

APROXIMACIÓN BAYESIANA AL MÉTODO DE ESTABILIDAD FENOTÍPICA DE EBERHART Y RUSSELL

Camilo Eduardo Quintero Jaramillo¹

José Miguel Cotes Torres^{2,4}

Elena Paola González Jaimes³

BAYESIAN APPROACH TO
EBERHART AND RUSSELL
PHENOTYPIC STABILITY
METHODOLOGY

Fecha de recepción: 4 de septiembre de 2010

Fecha de aceptación: 18 de diciembre de 2010

1 Ingeniero Agrónomo M.B.A. Facultad de Ciencias Agropecuarias. Universidad Nacional de Colombia, Sede Medellín.

2 Ingeniero Agrónomo Ph.D. Facultad de Ciencias Agropecuarias. Universidad Nacional de Colombia, Sede Medellín.

3 Ingeniero Agrónomo Ph. D. Facultad de Ciencias Agrarias. Politécnico Colombiano Jaime Isaza Cadavid. Medellín Colombia.

4 Autor para correspondencia: jmcotes@unal.edu.co

RESUMEN

Una metodología muy utilizada en el mejoramiento de plantas para determinar la adaptabilidad y la estabilidad de los genotipos es la propuesta de Eberhart y Russell a comienzos de la década de los 60's. Las pruebas de hipótesis para los parámetros de adaptabilidad y estabilidad bajo este modelo requieren un número elevado de ambientes, este supuesto es frecuentemente violado en el diseño de pruebas regionales. La estadística Bayesiana, con el uso de las distribuciones *a priori* resuelve con mayor precisión la inferencia del parámetro de interés que son los parámetros de estabilidad especialmente la predecibilidad de los genotipos. Para la obtención de dichos parámetros se implementó una función MACRO en el programa SAS v.8.2. que permitió simular escenarios de pruebas regionales y aplicar la metodología estadística tradicional y luego la metodología Bayesiana, con el fin de comparar estas dos metodologías implementando el modelo de Eberhart y Russell. También se utilizaron datos reales de pruebas regionales de maíz como otra vía para comparar estas metodologías. Se encontró que la metodología Bayesiana permite una inferencia más precisa cuando los genotipos son predecibles y estables, además la distribución *a priori* de Jeffreys presentó un mejor ajuste al verdadero valor del parámetro de estabilidad al ser comparada con la *a priori* Flat.

Palabras clave: Pruebas regionales, adaptabilidad fenotípica, interacción genotipo-ambiente, MCMC.

ABSTRACT

A widely used plant breeding methodology to determine the genotype stability and

adaptability is the proposed by Eberhart and Russell at the beginning of 60's decade. The hypothesis testes for the adaptability and stability parameters based on this model require a high number of environments; this assumption is frequently omitted in the design of trial test. The Bayesian statistics, using the *a priori* distributions resolve with high precision the estimates of stability and predictability parameters. For evaluate those parameters a MACRO function in the SAS v.8.2 program were implemented which allow to simulate sceneries of regional t rial in order to compared the traditional statistical methodology and then the Bayesian methodology, both methodologies using the Eberhart and Russell model. Real dataset of corn trial test was used as other way to compare both methodologies. It was found that Bayesian methodologies allow us an accurate inference when genotypes are predictable and stable, and the *Jeffreys* prior distribution showed a better adjust to the real value of the stability parameter when compared whit the *Flat* prior.

Key words: Regional trial, phenotypic adaptability, genotype by environment interaction, MCMC.

INTRODUCCIÓN

La selección de genotipos por medio de pruebas regionales es una práctica muy importante en la agronomía y el mejoramiento de plantas porque ayudan a estimar y predecir el rendimiento con alta precisión con una pequeña cantidad de datos experimentales y también determinan patrones de respuesta en diferentes ambientes que permiten seleccionar el mejor genotipo para años futuros y nuevas localidades (Crossa, 1990).

El valor fenotípico de un carácter cuantitativo incluye un componente genético, uno ambiental y una interacción entre el genotipo y el ambiente, que da como resultado una superposición de las clases genotípicas, que acentúa aún más la variación continua. (Poehlman, 2003). En general, es aceptado que la expresión fenotípica (F) de los diferentes caracteres, depende del genotipo (G), el ambiente (A) y de la interacción genotipo por ambiente (IGA).

El genotipo es la constitución hereditaria completa (expresada y latente) de un organismo, que comprende todos los genes localizados en los cromosomas y los factores de herencia citoplasmática; por tanto la respuesta a los diferentes ambientes dependerá del *pool* o paquete genético y la expresión de múltiples genes (Vallejo y Estrada, 2002).

El ambiente es el conjunto de todas las condiciones externas que afectan el crecimiento y desarrollo de un organismo. Incluyen factores ambientales predecibles (algunas características climáticas como radiación solar; tipo y fertilidad de suelo, fecha, densidad y método de siembra) y factores ambientales impredecibles (cantidad y distribución de las lluvias; temperatura y humedad relativa; presiones repentinas de insectos o enfermedades) (Kang y Magari, 1996; Hunh 1996; Vallejo y Estrada, 2002).

La IGA reduce la asociación entre los valores genéticos y fenotípicos, obligando a los fitomejoradores a considerar la estabilidad y/o adaptabilidad de los materiales.

La *estabilidad biológica, homeostática* o en sentido estático se refiere al comportamiento constante, sin variación; de un genotipo a través de todos los ambientes donde es evaluado, independiente de si las condiciones son favorables

o no para el cultivo y la *estabilidad agronómica* o en sentido dinámico es la capacidad de un genotipo de responder al potencial productivo ofrecido por cada ambiente donde es evaluado (Piepho, 1996; Vallejo y Estrada 2002).

Para cuantificar la estabilidad o adaptabilidad del rendimiento se han desarrollado numerosos y variados procedimientos estadísticos, dentro de las metodologías conocidas, las que se basan en la regresión son ampliamente difundidas entre los estudiosos de la IGA.

Finlay y Wilkinson (1963) propusieron realizar un análisis de regresión del rendimiento de cada variedad sobre el índice ambiental de cada localidad, con el fin de estimar la estabilidad de los genotipos. El índice ambiental es una estimación del potencial de rendimiento de cada localidad y puede ser calculado como la diferencia entre el rendimiento promedio de las variedades en cada localidad y el promedio de las variedades en todas las localidades evaluadas

Eberhart y Russell (1966) realizaron un trabajo paralelo a Finlay y Wilkinson, llegando casi al mismo modelo pero incorporando las desviaciones respecto a la línea de regresión como un criterio de estabilidad.

Así el coeficiente de regresión lineal o pendiente, comúnmente llamado de β_1 , es considerado como un parámetro de estabilidad, considerando como genotipos de adaptabilidad amplia aquellos que tienen una pendiente cercana a 1.0. En caso contrario, el genotipo es considerado como de adaptabilidad específica (Cruz y Regazzi, 1994).

El otro parámetro de estabilidad son los desvíos de la regresión que permite declarar un genotipo como predecible o impredecible, dependiendo de la magnitud de los desvíos (Vallejo y

Estrada, 2002). Esto también puede observarse a través del coeficiente de determinación o R^2 de cada recta de regresión. Mayores valores de R^2 indican mayor estabilidad del genotipo.

