

# SELECCIÓN DE GENOTIPOS CON ALTA RESPUESTA Y ESTABILIDAD FENOTÍPICA EN PRUEBAS REGIONALES: RECUPERANDO EL CONCEPTO BIOLÓGICO

Fecha de recepción: 14 de agosto de 2012 • Fecha de aceptación: 4 de octubre de 2012

## GENOTYPE SELECTION WITH BOTH HIGH PHENOTYPIC RESPONSE AND STABILITY THROUGH REGIONAL TRIALS: RECOVERING THE BIOLOGICAL CONCEPT

José Miguel Cotes Torres<sup>1,4</sup> • Elena Paola González Jaimes<sup>2</sup> • Alejandro Cotes Torres<sup>3</sup>

### RESUMEN

En los programas de mejoramiento genético de plantas es común evaluar la respuesta de los genotipos en varios ambientes con el fin de valorar la estabilidad fenotípica. Los primeros modelos para evaluar la estabilidad fueron llamados de varianza ambiental y se soportan en un modelo jerárquico que contempla el efecto principal de los genotipos, y el efecto de los genotipos dentro de los ambientes, lo cual se denomina el concepto biológico de la estabilidad fenotípica. Una de las grandes preocupaciones de los mejoradores es poder encontrar líneas, variedades e híbridos, con alta estabilidad y respuesta fenotípica. Con ese objetivo se han desarrollado numerosos métodos de análisis que involucran un modelo de dos vías con interacción; es decir: el efecto del genotipo, del ambiente y de la interacción genotipo ambiente. Se argumenta que estos modelos están más acorde con los intereses agronómicos que los modelos inicialmente usados, pero este artículo muestra como el modelo de varianza ambiental es un modelo más general y preferible para evaluar la estabilidad de la respuesta fenotípica, y con base en éste, a través de la metodología bayesiana, se desarrolla y propone un parámetro que no sólo permite la selección de genotipos con alta estabilidad y respuesta fenotípica, sino que además permite la incorporación de información previa de los genotipos testigos usados en las pruebas regionales.

**Palabras clave:** Pruebas multiambiente, selección simultánea, estabilidad fenotípica, algoritmo de GIBBS.

1 Profesor Asociado. PhD. Universidad Nacional de Colombia, Sede Medellín. Facultad de Ciencias Agrarias. Departamento de Agronomía. Carrera 59A # 63 -20, Medellín, Colombia

2 Profesor Asistente. PhD. Politécnico Colombiano Jaime Isaza Cadavid. Facultad de Ciencias Agrarias. Cr 48 # 7-151 Medellín, Colombia.

3 Profesor Asociado. PhD. Universidad Nacional de Colombia, Sede Bogotá. Facultad de Medicina Veterinaria y de Zootecnia. Departamento de Ciencias para la Producción Animal. Ciudad Universitaria, Edificio 561, Bogotá D.C., Colombia

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

## ABSTRACT

Plant breeding programs often evaluate genotype response across environments in order to assess phenotypic stability. Early models evaluating stability were called environmental variance models, and are based on a hierarchical model that includes the main effect of the genotype and the effect of the genotype within the environment, which is called the biological concept of phenotypic stability. One of the major concerns of plant breeders is to find lines, varieties, hybrids, etc., with both high stability and phenotypic response. With that goal in mind, several methods of analysis have been developed involving a two-way model with interaction; that is: the effect of the genotype, the effect of the environment, and genotype-environment interaction. It has been stated that these models are more in accordance with agronomic interests than the earlier models, but this article shows that the environmental variance model is a more general and convenient model to assess the stability of the phenotypic response. Also, based on the environmental variance model, and developed using Bayesian methodology, a parameter that not only allows selection of genotypes with both high stability and phenotypic response, but also allows to include previous information of reference genotypes used in regional trials is proposed.

**Keywords:** multi-environment trials, simultaneous selection, phenotypic stability, Gibbs algorithm.

---

## INTRODUCCIÓN

En mejoramiento genético de plantas es ampliamente conocido que la respuesta observable, es decir el fenotipo, es modelada en función del efecto genético, del efecto ambiental y de la interacción del genotipo con el ambiente (Falconer y Mackay, 1996). Este modelo teórico se ha utilizado en el análisis de pruebas regionales a través de modelos estadísticos que buscan la estimación de los tres efectos no observables, mencionados anteriormente. Posterior a su estimación, se han desarrollado metodologías de análisis que permitan seleccionar genotipos en presencia de la interacción genotipo por ambiente. Si esta es muy alta, se hace necesario seleccionar genotipos para un pequeño número de ambientes en cuyo caso se denomina que el genotipo tiene adaptación específica. En caso contrario, un genotipo puede ser seleccionado en un gran número de ambientes por lo que se le atribuye al genotipo una adaptabilidad amplia (Poelman y Allen, 2003).

Históricamente la interacción genotipo por ambiente es una de las principales preocupaciones de los mejoradores por dos razones (Kang y Magari, 1996): primero porque reduce el progreso de selección (Cooper y DeLacy, 1994; DeLacy *et al.*, 1996) y segundo porque hace imposible interpretar los efectos principales (debidos exclusivamente a los genotipos o al ambiente) (Kang y Magari, 1996). Esta interacción ocurre tanto en el corto plazo (tres o cuatro años de pruebas en una localidad) como en el largo plazo (varios años y varias localidades). Usualmente no se implementa el análisis de la interacción genotipo por ambiente en pruebas realizadas en el corto plazo, sin embargo en los últimos años esta tendencia ha cambiado (Kang y Pham, 1991), ya que se persiguen objetivos distintos más acordes con las limitaciones prácticas de las pruebas regionales en un programa de mejoramiento. Así se busca mejorar los sistemas de selección de

genotipos con el fin de obtener costos aceptables en la realización de dichas pruebas y poder realizar una adecuada selección en un tiempo razonable.

