

26. PRECISIÓN DE SELECCIÓN TEMPRANA USANDO MODELOS MIXTOS EN EL PROGRAMA DE MEJORAMIENTO GENÉTICO DE ARROZ DE INIA

S. Scheffel¹, F. Molina², F. Pérez de Vida³, J. E. Rosas⁴

PALABRAS CLAVE: BLUE, BLUP, valor genético

INTRODUCCIÓN

La precisión de selección de líneas experimentales en etapas tempranas de evaluación tiene gran impacto en la eficiencia de los programas de mejoramiento genético, ya que determina la intensidad de selección que puede aplicarse y por ende la cantidad de recursos requeridos para obtener líneas avanzadas significativamente superiores. Seleccionar en el primer ensayo de evaluación de rendimiento (E1) es desafiante dada la limitada representación de ambientes, lo que dificulta la estimación de efectos y componentes de varianza de los factores ambientales y su interacción con los efectos genotípicos. El análisis conjunto de múltiples ensayos y zafra permite una mejor estimación de estos efectos y componentes de varianza. Por otra parte, el uso de modelos mixtos que consideran el efecto genotípico (de las líneas) como aleatorio genera un acercamiento hacia la media de las predicciones lineales de estos efectos (*best linear unbiased prediction*, BLUP) que en general mejora la precisión de la estimación del valor genético. El Programa de Mejoramiento Genético de Arroz de INIA (PMGA) completó recientemente la consolidación de una base de datos en la que se integra

toda la información generada por el programa desde 1996-1997 a 2019-2020. Los datos consolidados facilitan la realización de análisis conjuntos usando información de varios ensayos y zafra.

Este trabajo tiene por objetivo evaluar la precisión de las estimaciones y predicciones del valor genético del análisis conjunto de múltiples ensayos y años.

MATERIALES Y MÉTODOS

Los datos utilizados provienen de la base de datos unificada del PMGA, la cual cuenta con toda la información disponible de los ensayos de evaluación fenotípica del programa desde 1996-1997 a 2019-2020 (Rebollo *et al.*, 2020). Se ajustaron modelos para estimar los valores genéticos de las líneas simulando diferentes disponibilidades de información al momento de seleccionar líneas en E1: 1) contando solamente con información de un ensayo (E), 2) contando con datos de todos los ensayos de una misma zafra (Z), 3) contando con todos los ensayos de cinco años incluyendo la zafra de evaluación y los cuatro anteriores (C5) y 4) con el conjunto total de los 23 años disponibles (CT). Se ajustó un modelo E por cada ensayo para estimar el valor genético de los genotipos modelados como efecto

¹ Sheila Scheffel, Ing. Agr. INIA. Estudiante de maestría Cs. Agrarias, F. Agronomía, UDELAR. sscheffel@inia.org.uy

² Federico Molina, Ph.D. INIA. Programa Nacional de Investigación en Producción de Arroz.

³ Fernando Pérez de Vida, Ph.D. INIA. Programa Nacional de Investigación en Producción de Arroz.

⁴ Juan E. Rosas, Dr. INIA. Programa Nacional de Investigación en Producción de Arroz. jrosas@inia.org.uy

fijo (BLUE): $y_{ij} = \mu + g_i + b_j + \varepsilon_{ij}$, donde y_{ij} es rendimiento, μ es el efecto fijo de la media general, g_i es el efecto fijo de genotipo, b_j es el efecto fijo de bloque, y ε_{ij} es el residual. El modelo Z fue: $y_{ijkl} = \mu + g_i + b_{j(k)} + e_k + l_l + (ge)_{ik} + \varepsilon_{ijkl}$, donde y_{ijkl} es rendimiento, μ es la media general, g_i es el efecto de genotipo, $b_{j(k)}$ el efecto bloque anidado en ensayo, e_k el efecto ensayo, l_l el efecto localidad, $(ge)_{ik}$ es la interacción genotipo por ensayo, y ε_{ijkl} es el residual del modelo, todos los efectos

fueron aleatorios. Los modelos C5 y CT se ajustaron con los mismos efectos aleatorios:

$$y_{ijklm} = \mu + g_i + b_{j(k)} + e_{k(m)} + l_l + z_m + (ge)_{ik} + (gz)_{im} + \varepsilon_{ijklm}$$

donde y_{ijklm} es rendimiento, μ es la media general, g_i es el efecto aleatorio de genotipo, $b_{j(k)}$ el efecto bloque anidado en ensayo, $e_{k(m)}$ el efecto ensayo anidado en zafra, l_l el efecto localidad, z_m el efecto zafra, $(ge)_{ik}$ la interacción genotipo por ensayo, $(gz)_{im}$ la interacción