Por otro lado, la metodología Bayesiana ha sido una opción bastante utilizada en el análisis de datos en los últimos años. Sin embargo, no fue siempre así y en otros tiempos; los resultados obtenidos por métodos Bayesianos eran frecuentemente vistos con cierta desconfianza una vez que tales métodos toman en consideración informaciones previas sobre parámetros que serán estimados, las cuales no provienen de datos experimentales de una variable de interés; denominada información *a priori* (Cotes et al., 2010).

La inferencia Bayesiana se basa en el Teorema de Bayes (Bayes, 1763), en el cual toda la información disponible para un parámetro de interés se representa mediante la distribución *a posteriori* la cual es proporcional a la información suministrada por los datos del nuevo estudio a través de la función de verosimilitud, y multiplicada por la información o conocimiento previo representado por la distribución *a priori*. Por lo tanto, el teorema de Bayes muestra cómo el conocimiento acerca del estado de la naturaleza representada por el parámetro es continuamente modificado a medida que nuevos datos (información) son adquiridos.

En el análisis Bayesiano se han implementado distribuciones *a priori* no informativas, se dice que una distribución *a priori* es no informativa cuando refleja una ignorancia total o un conocimiento muy limitado sobre el parámetro de interés.

El hecho de que la metodología Bayesiana no haya sido frecuentemente utilizada por tantos años, desde el trabajo de Bayes en

1763, y que sólo ahora haya ganado un fuerte impulso en las aplicaciones, se debe a las dificultades matemáticas que normalmente se encuentran en el cálculo de integrales, principalmente en problemas con más de un parámetro. A partir de la segunda mitad de la década de los 80's, con la facilidad de acceso a los computadores de gran capacidad de trabajo y cada vez más eficientes, además con el auxilio de programas de cómputo poderosos, fue posible resolver esas integrales por métodos numéricos y algoritmos fácilmente implementados (Cotes et al., 2010).

Una de las primeras ventajas que esta metodología presenta es la estimación no negativa de componentes de varianza (Box y Tiao, 1973), así como la presentación de los intervalos de credibilidad para algunos parámetros genéticos, los cuales no se pueden obtener por la metodología tradicional. Estimaciones no negativas pueden ser obtenidas por otras metodologías como máxima verosimilitud (ML) o máxima verosimilitud restringida (REML), pero en estas metodologías se realiza un ajuste, donde la estimación no negativa se transforma en estimación de varianza nula cambiando el modelo originalmente propuesto, lo cual puede afectar las estimativas de los demás componentes positivos de varianza (Searle et al., 1992; Cotes 2004).

Por otro lado, los algoritmos utilizados para realizar inferencia bayesiana en problemas de interacción genotipo por ambiente permiten modelar la heterogeneidad de varianzas entre los ambientes (Edwards y Jannink, 2006) o entre genotipos (Cotes et al., 2006), sin mayores dificultades computacionales, como ocurre con otros metodologías de estimación.

En la actualidad hay un número creciente de trabajos de aplicación de la metodología Bayesiana en mejoramiento animal, donde diferentes autores muestran varias ventajas de esta metodología, lo que ha motivado aún más el interés en su aplicación (Cotes et al., 2010). En el área de mejoramiento de plantas, aunque su uso empezó después de las aplicaciones de mejoramiento animal, es creciente el número de referencias que muestran un uso frecuente de la metodología Bayesiana en la toma de decisiones dentro de estos programas de mejoramiento.

Una de las principales ventajas de la metodología Bayesiana es la incorporación de la información *a priori* en la solución de determinado problema probabilístico, pero debe tenerse cuidado al escoger la distribución *a priori* pues ésta afectará directamente la distribución *a posteriori*, especialmente cuando se tienen pocos datos para realizar el análisis. Existen algunas reglas formales para la obtención de distribuciones *a priori* no informativas como es el caso de la propuesta realizada por Jeffreys (1961), con la cual se obtienen distribuciones *a priori* localmente uniformes. Sin embargo, esta regla debe ser tomada con precaución en el caso de estimación de varios parámetros ó multiparamétrico (Box y Tiao, 1973).

Esta investigación tuvo como objetivo comparar mediante simulación Monte Carlo la estimación de los parámetros de adaptabilidad y estabilidad fenotípica de la metodología de Eberhart y Russell estimados de forma tradicional y mediante inferencia bayesiana basada en dos diferentes tipos de distribuciones *a priori* no informativas.

MATERIALES Y MÉTODOS

Experimentos de Simulación

Para evaluar el método de Eberhart y Russell por la metodología estadística tradicional (Eberhart y Russell, 1966) y por la propuesta Bayesiana, se generaron escenarios de simulación considerando cinco (5) genotipos asumiendo igual rendimiento promedio para todos los genotipos, tomando un valor de $\beta_0 = 5$. Los coeficientes de regresión β_1 que indican estabilidad fenotípica del material incluyen los genotipos estables, los que rinden mejor a ambientes favorables y los que rinden mejor en ambientes desfavorables (Tabla 1).

Los escenarios fueron generados mediante la implementación de una MACRO en el sistema SAS v 8.2 (SAS, 2000), la cual se muestra en el Apéndice 1. De cada escenario fueron obtenidos 10000 ensayos equivalentes a 10000 estudios de pruebas regionales de cinco (5) genotipos por nueve (9) ambientes. Para los cinco genotipos simulados fueron considerados diferentes valores para el parámetro δ_i .

Datos de ensayos regionales

Fueron utilizados datos provenientes de la tesis de maestría de Freddy Salazar realizada en 1996 y publicados por Vallejo y Estrada (2002), relacionadas con la evaluación del rendimiento de ocho variedades (*Zea mays* L.) en cinco ambientes de suelos ácidos y uno normal localizado en Palmira (Colombia). Los datos se consignaron en una hoja de cálculo del programa Microsoft Excel ® y fueron exportadas al sistema SAS v 8.2, igual que para los datos simulados, los resultados de las rutinas programadas para correr los datos por la

Escenario	Genotipo	β_{0i}	β_{1i}	δ_i	Interpretación	
					Estabilidad	Predicibilidad
1	G1	5	1,5	1,00	Ambientes desfavorables	Poco predicibles
	G2	5	1,2	1,00		
	G3	5	1,0	1,00	Estable	
	G4	5	0,8	1,00	Ambientes favorables	
	G5	5	0,5	1,00		
2	G1	5	1,5	0,50	Ambientes desfavorables	Medianamente predicibles
	G2	5	1,2	0,50		
	G3	5	1,0	0,50	Estable	
	G4	5	0,8	0,50	Ambientes favorables	
	G5	5	0,5	0,50		
3	G1	5	1,5	0,25	Ambientes desfavorables	Muy predicibles
	G2	5	1,2	0,25		
	G3	5	1,0	0,25	Estable	
	G4	5	0,8	0,25	Ambientes favorables	
	G5	5	0,5	0,25		

β_0 = Rendimiento medio, β_1 = Pendiente de regresión, δ_i = Desvíos de regresión.

Tabla 1. Valores para cada uno de los parámetros de simulación utilizados para el modelo de Eberhart y Russell.

metodología tradicional y Bayesiana se pueden visualizar a través del programa SAS System Viewer v.9.1.3 (SAS, 2006).

Análisis de los ensayos simulados y datos reales de campo

Para el análisis tradicional y Bayesiano fue realizada una rutina MACRO en SAS v. 8.2, utilizando como base el procedimiento MIXED (Apéndice 2). La metodología tradicional se obtiene a través de la solución de los efectos fijos, con sus respectivos intervalos de confianza, mientras que el análisis bayesiano se obtiene con la opción PRIOR de este procedimiento. Para obtener los resultados con distintas distribuciones *a priori* se utilizaron las opciones *Flat* y la opción *Jeffreys*.