En este contexto se han desarrollado numerosas y variadas metodologías de análisis estadístico para el análisis de pruebas regionales que abarcan: a) análisis de regresión sugeridos inicialmente por Yates y Cochran (1938) y posteriormente desarrollados por Finlay y Wilkinson (1963), Eberhart y Russell (1966) y Perkins y Jinks (1968), además de los análisis propuestos por Plaisted y Peterson (1959), Tai (1971) y Wescott (1987); b) análisis de componentes de varianza propuestos por Shukla (1972), Becker (1981); c) análisis de patrones (DeLacy *et al.*, 1990), que pueden ser por agrupación (clasificación) y por ordenación, incluyendo estos últimos los denominados modelos AMMI (Additive Main Effects Multiplicative Interaction), SREG (Sites Regression), GREG (Genotypes Regression), COMM (Complete Multiplicative Modelling) desarrollados e implementados principalmente por Gauch y Zobel (1988), Gauch y Zobel (1996), Cornelius *et al.* (1996), Nachit *et al.* (1992), Yau

para los efectos de la interacción genotipo por ambiente propuestos por Piepho (1998), Smith *et al.* (2005), Kelly *et al.* (2007), Burgueño *et al.* (2008); f) análisis de rendimiento y estabilidad propuestos por Kang (1988), Kang (1991), Nassar *et al.* (1994); g) análisis bayesianos desarrollados por Besag y Higdon (1999), Foucteau y Dennis (2001), Theobald *et al.* (2002), Edwards y Jannick (2006), Cotes *et al.* (2006), Crossa *et al.* (2011).

Este elevado número de metodologías generan desconcierto y confusión para los mejoradores de plantas que, finalmente, no disponen de una herramienta confiable para el análisis de pruebas regionales, y el escoger una u otra metodología se hace con argumentos empíricos más que teóricos. Según DeLacy *et al.* (1996), ante la gran diversidad de métodos para el análisis de la interacción genotipo por ambiente, algunos mejoradores han optado por utilizar varios de ellos, con el fin de mejorar la eficiencia de la selección. Sin embargo, ello requiere gran cantidad de tiempo, mientras que las decisiones de selección deben ser tomadas en "tiempo real".

El elevado número de metodologías generan desconcierto y confusión para los mejoradores de plantas que, finalmente, no disponen de una herramienta confiable para el análisis de pruebas regionales, y el escoger una u otra metodología se hace con argumentos empíricos más que teóricos.

(1995), Crossa *et al.* (2004), Gauch (2006), Yan *et al.* (2007), Gauch *et al.* (2008), Yang *et al.* (2009); d) análisis no paramétricos propuestos por Nassar *et al.* (1994) y Hühn (1996); e) modelación tipo FA (Factor Analytic) de la estructura de varianzas y covarianzas

Los grupos de Investigación en Papa de la Universidad Nacional de Colombia - Sede Bogotá y de Mejoramiento y producción de especies andinas y tropicales de la Sede Medellín, al igual que muchos grupos de investigación en mejoramiento genético

han desarrollado numerosas pruebas regionales para la evaluación de sus materiales de mejoramiento. En ellas se ha implementado algunas de las metodologías de análisis de la interacción genotipo por ambiente mencionadas anteriormente y por razones empíricas se empezó a utilizar la metodología de rendimiento y estabilidad fenotípica propuesta por Kang (1991), ya que con ella se percibía que era aquella metodología que reflejaba mejor el comportamiento de los genotipos en las pruebas regionales. Sin embargo, en los cinco últimos años, ambos grupos de investigación han ampliado el espectro de características objetivo de mejoramiento tales como resistencia a varias enfermedades y varias variables de calidad de procesamiento, además de la tradicional característica de rendimiento que han desafiado los modelos de análisis de la interacción genotipo por ambiente, ya que en algunos casos genotipos muy promisorios no eran seleccionados a través de los análisis realizados. Por otro lado, en las variedades testigo utilizada, se disponía información de más de 15 años que no podía ser utilizada en los análisis de las nuevas pruebas. Esto hizo necesario replantear el modelo utilizado y desarrollarlo bajo la metodología de estimación bayesiana. Por lo anteriormente descrito, el presente artículo expone inicialmente el modelo estadístico que consideramos, bajo ciertas condiciones, debe ser utilizado en el análisis de pruebas regionales con el objetivo de seleccionar genotipos, posteriormente se presenta una forma de seleccionar genotipos con una alta respuesta y estabilidad fenotípicas, y finalmente se desarrolla la parte de estimación bayesiana del modelo considerado.

### **MODELO CONSIDERANDO LA INTERACCIÓN GENOTIPO-AMBIENTE**

En algunas variables que presentan herencia cuantitativa como el caso de resistencia a enfermedades, se encuentran al menos tres grandes grupos de

genotipos, los altamente resistentes, altamente susceptibles y los de resistencia intermedia. En los dos primeros casos, en ensayos multiambiente en papa realizados en los Departamentos de Cundinamarca, Boyacá y Antioquia, se observaba que estos genotipos tenían una respuesta consistente a lo largo de los ambientes (Cotes *et al.*, 2012), es decir, un genotipo resistente que desarrollaba la enfermedad a niveles muy bajos (menores del 1%), su respuesta era la misma en todos los ambientes, lo cual ocurría también con genotipos extremadamente susceptibles. Sin embargo, los genotipos con resistencia intermedia su respuesta variaba a lo largo de los ambiente presentando diferentes grados de expresión de la enfermedad dependiendo del ambiente evaluado.

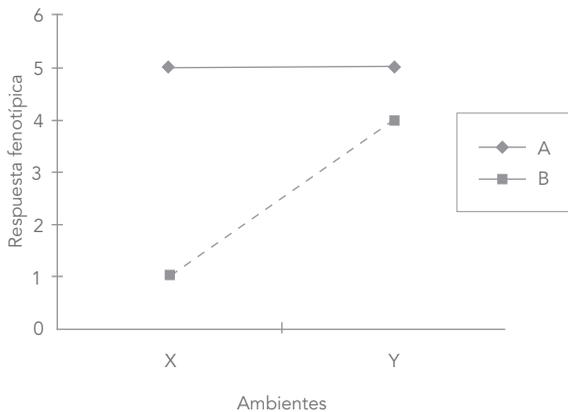
De igual forma, en otro estudio, evaluando la capacidad de producir hojuelas de papa aptas para fritura, ocurría un fenómeno similar al previamente mencionado (Ñustez, 2011); genotipos promisorios de papa para fritura sus respuesta en varios ambientes era inmodificable, es decir, exactamente la misma, al igual que los genotipos con poca habilidad para producir hojuelas para fritura. Sin embargo, genotipos con calidad de hojuela intermedia, mostraban variaciones de su respuesta, dependiendo del ambiente donde era probado el genotipo.

Estas respuestas consistentes, casi inmodificables, a lo largo de varios ambientes se pueden traducir en genotipos estables, mientras que la variación de la respuesta de los genotipos en los ambientes se puede categorizar como una respuesta inestable. Así, los genotipos estables serían deseables, siempre y cuando su respuesta fenotípica sea de interés, es decir, genotipos resistentes y estables en el caso de la valoración del grado de enfermedad, y genotipos aptos para fritura y estables en el caso de papas con dicha finalidad.