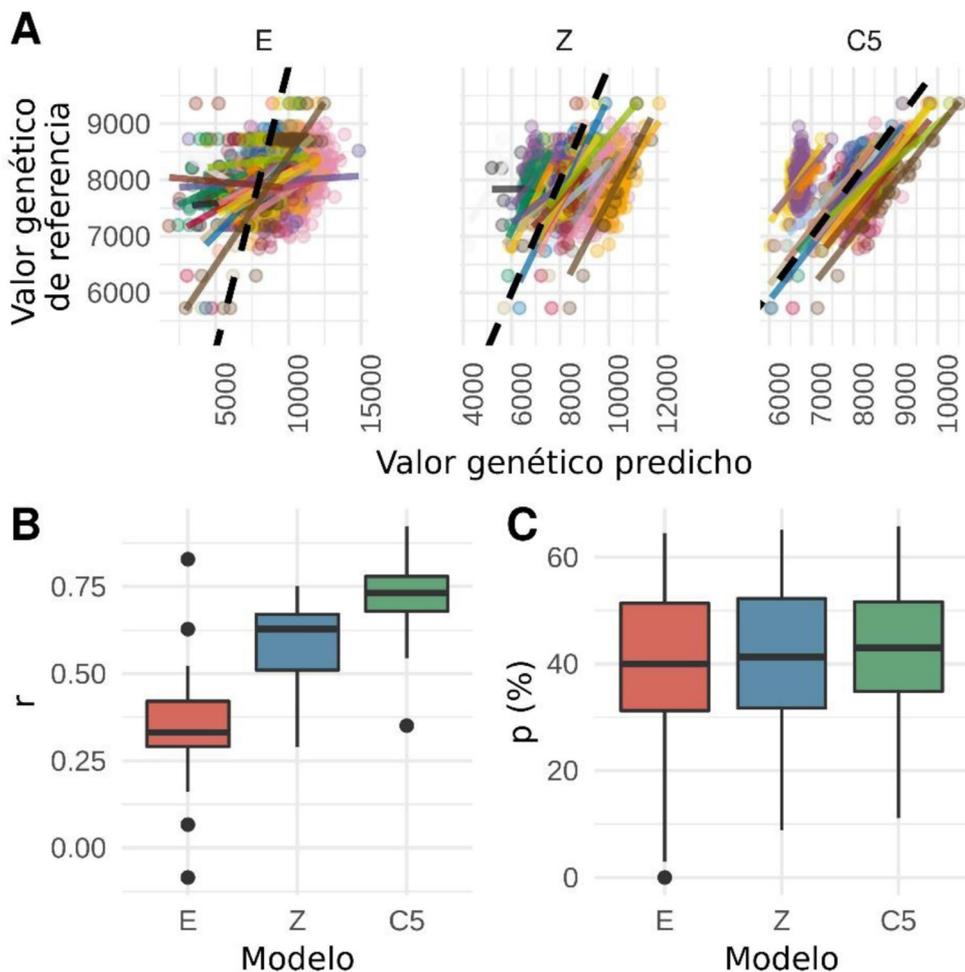


Figura 1. Relación entre los valores genéticos de referencia y los estimados o predichos por los modelos utilizando información de cada ensayo por separado (E), análisis multi ensayo por zafra (Z), análisis multi año con 5 años incluyendo la zafra de evaluación (C5). A) Regresión del valor genético de referencia en función del valor genético predicho por cada modelo. Los colores representan los años de evaluación (1997-2020), las líneas enteras la tendencia de regresión para cada año, y las líneas punteadas una relación 1:1. B) Valores de correlación (r) entre valores genéticos de referencia y los estimados o predichos por cada modelo. C) Porcentaje (p) de las mejores líneas seleccionadas de acuerdo con el valor genético de referencia que coinciden con las mejores líneas seleccionadas de acuerdo con cada modelo, aplicando la misma intensidad de selección.

genotipo por zafra, y ϵ_{ijklm} es el residual del modelo. Para los efectos aleatorios se asume $\sim N(0, \sigma^2)$ con varianzas independientes e idénticamente distribuidas.

El valor genético de referencia se definió como el BLUP obtenido con el modelo CT para aquellas líneas presentes en tres o más años en evaluaciones a campo. Como estimadores de la precisión de selección para cada modelo y zafra de evaluación se calculó: a) la correlación (r) entre los valores genéticos de referencia y los obtenidos con los modelos evaluados y b) el porcentaje (p) de líneas seleccionadas por cada modelo de líneas que coinciden con las que efectivamente fueron seleccionadas (población de referencia).