Para la obtención de muestras de la distribución *a posteriori* conjunta se consideró

el algoritmo *Independence Chain*, el cual es el predeterminado del procedimiento MIXED (Wolfinger y Kass, 2000). Por cada ensayo fueron obtenidas 10000 muestras de cada uno de los parámetros.

Después de obtenidas las muestras de la distribución *a posteriori* conjunta fueron obtenidas estimaciones de las distribuciones marginales usando el procedimiento KDE del sistema SAS v 8.2. (SAS, 1999). Como estimadores de Bayes fueron considerados la media, la mediana y la moda de estas distribuciones y como estimación por intervalos fueron obtenidos los HPD (*Highest Posterior Density* por sus siglas en ingles) ó Regiones de la Densidad Posterior Más Alta (RDPMA), con una probabilidad del 90%. Una MACRO (SAS, 2000) fue construida para facilitar estos cálculos, la cual se presenta en el Apéndice 3.

Las rutinas de simulación fueron ejecutadas bajo un sistema computacional con sistema operativo Windows XP profesional edition Service Pack 2, con un procesador Pentium IV (a 1,8 Ghz, con 512Mb de RAM) provistos por la Universidad Nacional de Colombia Sede Medellín a través del Posgrado en Ciencias Agrarias y programadas en el sistema SAS versión 8.2 bajo el modulo IML (SAS, 1999) licenciado para el mismo establecimiento (Ver apéndice 1 y 2) y empleando funciones Macro (SAS, 2000) para hacer más eficiente la estructura del programa de simulación.

Comparación Metodologías

El error cuadrático medio ó *ECM* es una vía objetiva para comparar ambas metodologías, la Bayesiana y la clásica (Cotes, 2004). Así, para este estudio se estimó el *ECM* de la siguiente manera:

$$ECM = \frac{\sum_{i=1}^{10000} (\theta_0 - \hat{\theta})^2}{10000}$$

donde θ son los parámetros β_0, β_1 y δ_1 simulados y $\hat{\theta}$ son los valores estimados. De esta forma se obtendrá un *ECM* para la metodología

tradicional y otro para la metodología Bayesiana. La metodología que produzca el menor valor de *ECM* será la mejor metodología, porque en promedio se acerca más al verdadero valor de θ .

RESULTADOS Y DISCUSIÓN

Datos simulados

Los resultados arrojados producto de la simulación se presentan a continuación según la distribución *a priori* empleada, es decir *Flat* o *Jeffreys*.

A priori Flat

La figura 1 muestra los valores de *ECM* bajo la distribución *Flat* para el parámetro β_0 (Promedio de los genotipos), se aprecia como en el primer escenario simulado los valores de *ECM* son los más altos, esto debido a que el modelo es poco predecible, es decir los valores de δ_i afectan fuertemente la inferencia que se realiza sobre los parámetros, aún en el caso de la estimación tradicional.

El parámetro de estabilidad (β_1) al igual que β_0 , se afecta fuertemente por el escenario en que se pretenda realizar la estimación del parámetro de estabilidad, la figura 2 muestra además que

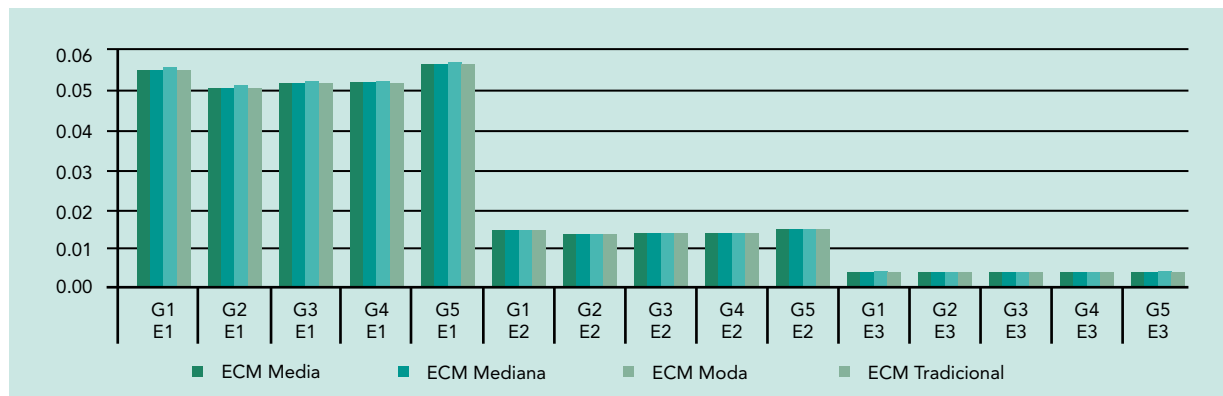


Figura 1. Valores *ECM* para los estimadores de β_0 encontrados vía simulación usando *Jeffreys* como distribución *a priori*.

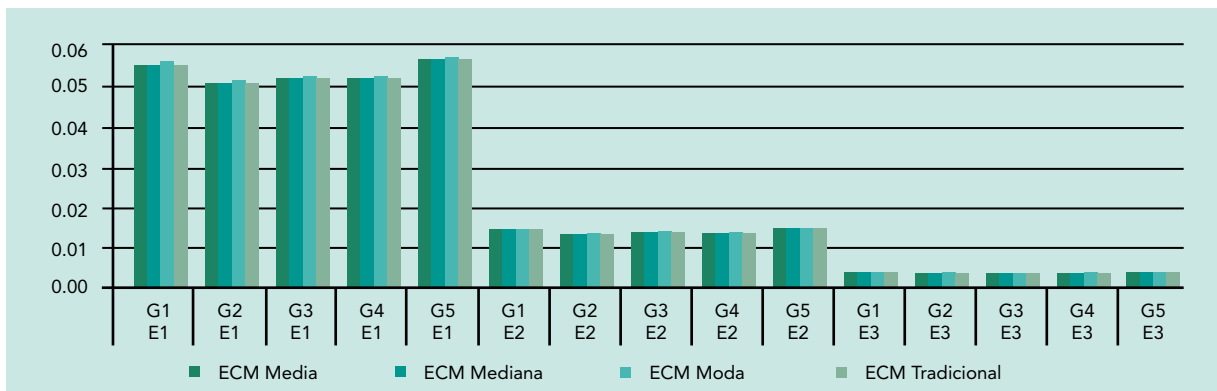


Figura 2. Valores *ECM* para los estimadores de β , encontrados vía simulación usando *Flat* como distribución *a priori*.

los genotipos estables, es decir con β_i igual ó cercano a 1, presentan menores valores de *ECM*.

Hasta ahora, los valores de *ECM* no difieren mucho entre sí, y no señalan con claridad que tipo de estimador es mejor, la metodología Bayesiana no se ha destacado debido a que la inferencia sobre medidas de tendencia central (primer momento al origen) son muy bien estimadas por ambos métodos.

Con relación a la estimación del parámetro δ_i , se aprecian altos valores de *ECM* para el estimador de la media por la metodología Bayesiana en

el primer escenario (Figura 3), y en general, este fue el escenario que mayores valores de *ECM* para la estimación de δ_i usando *Flat* como *a priori*.

En los otros ambientes no se revelan grandes diferencias debido a que los *ECM* en el primer escenario son altos disminuyendo la resolución para los otros dos escenarios.

En caso de querer estimar δ_i con la metodología cuando se tiene evidencia de altos valores en las desviaciones a la regresión (escenario 1) es preferible utilizar la moda ya que es el que menor valor de *ECM* presenta.