Sin embargo, el uso preliminar de la metodología de Kang (1991) en estos casos, la cual junta rendimiento y estabilidad, no seleccionaba dichos

genotipos, inclusive según la estimación de la varianza de estabilidad fenotípica Shukla (1972), que es utilizada por el método semiparamétrico de Kang (1991), reportaba algunos de esos genotipos deseables por su respuesta consistente entre los ambientes como inestables. Esta situación hizo reflexionar que tal vez el problema no era la metodología utilizada sino el modelo estadístico utilizado, el cual contempla un componente denominado interacción genotipo-ambiente.

Para conceptualizar qué es lo que ocurre en este caso analizaremos el siguiente ejemplo en un caso de interacción genotipo por ambiente 2 x 2 (Figura 1).



**Figura 1.** Interacción genotipo por ambiente en un análisis de 2 x 2. El genotipo A es invariante al ambiente mientras que el genotipo B cambia su respuesta en función del ambiente.

El genotipo A tiene una respuesta fenotípica consistente tanto en el ambiente X como en el Y, mientras que el genotipo B varía su respuesta de un ambiente para otro. Consideremos el tradicional modelo de interacción genotipo-ambiente, sin repeticiones:

$$y_{ij} = \mu + v_i + \alpha_j + (v\alpha)_{ij} \quad [1]$$

donde,  $y_{ij}$  es la respuesta fenotípica del genotipo  $i$  ( $i=1,2,\dots,v$ ) en el ambiente  $j$  ( $j=1,2,\dots,a$ ) es la media general,  $v_i$  es el efecto del genotipo  $i$ ,  $\alpha_j$  es el efecto del ambiente  $j$ , y  $(v\alpha)_{ij}$  es el efecto

de la interacción del genotipo  $i$  en el ambiente  $j$ . Los efectos de interacción genotipo-ambiente debidos al genotipo A están dados por las siguientes expresiones:

$$(v\alpha)_{AX} = y_{AX} - \mu - v_A - \alpha_X = 5 - \mu - v_A - \alpha_X$$

$$(v\alpha)_{AY} = y_{AY} - \mu - v_A - \alpha_Y = 5 - \mu - v_A - \alpha_Y$$

y para el genotipo B por las siguientes expresiones.

$$(v\alpha)_{BX} = y_{BX} - \mu - v_B - \alpha_X = 1 - \mu - v_B - \alpha_X$$

$$(v\alpha)_{BY} = y_{BY} - \mu - v_B - \alpha_Y = 4 - \mu - v_B - \alpha_Y$$

Como se puede observar los efectos de interacción genotipo-ambiente debidos al genotipo A solo varían debido al efecto de  $\alpha_j$ , es decir, del ambiente, lo cual no es un atributo del genotipo. Así el genotipo A que tiene una respuesta consistente, es decir estable, a lo largo de los ambientes, aparecerá como inestable por cualquier metodología de análisis estadístico de pruebas regionales que se base en los efectos de  $(v\alpha)_{ij}$ . En contraste, los efectos de interacción genotipo-ambiente para el genotipo B varían debido a la respuesta fenotípica  $y_{ij}$  y a los efectos de  $\alpha_j$ , por lo que solamente una parte del efecto  $(v\alpha)_{ij}$  es debida a la expresión del genotipo. De esta forma cualquier metodología que mida la estabilidad fenotípica basada en el efecto  $(v\alpha)_{ij}$  no sería apropiada para tal fin ya que genotipos deseables como el genotipo A podrían no ser seleccionados.

## MODELO PARA EVALUAR LA RESPUESTA FENOTÍPICA EN VARIOS AMBIENTES

Una solución para el problema expuesto anteriormente para el modelo de interacción genotipo por ambiente es considerar un modelo jerárquico

envés de un modelo de dos vías con interacción. Así, para el ejemplo de la Figura 1, podemos considerar el siguiente modelo:

$$y_{ij} = \mu + v_i + \gamma_{ij}$$

donde,  $y_{ij}$ ,  $\mu$ ,  $v_i$  son como definidos anteriormente en el modelo [1] y  $\gamma_{ij}$  es el efecto del genotipo  $i$  dentro del ambiente  $j$ . Así, los efectos de la respuesta de genotipo A al ambiente están dados por las siguientes expresiones:

$$\gamma_{AX} = y_{AX} - \mu - v_A = 5 - \mu - v_A$$

$$\gamma_{AY} = y_{AY} - \mu - v_A = 5 - \mu - v_A$$

y para el genotipo B por las siguientes expresiones.

$$\gamma_{BX} = y_{BX} - \mu - v_B = 1 - \mu - v_B$$

$$\gamma_{BY} = y_{BY} - \mu - v_B = 4 - \mu - v_B$$

Como se puede observar los efectos del genotipo A en los diferentes ambientes es el mismo, teniendo una variación nula, lo cual está acorde con lo observado en la expresión fenotípica de este genotipo la cual es consistente tanto en el ambiente X como en el ambiente Y. En contraste, los efectos del genotipo B en los diferentes ambientes es variable, pero esta variación se debe exclusivamente a la respuesta fenotípica y no a otro tipo de efectos como ocurre en el caso del modelo con interacción.

El modelo [2] es el denominado de variancia ambiental (Lin *et al.*, 1986; Becker y Leon, 1988; Piepho, 1999) y es acorde el denominado concepto biológico de la estabilidad (Becker, 1981), al concepto Tipo 1 u homeostasis (Lin *et al.*, 1986), y al concepto estático de la estabilidad (Becker y Leon, 1988; Piepho, 1996).

Según Becker (1981), este concepto de biológico de la estabilidad fue ampliamente aplicado por los

mejoradores de plantas desde 1917, pero posteriormente reemplazado por el concepto agronómico o dinámico basados en el modelo [1], donde la estabilidad de los genotipos es relativa al potencial del ambiente basada en la media de los genotipos en cada ambiente.

Sin embargo, aquí surge otra dificultad con el modelo [1]. Este potencia del ambiente es medido a través de un índice ambiental el cual es una función de  $\alpha_j$  en el modelo de dos vías con interacción (Becker y Leon, 1988), es decir, el modelo [1]. Lo anterior implica que la calificación del ambiente depende del desempeño de los genotipos que allí se evaluaron, es decir la evaluación del efecto ambiental, no es independiente del desempeño de los genotipos. Pero ¿qué ocurre si parte o la totalidad de los genotipos cambia? Si eso ocurre, entonces el índice ambiental cambia y por consiguiente la calificación de estabilidad fenotípica de los genotipos basada en  $(v\alpha)_{ij}$  del modelo [1]. Es decir, en el modelo [1], un genotipo es estable o inestable, dependiendo de los demás genotipos que ingresan en las pruebas regionales, por lo que la estabilidad se convierte en un atributo de las pruebas regionales y no en un atributo *per se* del genotipo. En el modelo [2] estos problemas no ocurren debido a que el efecto del ambiente no se mide independiente del efecto del genotipo en cada ambiente, y los efectos  $\gamma_{ij}$  para el genotipo  $i$  son útiles para evaluar la estabilidad fenotípica de un determinado genotipo, convirtiéndose en este caso en un atributo *per se* del mismo.