RESULTADOS DE LA INVESTIGACIÓN

Las precisiones de predicción del valor genético se muestran en la figura 1A, donde se grafica la relación entre el valor genético de referencia (BLUP con CT) y los valores genéticos predichos por E (BLUE), Z y C5

(BLUP). A medida que los modelos usan más información, los valores genéticos predichos se encogen hacia la media, una propiedad característica de los BLUPs, la cual aumenta la precisión de la estimación (Piepho *et al.*, 2008). También a medida que aumenta el uso de información, r y p son más altos y menos dispersos, aunque p de forma menos pronunciada (Figura 1B y 1C). Esta mejora progresiva en la predicción del valor genético se debe también a un más adecuado modelado del efecto ambiental y de la interacción entre genotipo y ambiente, ambos factores clave para estimar en forma más precisa los valores genéticos (van Eeuwijk *et al.*, 2016). En el modelo CT (Figura 2) la variación atribuida al ambiente está representado por los factores ensayo y zafra, cuya contribución al desvío estándar es mayor a 2.000 kg/ha, y la de los factores que representan la interacción genotipo por ambiente (Genotipo por Ensayo y Genotipo por Zafra) suma más de 900 kg/ha. En el modelo E la información disponible no permite modelar ninguno de estos efectos, mientras que en el modelo Z se puede modelar el efecto de ensayo y su interacción

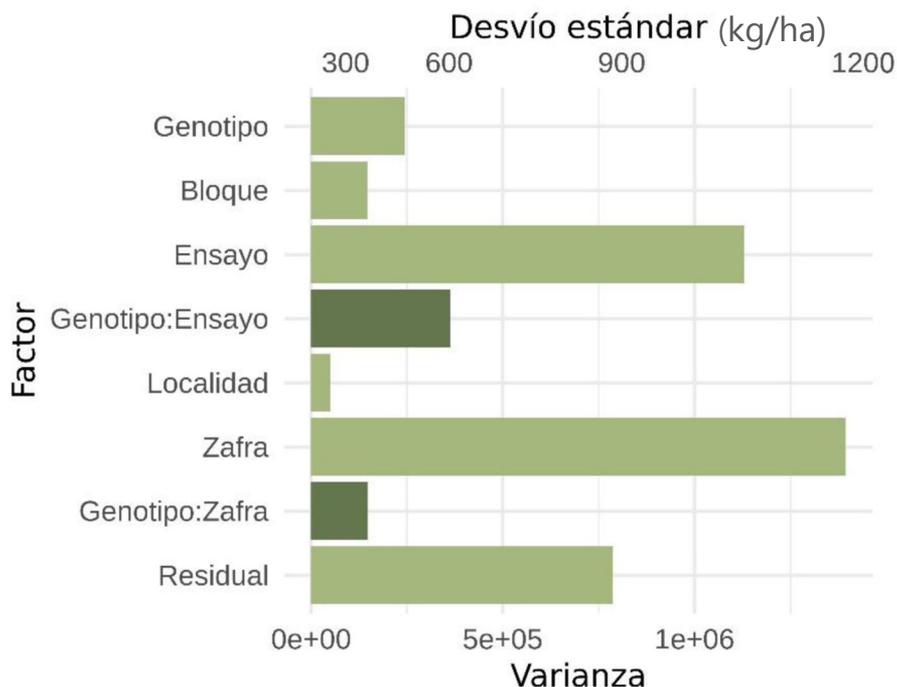


Figura 2. Distribución de varianzas y desvíos estándar en el total de los datos (1997-2020) ajustados en el modelo conjunto (CT).

con genotipo, y en el modelo G5 se pueden modelar todos los efectos.

CONCLUSIONES

La precisión lograda en la estimación de los valores genéticos de las líneas en evaluación temprana fue mayor a medida que se incorporó la modelación de los efectos ambientales correspondientes a ensayo, año y de la interacción del genotipo con estos factores. Esto se corresponde con la alta proporción de la variación fenotípica que corresponde a efectos ambientales y de interacción genotipo por ambiente.

Caracterizar estos efectos ambientales y su interacción con el genotipo es clave para predecir el comportamiento de las líneas experimentales en futuros ambientes. Esto será abordado en futuros trabajos incorporando al modelo covariables climáticas en las etapas críticas de construcción del rendimiento. Por otra parte, los modelos utilizados en este trabajo asumen covarianzas nulas entre genotipos, por lo que incorporar las relaciones

de parentesco entre ellos podría disminuir la varianza residual y mejorar aún más las precisiones de predicción.

BIBLIOGRAFÍA

Piepho, H.; Mohring, J.; Melchinger, A.; Buchse, A. 2008. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*, v. 161, n. 1-2, p. 209-228. Doi: <https://doi.org/10.1007/s10681-007-9449-8>

Rebollo, I.; Scheffel, S.; Iriarte, W.; Blanco, P.; Molina, F.; Pérez De Vida, F.; Rosas, J. 2020. Consolidación de los datos históricos del Programa de Mejoramiento Genético de Arroz en una base de datos. In: Terra, J. A.; Martínez, S.; Saravia, H.; Mesones, B.; Álvarez, O. (Eds.) *Arroz 2020*. Montevideo: INIA. p. 5-8. (Serie Técnica 257). doi: <http://doi.org/10.35676/INIA/ST.257>.

Van Eeuwijk, F.; Bustos-korts, D., Malosetti M. 2016 What should students in plant breeding know about the statistical aspects of genotype × environment interactions? *Crop Science*, v. 56, no. 5, p. 2119–40. Doi: <https://doi.org/10.2135/cropsci2015.06.0375>