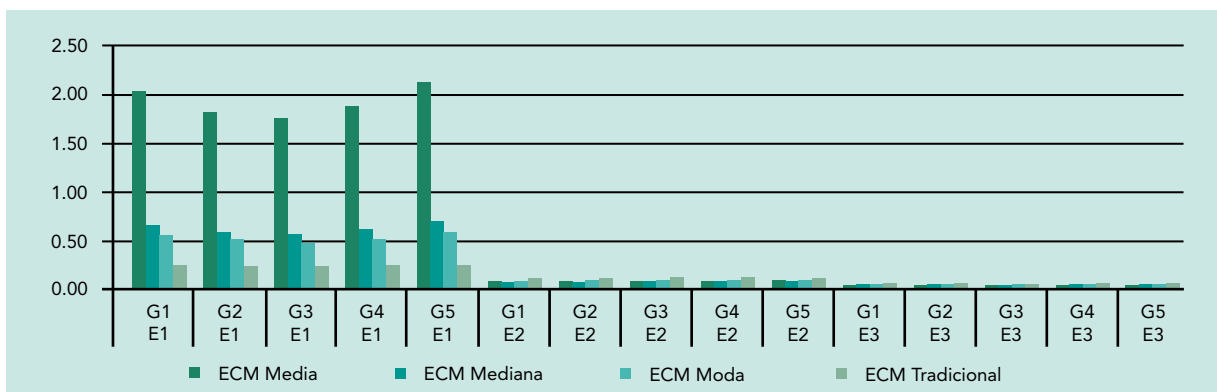


Figura 3. Valores *ECM* para los estimadores de δ_i encontrados vía simulación usando *Flat* como distribución *a priori*.

Cuando los genotipos son más predecibles, el estimador Bayesiano para la media, supera el estimador tradicional. Nuevamente, las estimaciones Bayesianas son preferibles para estimar parámetros de variabilidad o de dispersión como δ_i , (segundo momento al origen, que implica términos cuadráticos).

La metodología Bayesiana es muy sensible a los valores altos de δ_i que reflejan el poco ajuste de los datos al modelo lineal propuesto por Eberhart y Russell, pero el comportamiento de los estimadores Bayesianos es diferente cuando se analizan los escenarios 2 y 3, en ellos las desviaciones a la regresión son menores, el estimador Bayesiano para la media y en general todos los estimadores Bayesianos (como se muestra en la figura 4 los escenarios 2 y 3), son mejores que el estimador tradicional cuando se trata de estimar la variabilidad de los datos respecto al modelo.

A priori de Jeffreys

Los mismos estimadores se analizan teniendo en cuenta que se utilizó la distribución *a priori* de *Jeffreys* permitiendo luego comparar entre ambas distribuciones también con los valores de *ECM*.

Los valores de *ECM* al estimar β_0 no difieren mucho de los valores alcanzados usando *Jeffreys* como *a priori*. Nótese que los genotipos estables como el G3 en el escenario 1 (figura 5) presentan menores valores de *ECM*, la misma situación que se presentó en la estimación usando *Flat*.

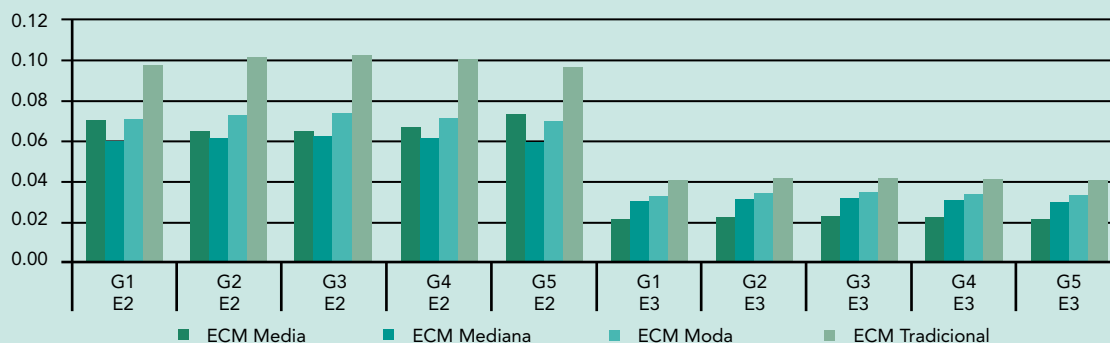


Figura 4. Valores *ECM* para los estimadores de δ_i encontrados vía simulación en los escenarios 2 y 3 usando *Flat* como distribución *a priori*.

Al evaluar los valores de *ECM* al estimar el parámetro de estabilidad β_1 , se obtiene una situación similar para el caso de β_0 .

Como se muestra en la figura 6, los valores de *ECM* son muy similares para los diferentes estimadores de β_1 , por tanto el estimador tradicional y los Bayesianos presentan una capacidad predictiva similar al estimar el parámetro de estabilidad β_1 .

Al igual que en la distribución *Flat*, usando *Jeffreys* se obtienen valores altos de *ECM* en el escenario 1 con la Media Bayesiana al inferir δ_i , aunque los otros estimadores no se alejan

tanto del valor de *ECM* del estimador tradicional, esto prueba de nuevo que la metodología Bayesiana es muy sensible al ajuste que tengan los datos al modelo lineal de Eberhart y Russell, es decir; ante situaciones con genotipos poco predecibles como se nota en la figura 7.

Cuando los genotipos son más predecibles, el estimador Bayesiano para la media, supera el estimador tradicional como se evidencia en la Figura 8. Nuevamente, las estimaciones Bayesianas son preferibles para estimar parámetros de variabilidad o de dispersión como δ_i (segundo momento al origen, que implica términos cuadráticos).

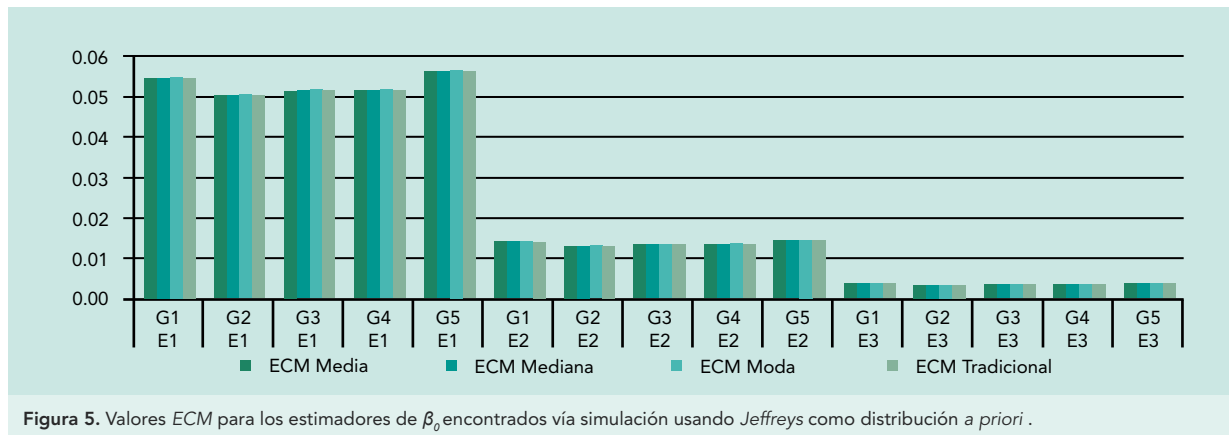


Figura 5. Valores *ECM* para los estimadores de β_0 , encontrados vía simulación usando *Jeffreys* como distribución *a priori*.