Una solución a este último problema mencionado para el modelo [1] es considerar el efecto del ambiente como una función lineal de covariables ambientales (p.e. temperatura, humedad, días de sequía, entre otros), como lo realizado por Theobald *et al.* (2002) quienes usaron como covariable los grados día térmicos en maíz. Otra solución sería considerar, una regresión no paramétrica de varias variables de clima en la construcción de un *kernel* para el efecto

ambiental e incorporarlo en la estimación de parámetros de un modelo mixto, tal como se presenta en el trabajo de Gianola y van Kaam (2008).

Becker y Leon (1988), mencionan que el concepto estático de la estabilidad puede ser útil en el caso de características cualitativas, resistencia a enfermedades o respuestas en condiciones de estrés, mientras que el concepto dinámico puede ser aplicado a otras características cuantitativas como es el caso de rendimiento. Asimismo, argumentan que los modelos de estabilidad fenotípica basados en el modelo [1], es decir, en sentido dinámico o de interés agronómico, surgen por la necesidad de los mejoradores de seleccionar líneas con altos rendimientos y algún grado de estabilidad, ya que es bien conocido que genotipos con rendimientos bajos tienen mayor estabilidad que genotipos con rendimientos altos. Sin embargo, no es totalmente clara la argumentación de ¿por qué en la variable rendimiento, las deficiencias mostradas en el modelo [1] son superadas, mientras que en las demás variables el modelo [2] si puede ser aplicable?, ¿por qué el modelo para evaluar la estabilidad fenotípica no es el mismo para todas las variables cuantitativas?

El punto central del problema tal vez no es si el concepto de estabilidad es biológico o agronómico, estático o dinámico, sino el de seleccionar genotipos teniendo en cuenta la magnitud del efecto del genotipo y la estabilidad del mismo a través de los ambientes, lo que es conocida como selección por rendimiento-estabilidad (Kang, 1988; Kang, 1991; Kang y Pham, 1991; Nassar *et al.*, 1994; Hühn, 1996; Kang y Magari, 1996).

## METODOLOGÍA PARA LA SELECCIÓN DE GENOTIPOS

Basados en el modelo [2] combinar la selección de genotipos con, por ejemplo, altos rendimientos o resistencia a enfermedades y buena estabilidad

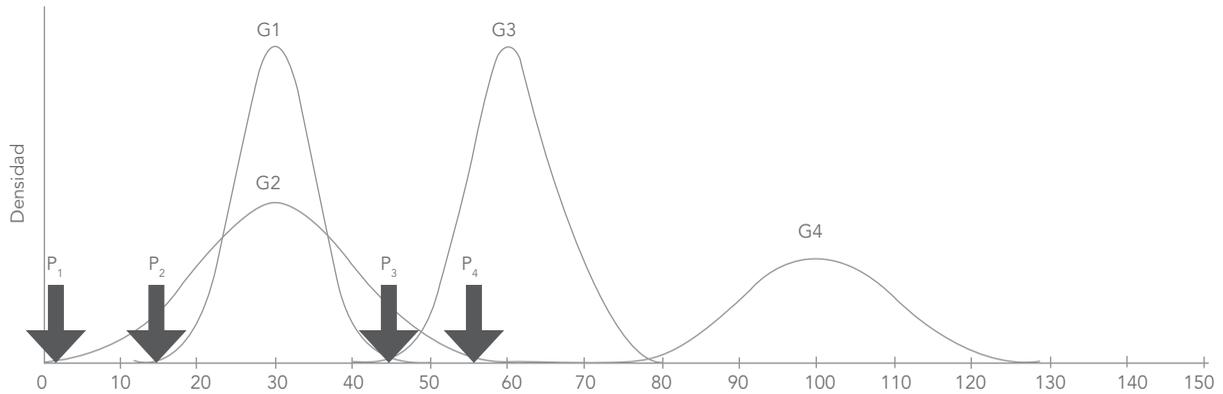
equivale a combinar en la metodología de selección los efectos  $u_i$  y  $\gamma_{ij}$ . Supóngase que el efecto de los genotipos es fijo  $u_i$ , así el desempeño medio es genotipo está dado por  $\phi_i = \mu + u_i$  y que los efectos  $\gamma_{ij}$  son aleatorios que siguen una distribución normal con media cero y varianzas  $\sigma_i^2$  entonces un genotipo deseable es aquel que tenga un alto valor para  $g_i$  y una varianza de estabilidad fenotípica baja  $\sigma_i^2$ .

En una distribución normal la diferencia entre la media y tres veces la raíz cuadrada de la varianza equivale al límite por debajo del cual es poco probable ( $p < 0,0013$ ) obtener valores de dicha variable aleatoria. Así, desde el punto de vista agronómico, si p.e. evaluamos rendimiento, la cantidad  $p_i = \phi_i - 3\sigma_i$  puede verse como el menor rendimiento posible. Los genotipos deseables serían aquellos con valores superiores de  $p_i$ . Para clarificar su uso, suponga que se evalúa el rendimiento en cuatro genotipos y tenemos valores de 30, 30, 60 y 100 para los efectos de  $\phi_i$ , y con valores de estabilidad fenotípica de 100, 25, 25 y 225. Así, los genotipos G1 y G2 son los que presentan bajos rendimientos y el genotipo G4 presenta el mayor rendimiento. Asimismo, los genotipos G2 y G3 presentan alta estabilidad y el genotipo G5 presenta la más baja estabilidad (Figura 2)

Los valores de  $p_i$  que juntan el potencial de rendimiento y la estabilidad del mismo son 0, 15, 45 y 55 para los genotipos G1, G2, G3 y G4, respectivamente. El genotipo a seleccionar sería el genotipo G4 porque posee mayor valor de  $p_i$ , siendo un genotipo altamente rendidor y con baja estabilidad fenotípica, sin embargo se esperaran rendimiento en campo superiores a los demás genotipos, ya que el valor de 55 indicaría el valor mínimo esperado para ese genotipo.

