.02.0057>