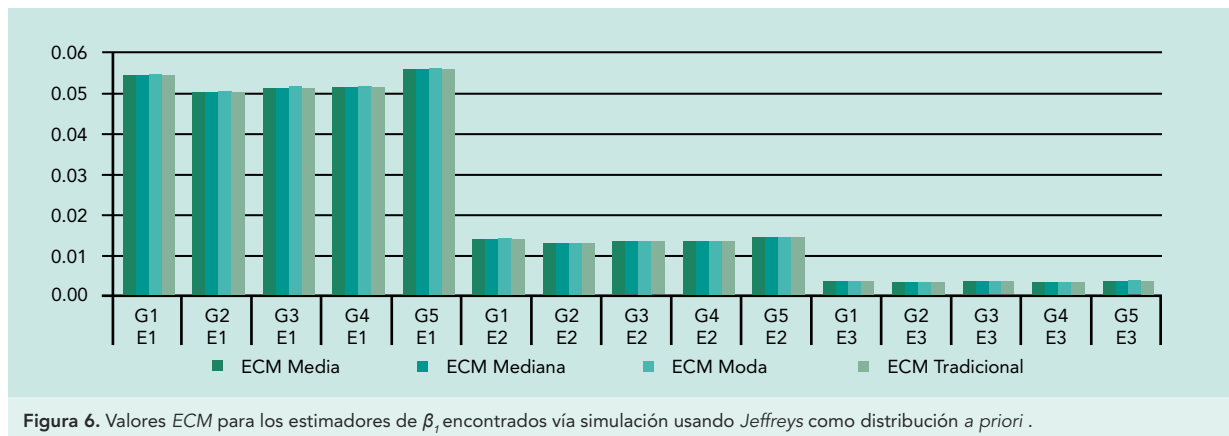


Figura 6. Valores *ECM* para los estimadores de β_1 , encontrados vía simulación usando *Jeffreys* como distribución *a priori*.

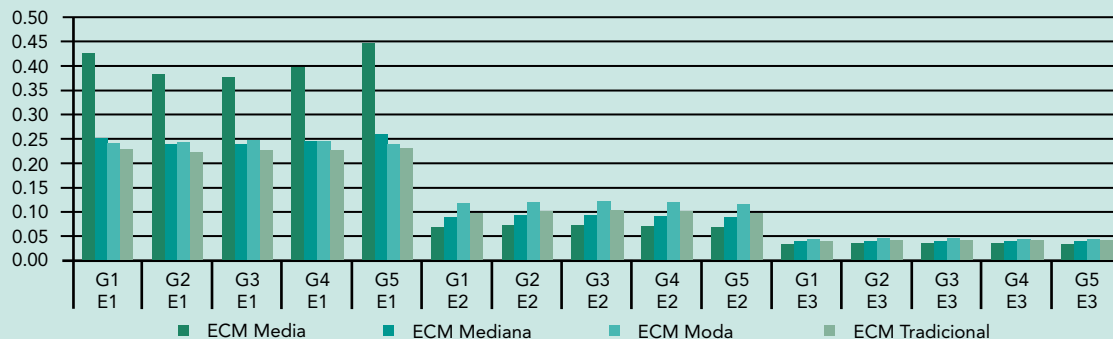


Figura 7. Valores ECM para los estimadores de δ_i encontrados vía simulación usando *Jeffreys* como distribución *a priori*.

Tomando en cuenta los resultados al analizar δ_i se tiene un escenario en el cual los genotipos no son predecibles (escenario 1), desde el punto de vista del mejorador un genotipo no predecible no es un buen objetivo en un programa de mejoramiento, ni cuando se piensa en un genotipo para la liberación de germoplasma en una región, ya que se buscan buenas producciones pero estables.

Luego, con genotipos predecibles, que brindan resultados confiables en los programas de mejoramiento, las metodologías Bayesianas son mucho más acertadas, en especial al evaluar el estimador de la media.

Por medio de los valores de ECM también es posible comparar objetivamente las *a priori Flat* y *Jeffreys*.

En la tabla 2 se reportan los valores usando las dos distribuciones *a priori* y no se detectan diferencias numéricas entre los diferentes estimadores de β_0 tomando en cuenta los genotipos y los escenarios. La tabla 3 refleja la misma situación al comparar las dos distribuciones a través de los escenarios y los genotipos para el parámetro β_1 .

Finalmente, al tabla 4, muestra numéricamente cómo los valores de ECM calculados para las distribuciones, difieren al evaluar los diferentes

Las densidades de probabilidad de los parámetros brindan información acerca de cuán preciso es el cálculo de los estimadores en el caso bayesiano, de modo que permitirá comparar las distribuciones *a priori* que pautan la probabilidad para las estimaciones de los parámetros de estabilidad.

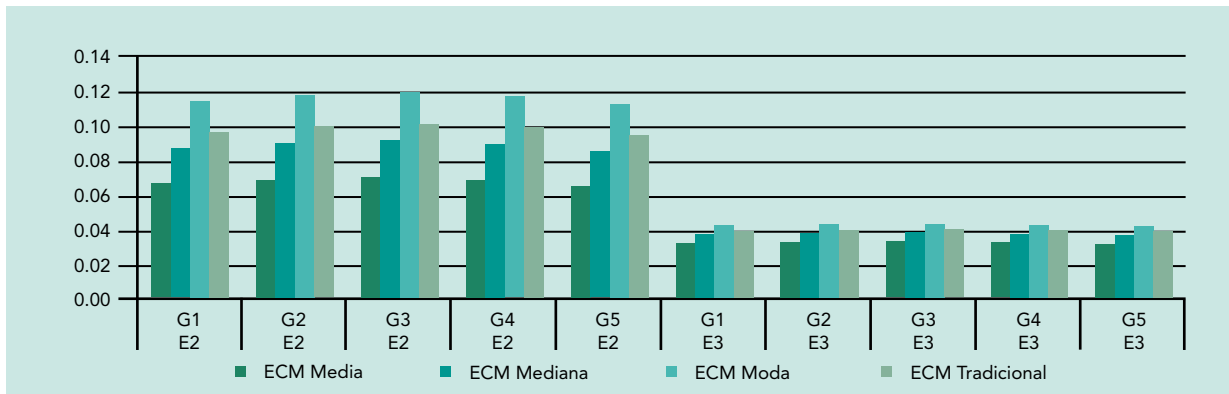


Figura 8. Valores ECM para los estimadores de δ_j encontrados vía simulación en los escenarios 2 y 3 usando Jeffreys como a priori.