## INCORPORACIÓN DE INFORMACIÓN PREVIA Y ESTIMACIÓN DE PARÁMETROS

Hasta aquí, el artículo ha presentado la utilización del modelo de variancia ambiental como el más



**Figura 2.** Densidades normales para cuatro genotipos con diferente respuesta y estabilidad fenotípica. Las flechas en negro resaltan la posición del parámetro de rendimiento y estabilidad  $p_i$ .

apropiado para evaluar la estabilidad fenotípica en pruebas regionales. El objetivo del artículo es incorporar la información previa que se tiene en los programas de mejoramiento de algunos genotipos, especialmente de variedades testigo, donde la información recopilada puede ser de más de 15 años en algunos casos.

Aunque existen varios métodos de estimación estadísticas, la metodología de estimación basadas en el teorema de Bayes es la únicas en el momento donde la información previa o a priori puede ser considerada en el modelo probabilístico. Esta metodología se ha aplicado en la estimación de parámetros basados en el modelo [1] (Besag y Higdon, 1999; Foucteau y Dennis ,2001; Theobald et al., 2002, Edwards y Jannick, 2006; Cotes et al., 2006; Crossa et al., 2011), sin embargo, utilizar información previa para los parámetros de estabilidad fenotípica o para los efectos de interacción genotipo-ambiente es una dificultad en la aplicación de la metodología Bayesiana. Para el modelo [2] esto no ocurre debido a que tanto el parámetro de estabilidad como la magnitud del efecto del genotipo, son un atributo *per se* del genotipo sin estar contaminado por otros efecto, como ya se

mostró previamente.

Así, consideremos el siguiente modelo para análisis de pruebas regionales basado en el modelo [2]:

$$y_{ijk} = \mu + v_i + \gamma_j + \rho_{jk} + \epsilon_{ijk}$$

donde,  $y_{ijk}$  es la respuesta fenotípica del genotipo  $i$  ( $i=1,2,\dots,v$ ) del ambiente  $j$  ( $j=1,2,\dots,a$ ) en el bloque  $k$  ( $k=1,2,\dots,b$ ),  $\mu$ ,  $v_i$ ,  $\gamma_j$  son como definidos en el modelo [2],  $\rho_{jk}$  es el efecto del bloque  $k$  dentro del ambiente  $j$  y  $\epsilon_{ijk}$  es el error residual asociado al genotipo  $i$ , del ambiente  $j$ , en el bloque  $k$ .

En notación matricial considerando el efecto  $\phi_i = \mu - v_i$ , y el posible desbalance de los datos podemos expresar el modelo [3] como:

$$y = X\phi + Z_1\rho + \sum_{i=1}^v Z_{2(i)}\gamma_{(i)} + e$$

donde, en su orden,  $X$  y  $\phi = [\phi_1, \dots, \phi_v]^t$  son la matriz de incidencia  $n \times v$  y el vector  $v \times 1$  de efectos fijos de los genotipos siendo  $n$  el número total de observaciones,  $Z_1$  y  $\rho = [\rho_{11}, \dots, \rho_{av}]$  son a matriz de incidencia  $n \times \sum_j^a b_j$  y el vector  $\sum_j^a b_j \times 1$  de efectos de los bloques  $j$  dentro de los  $j$  ambientes y  $b_j$  indica el número de bloques en cada ambiente,

$\mathbf{Z}_{2(i)}$  y  $\boldsymbol{\gamma}_{(i)} = [\gamma_{i1}, \dots, \gamma_{ia}]$  son la matriz de incidencia  $n \times n_i$  y el vector  $n_i \times 1$  de efectos de los genotipos dentro de los ambientes para o  $i$ -ésimo genotipo, siendo  $n_i$  el número de observaciones para un genotipo. El vector  $\mathbf{e} = [\mathbf{e}'_1, \dots, \mathbf{e}'_a]'$ , donde  $\mathbf{e}_j$  es el vector  $n_j \times 1$  de efectos residuales para el  $j$ -ésimo ambiente, siendo  $n_j \times 1$  el número de observaciones en un ambiente.

Para realizar la estimación Bayesiana de los parámetros del modelo [3], se supone que la distribución condicional de  $\mathbf{y}$  dado  $\boldsymbol{\Phi}, \boldsymbol{\rho}, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_v, \sigma_\rho^2, \sigma_{\gamma(1)}^2, \dots, \sigma_{\gamma(v)}^2, \sigma_{e(1)}^2, \dots, \sigma_{e(a)}^2$  sigue una distribución normal multivariada, de la forma:

$$\mathbf{y} | \boldsymbol{\Phi}, \boldsymbol{\rho}, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_v, \sigma_\rho^2, \sigma_{\gamma(1)}^2, \dots, \sigma_{\gamma(v)}^2, \sigma_{e(1)}^2, \dots, \sigma_{e(a)}^2 \\ \sim N \left( \mathbf{X}\boldsymbol{\Phi} + \mathbf{Z}_1\boldsymbol{\rho} + \sum_{i=1}^v \mathbf{Z}_{2(i)}\boldsymbol{\gamma}_{(i)}, \mathbf{R} \right) \\ \mathbf{R} = \left\{ \sigma_{e(j)}^2 \mathbf{I}_{n_j} \right\}$$

donde  $\mathbf{I}_{n_j}$  es la matriz de identidad de orden  $n_j$ . La información *a priori* para cada uno de los parámetros es dada por:

$$\boldsymbol{\Phi} | \boldsymbol{\Phi}_0, \sigma_0^2 \sim N(\boldsymbol{\Phi}_0, \sigma_0^2) \\ \boldsymbol{\rho} | \sigma_\rho^2 \sim N(\mathbf{0}, \mathbf{I}_b \sigma_\rho^2) \\ \boldsymbol{\gamma}_i | \sigma_{\gamma(i)}^2 \sim N(\mathbf{0}, \mathbf{I}_{n_i} \sigma_{\gamma(i)}^2) \\ \sigma_\rho^2 | \nu_\rho, s_\rho^2 \sim \text{Inv-Scaled-}\chi^2(\nu_\rho, s_\rho^2) \\ \sigma_{\gamma(i)}^2 | \nu_{\gamma(i)}, s_{\gamma(i)}^2 \sim \text{Inv-Scaled-}\chi^2(\nu_{\gamma(i)}, s_{\gamma(i)}^2)$$

$$\sigma_{e(j)}^2 | \nu_{e(j)}, s_{e(j)}^2 \sim \text{Inv-Scaled-}\chi^2(\nu_{e(j)}, s_{e(j)}^2)$$

donde, en su orden,  $\boldsymbol{\Phi}_0$  y  $\sigma_0^2$  son el vector de medias de la matriz diagonal de varianzas *a priori* del vector  $\boldsymbol{\Phi}$  de efectos fijos (medias de los genotipos),  $\nu_\rho$  y  $s_\rho^2$  son los grados de credibilidad y el factor de escala para el componente de variancia  $\sigma_\rho^2$ ,  $\nu_{\gamma(i)}$  y  $s_{\gamma(i)}^2$  son los grados de credibilidad y el factor de escala para el componente de variancia  $\sigma_{\gamma(i)}^2$  y, finalmente,  $\nu_{e(j)}$  y  $s_{e(j)}^2$  son los grados de credibilidad y el factor de escala para a variancia de  $\sigma_{e(j)}^2$ . La notación *Inv-Scaled-}\chi^2* (Gelman et al., 2004) indica a distribución ji-cuadrado invertida con un parámetro de escala (*scaled*, en inglés). Los valores  $\boldsymbol{\Phi}_0, \sigma_0^2, \nu_\rho, s_\rho^2, \nu_{\gamma(i)}, s_{\gamma(i)}^2, \nu_{e(j)}$  y  $s_{e(j)}^2$  son denominados de hiper-parámetros. En este modelo la información previa sobre la media de los genotipos puede ser plasmada a través de la distribución *a priori* de  $\boldsymbol{\Phi}$ , mientras que la de la variancia de estabilidad puede expresarse a través de la distribución *a priori* de  $\sigma_{\gamma(i)}^2$ . Es de notar que este modelo también considera que la variancia del error residual pueda ser diferente de un ambiente para otro, es decir, heteroscedasticidad de variancia para el término del error, lo cual permite mayor flexibilidad al modelo. Sin embargo, considerar el modelo con varianzas homogéneas es factible y simplifica algunas expresiones de las anteriormente presentadas.

$$p(\boldsymbol{\Phi}, \boldsymbol{\rho}, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_v, \sigma_\rho^2, \sigma_{\gamma(1)}^2, \dots, \sigma_{\gamma(v)}^2, \sigma_{e(1)}^2, \dots, \sigma_{e(a)}^2 | \mathbf{y}, \mathbf{h}) \\ \propto \exp \left[ -\frac{1}{2\sigma_0^2} (\boldsymbol{\Phi}_0 - \boldsymbol{\Phi})' (\boldsymbol{\Phi}_0 - \boldsymbol{\Phi}) \right] \times \left( \prod_{i=1}^a (\sigma_{e(i)}^2)^{-\frac{\nu_j + \nu_{e(i)} - 1}{2}} \exp \left[ -\frac{1}{2\sigma_{e(i)}^2} \nu_{e(i)} s_{e(i)}^2 \right] \right) \quad [5]$$

$$\times \exp \left[ -\frac{1}{2} \left( \mathbf{y} - \mathbf{X}\boldsymbol{\phi} + \mathbf{Z}_1\boldsymbol{\rho} + \sum_{j=1}^v \mathbf{Z}_{2(j)}\boldsymbol{\gamma}_{(j)} \right)' \mathbf{R}^{-1} \left( \mathbf{y} - \mathbf{X}\boldsymbol{\phi} + \mathbf{Z}_1\boldsymbol{\rho} + \sum_{j=1}^v \mathbf{Z}_{2(j)}\boldsymbol{\gamma}_{(j)} \right) \right]$$

$$\times (\sigma_\rho^2)^{\frac{\left(\sum_{j=1}^a b_j\right) + v_\rho}{2} - 1} \exp \left[ -\frac{1}{2\sigma_\rho^2} \boldsymbol{\rho}' \mathbf{I}_b \boldsymbol{\rho} + v_\rho S_\rho^2 \right] \times \left( \prod_{j=1}^v (\sigma_{\gamma(j)}^2)^{\frac{n_j + v_{\gamma(j)}}{2} - 1} \exp \left[ -\frac{1}{2\sigma_{\gamma(j)}^2} \boldsymbol{\gamma}' \mathbf{I}_{n_j} \boldsymbol{\gamma} + v_{\gamma(j)} S_{\gamma(j)}^2 \right] \right)$$

Multiplicando las distribuciones a priori por la distribución condicional de los datos o función de verosimilitud, se obtiene la distribución a posteriori conjunta de la forma:

donde  $\mathbf{h}$  representa el vector de los hiper-parámetros.

El problema con la estimación bayesiana es que esta distribución a posteriori de difícil manipulación matemática, y con el conocimiento actual en el área de cálculo y matemáticas es imposible la obtención de las distribuciones marginales de cada uno de los parámetros. Así es necesario acudir a un método numérico denominado algoritmo de GIBBS.

**Algoritmo de GIBBS**

Para la implementación del algoritmo es necesario reconocer las distribuciones condicionales de cada parámetro con base en la distribución a posteriori conjunta [5]. Por conveniencia, y debido a un mejor comportamiento del método numérico, es común utilizar la siguiente reparametrización:

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\phi} \\ \boldsymbol{\rho} \\ \boldsymbol{\gamma}_1 \\ \vdots \\ \boldsymbol{\gamma}_v \end{bmatrix}, \boldsymbol{\theta}_0 = \begin{bmatrix} \boldsymbol{\phi}_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \mathbf{D} = \begin{bmatrix} \sigma_{\phi_0}^2 & & & & \\ & \mathbf{I}_{\sigma_\rho^2} & & & \\ & & \mathbf{I}_{n_1} \sigma_{\gamma(1)}^2 & & \\ & & & \ddots & \\ & & & & \mathbf{I}_{n_v} \sigma_{\gamma(v)}^2 \end{bmatrix}$$

e,  $\mathbf{W} = [\mathbf{X} \quad \mathbf{Z}_1 \quad \dots \quad \mathbf{Z}_{2(v)}]$

$$p(\boldsymbol{\theta} | \sigma_\rho^2, \sigma_{\gamma(1)}^2, \dots, \sigma_{\gamma(v)}^2, \sigma_{e(1)}^2, \dots, \sigma_{e(a)}^2, \mathbf{y}, \mathbf{h})$$

$$\propto \exp \left[ -\frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' (\mathbf{W}' \mathbf{R}^{-1} \mathbf{W} + \mathbf{D}^{-1}) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right]$$

Después de aplicar las reglas de álgebra de matrices, la distribución condicional a posteriori para  $\boldsymbol{\theta}$  es (Ver Apéndice A),

donde,

$$\hat{\boldsymbol{\theta}} = (\mathbf{W}' \mathbf{R}^{-1} \mathbf{W} + \mathbf{D}^{-1})^{-1} (\mathbf{W}' \mathbf{R}^{-1} \mathbf{y} + \mathbf{D}^{-1} \boldsymbol{\theta}_0)$$