IGA		Media		Mediana		Moda		Método
Escenario	Genotipo	Flat	Jeffreys	Flat	Jeffreys	Flat	Jeffreys	Tradicional
E1	G1	54,23	54,24	54,24	54,23	55,02	54,48	54,22
E1	G2	49,9	49,89	49,92	49,9	50,51	50,2	49,9
E1	G3	51,04	51,05	51,06	51,05	51,59	51,41	51,05
E1	G4	51,26	51,21	51,24	51,2	51,67	51,5	51,21
E1	G5	55,85	55,86	55,86	55,84	56,29	56,09	55,85
E2	G1	13,96	13,95	13,96	13,95	14,03	14,05	13,96
E2	G2	12,93	12,93	12,93	12,93	13,04	12,99	12,93
E2	G3	13,22	13,23	13,22	13,23	13,36	13,3	13,23
E2	G4	13,23	13,23	13,23	13,23	13,35	13,32	13,23
E2	G5	14,23	14,22	14,22	14,22	14,26	14,32	14,22
E3	G1	3,51	3,52	3,51	3,52	3,55	3,54	3,51
E3	G2	3,26	3,26	3,26	3,26	3,3	3,28	3,26
E3	G3	3,33	3,33	3,33	3,33	3,35	3,36	3,33
E3	G4	3,33	3,33	3,33	3,33	3,36	3,35	3,33
E3	G5	3,56	3,56	3,56	3,56	3,59	3,59	3,56

Tabla 2. Valores de ECM ($\times 10^3$) para β_0 comparando distribuciones a priori Flat vs. Jeffreys.

escenarios y genotipos. De modo que es preferible emplear la a priori de Jeffreys, poniendo especial atención en las desviaciones a la regresión, porque si son altas la inferencia deja de ser precisa.

Los datos reales de campo podrán ampliar la discusión entorno a cual a priori es más indicada para evaluar los datos de pruebas regionales.

IGA		Media		Mediana		Moda		Método
Escenario	Genotipo	Flat	Jeffreys	Flat	Jeffreys	Flat	Jeffreys	Tradicional
E1	G1	54,23	54,24	54,24	54,23	55,02	54,48	54,22
E1	G2	49,9	49,89	49,92	49,9	50,51	50,2	49,9
E1	G3	51,04	51,05	51,06	51,05	51,59	51,41	51,05
E1	G4	51,26	51,21	51,24	51,2	51,67	51,5	51,21
E1	G5	55,85	55,86	55,86	55,84	56,29	56,09	55,85
E2	G1	13,96	13,95	13,96	13,95	14,03	14,05	13,96
E2	G2	12,93	12,93	12,93	12,93	13,04	12,99	12,93
E2	G3	13,22	13,23	13,22	13,23	13,36	13,3	13,23
E2	G4	13,23	13,23	13,23	13,23	13,35	13,32	13,23
E2	G5	14,23	14,22	14,22	14,22	14,26	14,32	14,22
E3	G1	3,51	3,52	3,51	3,52	3,55	3,54	3,51
E3	G2	3,26	3,26	3,26	3,26	3,3	3,28	3,26
E3	G3	3,33	3,33	3,33	3,33	3,35	3,36	3,33
E3	G4	3,33	3,33	3,33	3,33	3,36	3,35	3,33
E3	G5	3,56	3,56	3,56	3,56	3,59	3,59	3,56

Tabla 3. Valores de ECM ($\times 10^3$) para β_1 , comparando distribuciones a priori Flat vs. Jeffreys.

IGA		Media		Mediana		Moda		Método
Escenario	Genotipo	Flat	Jeffreys	Flat	Jeffreys	Flat	Jeffreys	Tradicional
E1	G1	2,0123	0,4248	0,6426	0,2503	0,5368	0,2390	0,2270
E1	G2	1,8030	0,3812	0,5713	0,2365	0,4947	0,2418	0,2208
E1	G3	1,7502	0,3737	0,5572	0,2369	0,4621	0,2456	0,2233
E1	G4	1,8650	0,3958	0,5945	0,2430	0,4971	0,2425	0,2252
E1	G5	2,1107	0,4461	0,6771	0,2575	0,5706	0,2389	0,2306
E2	G1	0,0691	0,0670	0,0589	0,0871	0,0694	0,1146	0,0965
E2	G2	0,0638	0,0694	0,0602	0,0904	0,0715	0,1182	0,0999
E2	G3	0,0636	0,0708	0,0614	0,0919	0,0728	0,1192	0,1014
E2	G4	0,0660	0,0692	0,0603	0,0899	0,0707	0,1173	0,0993
E2	G5	0,0717	0,0661	0,0584	0,0858	0,0689	0,1130	0,0951
E3	G1	0,0205	0,0326	0,0294	0,0378	0,0322	0,0430	0,0397
E3	G2	0,0213	0,0335	0,0302	0,0385	0,0330	0,0437	0,0404
E3	G3	0,0217	0,0338	0,0306	0,0388	0,0334	0,0439	0,0407
E3	G4	0,0212	0,0333	0,0301	0,0384	0,0329	0,0435	0,0403
E3	G5	0,0202	0,0323	0,0290	0,0375	0,0319	0,0428	0,0394

Tabla 4. Valores de ECM para δ_1 , comparando distribuciones a priori Flat vs. Jeffreys

Datos de pruebas regionales.

Las densidades de probabilidad de los parámetros brindan información acerca de cuán preciso es el cálculo de los estimadores en el caso bayesiano, de modo que permitirá comparar las distribuciones *a priori* que pautan la probabilidad para las estimaciones de los parámetros de estabilidad.

Se muestran en la figura 9 las densidades de probabilidades usando diferentes *a priori*, es evidente que la distribución *Flat* resulta mejor que *Jeffreys* para hacer estimaciones de β_0 , por ser menos dispersa.

La densidad de probabilidad para el parámetro β_1 (figura 10) refleja que las estimaciones usando la distribución no informativa *a priori*

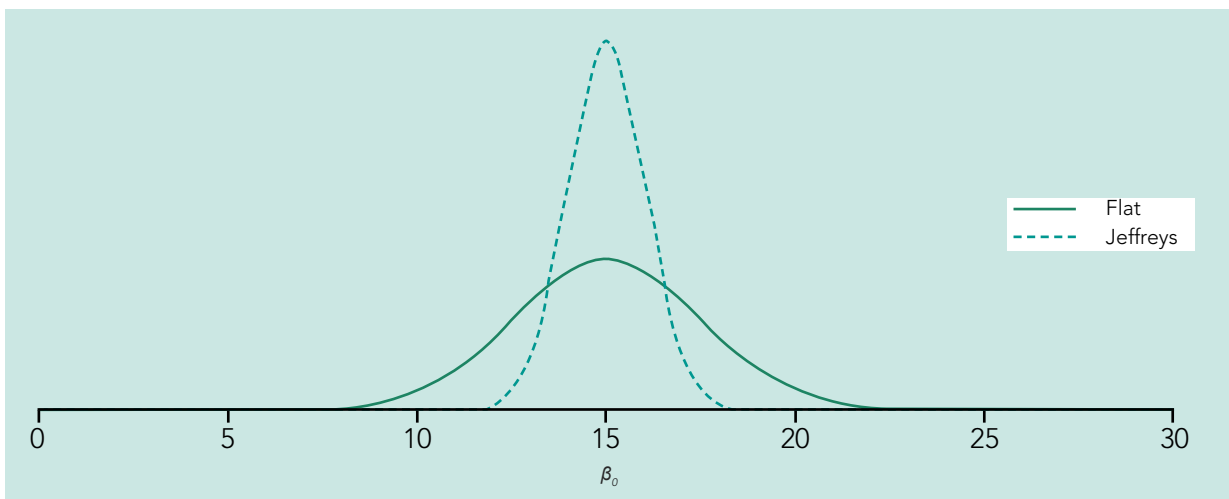


Figura 9. Densidad de probabilidad *a posteriori* del parámetro β_0 (kg ha⁻¹) usando las distribuciones *a priori* Flat y Jeffreys para el cultivar CMS