Esto es,  $\boldsymbol{\theta}$  se distribuye condicionalmente como una normal multivariada, según la siguiente expresión.

$$\boldsymbol{\theta} | \sigma_{u_1}^2, \sigma_{u_2(a)}^2, \dots, \sigma_{u_2(a)}^2, \sigma_{e_1}^2, \dots, \sigma_{e_a}^2, \mathbf{y}, \mathbf{h} \quad [6]$$

$$\sim N(\hat{\boldsymbol{\theta}}, (\mathbf{W}' \mathbf{R}^{-1} \mathbf{W} + \mathbf{D}^{-1})^{-1})$$

Las distribuciones condicionales a posteriori para  $\sigma_{e(j)}^2$  se identifican como una ji-cuadrado invertida escalonada, esto es:

$$\sigma_{e(j)}^2 | \boldsymbol{\theta}, \sigma_\rho^2, \sigma_{\gamma(1)}^2, \dots, \sigma_{\gamma(v)}^2, \mathbf{y}, \mathbf{h} \quad [7]$$

$$\sim \text{Inv-Scaled-}\chi^2(\hat{v}_{e(j)}, \hat{S}_{e(j)}^2)$$

donde

$$\hat{v}_{e(j)} = v_{e(j)} + n_j \quad \text{y} \quad \hat{S}_{e(j)}^2 = \frac{\mathbf{e}'_j \mathbf{e}_j + v_{e(j)} S_{e(j)}^2}{v_{e(j)} + n_j}$$

De igual forma, las distribuciones condicionales a posteriori para  $\sigma_\rho^2$  son también identificadas como una ji-cuadrado invertida escalonada, es decir:

$$\sigma_{\rho}^2 | \boldsymbol{\theta}, \sigma_{\alpha}^2, \dots, \sigma_{\epsilon}^2, \mathbf{y}, \mathbf{h} \sim \text{Inv-Scaled} - \chi^2(\hat{v}_{\rho}, \hat{s}_{\rho}^2) \quad [8]$$

donde

$$\hat{v}_{\rho} = v_{\rho} + \sum_{j=1}^a b_j \quad \text{y} \quad \hat{s}_{\rho}^2 = \frac{\mathbf{p}'\mathbf{p} + v_{\rho}s_{\rho}^2}{v_{\rho} + \sum_{j=1}^a b_j}$$

Asimismo, las distribuciones condicionales a posteriori para  $\sigma_{\gamma(l)}^2$  son también identificadas como una ji-cuadrado invertida escalonada, es decir:

$$\sigma_{\gamma(l)}^2 | \boldsymbol{\theta}, \sigma_{\alpha}^2, \dots, \sigma_{\epsilon}^2, \mathbf{y}, \mathbf{h} \sim \text{Inv-Scaled} - \chi^2(\hat{v}_{\gamma(l)}, \hat{s}_{\gamma(l)}^2) \quad [9]$$

donde

$$\hat{v}_{\gamma(l)} = v_{\gamma(l)} + n_l \quad \text{y} \quad \hat{s}_{\gamma(l)}^2 = \frac{\boldsymbol{\gamma}'\boldsymbol{\gamma}_l + v_{\gamma(l)}s_{\gamma(l)}^2}{v_{\gamma(l)} + n_l}$$

Así, para generar un número  $r$  de vectores  $[\boldsymbol{\phi}, \boldsymbol{\rho}, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_v, \sigma_{\rho}^2, \sigma_{\gamma(1)}^2, \dots, \sigma_{\gamma(v)}^2, \sigma_{e(1)}^2, \dots, \sigma_{e(a)}^2]^t$  de la distribución a posteriori conjunta [5] es necesario seguir los siguientes pasos:

- Colocar valores iniciales para cada uno de los parámetros  $\sigma_{\rho}^2, \sigma_{\gamma(1)}^2, \dots, \sigma_{\gamma(v)}^2, \sigma_{e(1)}^2, \dots, \sigma_{e(a)}^2$
- Generar un vector de números aleatorios para acorde con la distribución condicional a posteriori [6] y actualizar;
- Generar un vector de números aleatorios para cada uno de los  $\sigma_{\epsilon_j}^2$  acorde con la distribución condicional a posteriori [7] y actualizar
- Generar un vector de números aleatorios para cada uno de los  $\sigma_{\rho}^2, \sigma_{\gamma(1)}^2, \dots, \sigma_{\gamma(v)}^2, \sigma_{e(1)}^2, \dots, \sigma_{e(a)}^2$  acorde con la distribución condicional a posteriori [8] y [9] y actualizar;
- Con los valores actualizados, se repiten los pasos 2,3 y 4 otras  $r-1$  veces.

El anterior algoritmo es parte de los algoritmos denominado MCMC, por las siglas en inglés de *Markov Chain Monte Carlo*. Estos métodos

numéricos basados en cadenas de Markov, presentan un grado de autocorrelación entre la muestra  $r$  y la muestra  $r+1$ , y debido a que los valores iniciales de los parámetros son aleatorios requieren de un determinado número de iteraciones o muestras de vectores antes de alcanzar la estabilidad numérica deseada. En el primer caso, para evitar que la autocorrelación de las muestras afecte la estimación de los parámetros es común seleccionar una a cada  $r_t$  muestras, con la finalidad de garantizar que la muestra  $r$  y la muestra  $r_t$  no estén correlacionadas, lo que se conoce como el *thin* por sus siglas en inglés o el espaciamiento entre muestra. Para el segundo caso, es común no considerar las primeras  $r_b$  muestras, lo que es conocido como período de *burn-in* por la expresión en inglés o período de calentamiento. Tanto el espaciamiento como el período de calentamiento dependen de los datos, por lo que para cada conjunto de datos hay que tomar decisiones sobre estos parámetros del método numérico.

Debido al período de calentamiento y al espaciamiento entre las muestras para evitar autocorrelación, de los  $r$  vectores generados quedan solamente  $s < r$  vectores, siendo  $s < r$ . Finalmente, muestras de las distribuciones marginales a posteriori,  $p(\boldsymbol{\theta} | \mathbf{y})$ ,  $p(\sigma_{\rho}^2 | \mathbf{y})$ ,  $p(\sigma_{\gamma(l)}^2 | \mathbf{y})$  y  $p(\sigma_{e(j)}^2 | \mathbf{y})$  son obtenidos de los  $s$  vectores generados, utilizando los componentes  $\boldsymbol{\theta}$ ,  $\sigma_{\rho}^2$ ,  $\sigma_{\gamma(l)}^2$  y  $\sigma_{e(j)}^2$  de cada uno de esos vectores, respectivamente. Para la obtención de muestras de la distribución marginal del parámetro de rendimiento estabilidad  $p_i$  este se realiza con base en los vectores generados de las distribuciones marginales de  $p(\boldsymbol{\phi} | \mathbf{y})$  y  $p(\sigma_{\gamma(l)}^2 | \mathbf{y})$ .