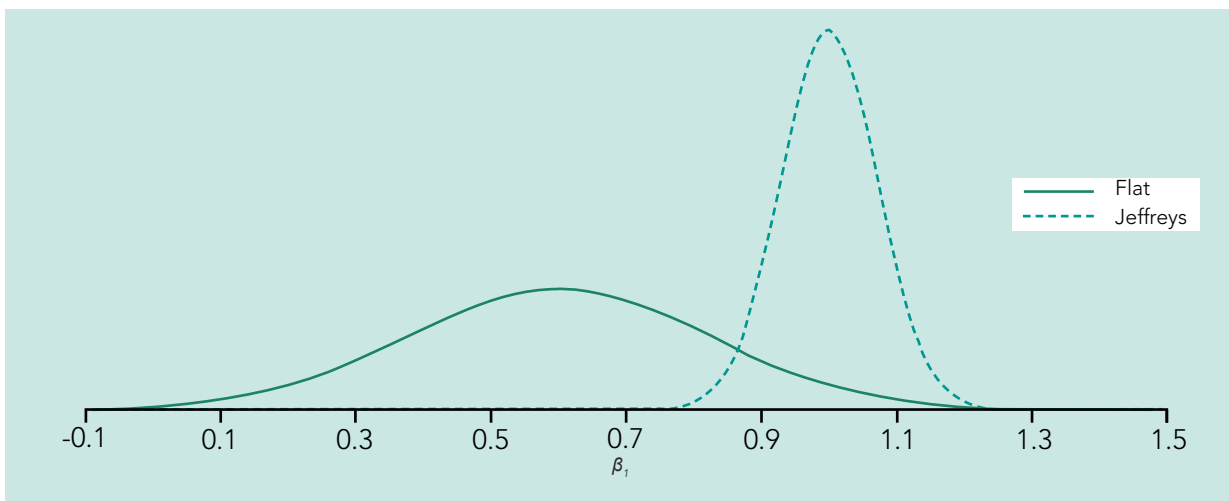


Figura 10. Densidad de probabilidad *a posteriori* del parámetro β_1 usando las distribuciones *a priori* Flat y Jeffreys para el cultivar CMS

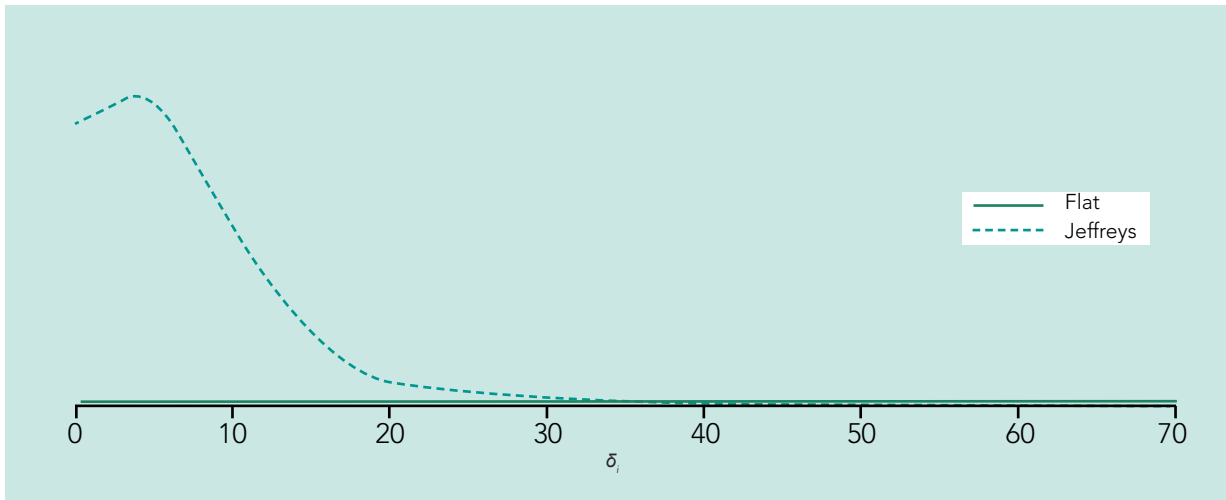


Figura 11. Densidad de probabilidad a posteriori del parámetro δ_i usando las distribuciones a priori Flat y Jeffreys para el cultivar CMS

de *Jeffreys* es más apuntada (Leptocúrtica) que la de *Flat*, por tanto infiere sobre la media de la distribución a priori con mayor precisión.

La distribución a priori de *Jeffreys* también sobresale al compararse con la *Flat* usando datos reales pues estima con menor insesgamiento los valores de δ_i , esto se ve reflejado en la densidad de probabilidad a priori, que se caracteriza por ser menos dispersa.

La figura 14 muestra la densidad de probabilidad del parámetro δ_i , se observa que ambas distribuciones se diferencian por su dispersión, de modo que la a priori *Flat* es más dispersa en comparación con la a priori de *Jeffreys*, la densidad de probabilidad usando la a priori *Flat* muestra un mejor ajuste a los valores de δ_i marcando una región estrecha con mayor densidad de probabilidad.

CONCLUSIONES

Los métodos estadísticos de regresión son muy potentes y sensibles, un desarrollo metodológico Bayesiano se convierte en una alternativa precisa y confiable para estimar los parámetros comúnmente evaluados en los modelos de estabilidad Fenotípica, especialmente cuando se buscan genotipos de interés agronómico, es decir que sean estables y altamente predecibles.

Los eventos de simulación y los datos reales respaldan la metodología Bayesiana además que proponen el uso de distribuciones a priori

Los métodos estadísticos de regresión son muy potentes y sensibles, un desarrollo metodológico Bayesiano se convierte en una alternativa precisa y confiable para estimar los parámetros comúnmente evaluados en los modelos de estabilidad Fenotípica.

de *Jeffreys* para la estimación de los parámetros de estabilidad y adaptabilidad ya que la densidad de la distribución *a posteriori* es más ajustada y precisa al verdadero valor del parámetro.

Después de analizar los tres parámetros se observó que las distribuciones *a priori* afectan fuertemente la distribución *a posteriori*, aún siendo ambas no informativas, y además se observó

que la *a priori* de *Jeffreys* es más ajustada y precisa para inferir los parámetros de estabilidad.

Desde el punto de vista de toma de decisiones, la metodología Bayesiana a través de las regiones de credibilidad o HPD brinda una ventaja considerable ya que calcula los intervalos más pequeños posibles, lo que refleja un mejor ajuste al verdadero valor del parámetro.

BIBLIOGRAFÍA

1. Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society of London, 53: 370-418.
2. Box, G.E.P. y G.C. Tiao. 1973. Bayesian inference in statistical analysis. Wiley Classical library. New York. 588p.
3. Cotes Torres, J. M. 2004. Análise bayesiana da interação genótipo-ambiente na seleção de genótipos por estabilidade fenotípica e por rendimento-estabilidade. 63h. Tesis Doctoral en Agronomía (Genética y Mejoramiento de Plantas). Universidade Estadual Paulista. Faculdade de Ciências Agrárias e Veterinárias. 76 p.
4. Cotes J.M., J. Crossa, A. Sanches y Cornelli P.L. A Bayesian Approach for Assessing the Stability of Genotypes. Crop Science, 46: 2654-2665
5. Crossa, J. 1990. Statistical analyses of multilocation trials. Advances in Agronomy, 44: 55-85.
6. Cruz, C. D. y A. J. Regazzi. 1994. Modelos biométricos aplicados ao melhoramento genético. Imprensa Universitária UFV. Viçosa – MG. 390p.
7. Eberhart S. A. y W.A. Russell. 1966. Stability parameters for comparing varieties. Crop Science, 6: 36-40.
8. Edwards J. W. y J.L. Jannink. 2006. Bayesian modeling of heterogeneous error and genotype x environment interaction variances. Crop Science, 46: 820-833.
9. Finlay, K. W. y F.N. Wilkinson. 1963. The analysis of adaptation in a plant-breeding programme. Australian Journal Agricultural Resource, 14: 742-754.
10. Hühn, M. 1996. Nonparametric analysis of genotype x environment interactions by ranks. p. 235-271. En: Kang, M. S. y H. G. Gauch (Eds.). Genotype by environment interaction. CRC Press. New York. 416p.
11. Jeffreys, H. 1961. Theory of Probability. Tercera Edición. Clarendon Press. New York. 459p.
12. Kang, M. S. y R. Magari, R. 1996. New development in selecting for phenotypic stability in crop breeding. p. 1-14. En: Kang, M. S. y H. G. Gauch (Eds.). Genotype by environment interaction. CRC Press. New York. 416p.
13. Piepho, H. P. Analysis of genotype-by-environment interaction and phenotypic stability. p. 152-174. En: Kang, M. S. y H. G. Gauch (Eds.). Genotype by environment interaction. CRC Press. New York. 416p.
14. Poehlman A. 2003. Mejoramiento genético de las cosechas. Editorial Limusa. México. 459p.
15. SAS Institute Inc. 1999. SAS/IML. User's Guide. Version 8. SAS Institute Inc. Cary NC. 846p.
16. Searle, S. R., G. Casella y C.E. McCulloch. 1992. Variance components. John Wiley & Sons, Inc. New York. 501p

17. Vallejo F. A. y E. I. Salazar. 2002. Mejoramiento genético de plantas. Editorial Feriva S.A. Cali. 350p.
18. Wolfinger, R. D.; R. E. Kass. 2000. Non-conjugate Bayesian analysis of variance component models. *Biometrics*, 56: 768-774.

CONSULTA VIRTUAL

19. Cotes, J. M.; A. Sánchez y E.P. González. 2010. Introducción al análisis Bayesiano en el mejoramiento de plantas. 54p. <http://www.docentes.unal.edu.co/jmcotes/docs/>, consulta noviembre 2010
20. SAS Institute Inc. 2000. SAS Macros and Functions. SAS OnLineDoc. Version 8, 20p. <http://www.sas.com/apps/demosdownloads/setupcat.jsp?cat=SAS+OnlineDoc>, consulta Junio 2006.
21. SAS Institute Inc. 2006. SAS SYSTEM VIEWER, Software Downloads. <http://www.sas.com/apps/demosdownloads/setupcat.jsp?cat=SAS+System+Viewer>, consulta 1 Junio 2006.

APÉNDICE 1. Macro programada en SAS para la simulación de los escenarios.

```
proc iml;
*reset print;
start simular_ER(medias,beta,sigmai,indices,out1,ont);
*****INPUT: Entrada de datos*****
* Es necesario indicar los beta (o Adaptabilidad), los
  índices ambientales*
* y los sigma(o estabilidad) de cada genotipo *
*****;

****Obtención de información básica;
ngen=nrow(beta);
nenv=nrow(indices);
nt=ngen*nenv;

****Construcción de la matriz de diseño X;
*b0=diag(medias)@j(nenv,1,1);
b0=i(ngen)@j(nenv,1,1);
b1=i(ngen)@indices;
X=b0||b1;

****Obtención de los errores aleatorios de una distribución Normal;
```

```
z=j(nt,1,0);
do i=1 to nt;
  z[i]=rannor(1234);
end;
tmp=sigmai@j(nenv,1,1);
e=z#tmp;

b=medias//beta;
****Construcción del modelo Eberhart y Russell;
y=X*b+e;

****Se generan los valores para el archivo de salida;
gen=t(1:ngen)@j(nenv,1,1);
env=j(ngen,1,1)@t(1:nenv);

out1=gen||env||y;

ont=nt;
finish;
store module=_all_;
quit;

*=====MACRO sim_er=====
=====
Es necesario indicar los beta (o Adaptabilidad), los índices
ambientales y los sigma(o estabilidad) de cada genotipo
=====;
=====;
```

APÉNDICE 2. Macro programada en SAS para el análisis tradicional y bayesiano basada en el procedimiento Mixed.

```
*=====MACRO ER_Ba-
yes=====
Recibe la información de la Macro hpd y la procesa para
cada experimento
=====
=====;

%macro ER_Bayes(data,nextpos,ngen, prior);
libname Camilo "C:\Camilo\&data";
proc delete data= Camilo.tradicionalBetas Camilo.tradi-
cionalSigmai
Camilo.BayesBetas Camilo.BayesSigmai;
run;
quit;
%do expto=1 %to &nextpos;
Data work;
```

```

set Camilo.&data;
where expto=&expto;
run;

proc means data=work noprint;
class env;
var y;
output out=prom mean=my;

run;
data
mediast(where=(l_type_l=1))
mediast(where=(l_type_l=0) rename=(my=mt));
set prom;
i=1;

run;
data indice;
merge mediasg(keep=my env i) me-
diast (keep=mt i);
by i;
Indice=my-mt;
keep env indice;

run;
proc sort data=work out=work; by env gen; run;
data work;
merge work indice;
by env;
array g (&ngen) g1-g&ngen;
do i=1 to &ngen;
if gen=i then g{i}=1; else g{i}=0;
end;

run;

proc mixed data=work cl;
class gen;
model y=g1-g&ngen gen*indice/no-
int solution cl;
repeated /group=gen;
prior &prior /out=posterior nsample=10000;
ods output SolutionF=SolutionF;
ods output CovParms=CovParms;
by expto;

run;

proc append base=Camilo.tradicionalBetas
data=SolutionF;
run;
quit;

proc append base=Camilo.tradicionalSigmaI
data=CovParms;
run;
quit;

&end;
&mend ER_Bayes;

```

APÉNDICE 3. Macro programada en SAS para la obtención de los intervalos HPD.

```

*=====MACRO hpd=====
Calcula los intervalos de los intervalos de Highest Poste-
rior Density
=====
;

%macro hpd(nome,ngen,title,options,salida);
*ods rtf exclude all;
proc delete data=&nome._IC;run;
proc delete data=&nome._den;run;
data &nome._den;
run;
%do i=1 %to &ngen;

*Estimador Kernel de la densidad de probabilidad;

proc kde data=BayesEstimator out=Bayes_Kde ng=200
bwm=1 Levels=100 10 &options;
var &nome&i;
ods output Levels=Levels;
ods output statistics=Statistics(where=(Descr=
'Mean') rename=(&nome&i=Bayes_Mean));
ods output Percentiles=Percentiles(where=(Per-
cent=50) rename=(&nome&i=Bayes_Median));
by expto;

run;
data Levels; Informat Variavel $12.; merge levels(where=
(percent=100) rename=(lower=Bayes_Mode))
levels(where=(percent=10)) statistics(drop=Descr)
Percentiles(drop=Percent);
Variavel="&nome&i";
proc append base=&nome._IC data=Levels(drop=density
percent);
run;
proc delete data=levels statistics percentiles;
data &nome._den;
merge &nome._den Bayes_kde(rename=(Density=Den_
&nome&i));
run;
&end;
*ods rtf select all;
proc print data=&nome._IC;
title "Estimativas de Bayes e HPD";
title2 "&title";
run;
title;

Data Camilo.&nome._den&expto;
set &nome._den;

run;

&mend hpd;

```