La estimación puntual se hace a través de los estimadores de bayes. Una amplia discusión sobre estos estimadores es la realizada por Box y Tiao (1973), quienes afirman que la estimación puntual en el contexto de inferencia bayesiana debería ser utilizada. Sin embargo, la media y la mediana de las distribuciones marginales a posteriori de cada uno de los

parámetros, son frecuentemente utilizadas como estimativas de bayes. La media es considerada aquella estimativa de bayes, que minimiza el riesgo de bayes, utilizando la función de pérdida cuadrática, en el contexto de teoría de la decisión. La mediana, minimiza el riesgo de bayes utilizando la función de pérdida absoluta (Mood et al., 1974).

La estimación por intervalo puede realizar a través de los intervalos de alta densidad *a posteriori* (HPD, por las siglas del inglés *Highest Posterior Density*) o por los intervalos de credibilidad. Se prefiere los intervalos HPD debido a que éstos son los intervalos de menor longitud entre todos los posibles intervalos de probabilidad  $(1-\alpha)\%$ , siendo  $\alpha$  el nivel de precisión deseado (Box y Tiao, 1973).

En cuanto a la parte computacional, en el ambiente de análisis estadístico R (R Core Team, 2012), el cual es facilitado bajo la licencia GNU y de fácil acceso por la comunidad académica, se encuentra el paquete MCMCglmm (Hadfield, 2010) en el cual se pueden obtener los  $s$  vector de las distribuciones marginales *a posteriori*, la estimación puntual y por intervalo del modelo [4] presentado en este trabajo. El paquete CODA (Plummer et al. 2006) puede ser útil en el establecimiento del período de calentamiento y el espaciamiento entre las muestras.

### Optimización del algoritmo de GIBBS

La principal dificultad en la implementación del algoritmo de GIBBS, para obtener muestras de la distribución *a posteriori* conjunta [5], está en el segundo paso en la generación de vectores aleatorios de la distribución condicional de  $\theta$ [6]. En este paso es necesaria la inversión de la matriz  $(\mathbf{W}'\mathbf{R}^{-1}\mathbf{W}+\mathbf{D}^{-1})^{-1}$ , la cual es necesaria tanto para la media  $\hat{\theta}$  como para la varianza. Así, el proceso de inversión genera errores numéricos computacionales los cuales afectan visiblemente al menos estos dos componentes. Con el fin de disminuir

la propagación de estos errores en la obtención de vectores aleatorios de la distribución  $\theta$ [6], esta puede ser escrita como:

$$\theta | \sigma_{u_1}^2, \sigma_{u_{2(1)}}^2, \dots, \sigma_{u_{2(i)}}^2, \sigma_{\eta_1}^2, \dots, \sigma_{\eta_s}^2, \mathbf{y}, \mathbf{h} \\ \sim N(\mathbf{C}^{-1}(\mathbf{W}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{D}^{-1}\theta_0), \mathbf{C}^{-1})$$

donde  $\mathbf{C}^{-1} = (\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{D}^{-1})^{-1}$ . Luego realizando una transformación de variable podemos obtener:

$$\mathbf{C}\theta | \sigma_{u_1}^2, \sigma_{u_{2(1)}}^2, \dots, \sigma_{u_{2(i)}}^2, \sigma_{\eta_1}^2, \dots, \sigma_{\eta_s}^2, \mathbf{y}, \mathbf{h} \\ \sim N(\mathbf{C}\mathbf{C}^{-1}(\mathbf{W}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{D}^{-1}\theta_0), \mathbf{C}\mathbf{C}^{-1}\mathbf{C}')$$

Porque  $\mathbf{C} = \mathbf{C}'$ , y simplificando la expresión se obtiene.

$$\theta^* | \sigma_{u_1}^2, \sigma_{u_{2(1)}}^2, \dots, \sigma_{u_{2(i)}}^2, \sigma_{\eta_1}^2, \dots, \sigma_{\eta_s}^2, \mathbf{y}, \mathbf{h} \\ \sim N((\mathbf{W}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{D}^{-1}\theta_0), \mathbf{C}) \quad [10]$$

donde  $\theta^* = \mathbf{C}\theta$ . Así el segundo paso del algoritmo de GIBBS puede ser reemplazado por:

- a. Generar un vector de números aleatorios para  $\theta^*$  acorde con la distribución condicional *a posteriori* [10]
- b. Obtener  $\theta = \mathbf{C}^{-1}\theta^*$

La realización de estos dos pasos son más eficientes computacionalmente que la obtención de vectores aleatorios de la distribución condicional de  $\theta$ , mejorando la rapidez del algoritmo, pero lo importante es que solamente se realiza una sola vez el proceso de inversión de la matriz  $(\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{D}^{-1})^{-1}$ , con lo cual se disminuyen los errores numéricos computacionales.

Por otro lado, encontrar una forma eficiente de invertir la matriz  $(\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{D}^{-1})^{-1}$  es crucial en la programación del algoritmo, ya que de ello depende en gran medida la rapidez del mismo. La matriz  $(\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{D}^{-1})^{-1}$  puede escribirse como:

$$\begin{bmatrix} \mathbf{X}\mathbf{R}^{-1}\mathbf{X} + \mathbf{I}/\sigma_{\alpha(k)}^2 & \mathbf{X}\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{X}\mathbf{Z}_{2(1)} & \dots & \mathbf{X}\mathbf{Z}_{2(v)} \\ \mathbf{Z}'_1\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'_1\mathbf{R}^{-1}\mathbf{Z}_1 + \mathbf{I}\frac{1}{\sigma_{\rho}^2} & \mathbf{Z}'_1\mathbf{R}^{-1}\mathbf{Z}_{2(1)} & \dots & \mathbf{Z}'_1\mathbf{R}^{-1}\mathbf{Z}_{2(v)} \\ \mathbf{Z}'_{2(1)}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'_{2(1)}\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{Z}'_{2(1)}\mathbf{R}^{-1}\mathbf{Z}_{2(1)} + \mathbf{I}\frac{1}{\sigma_{\gamma(1)}^2} & \dots & \mathbf{Z}'_{2(1)}\mathbf{R}^{-1}\mathbf{Z}_{2(v)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{Z}'_{2(v)}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'_{2(v)}\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{Z}'_{2(v)}\mathbf{R}^{-1}\mathbf{Z}_{2(1)} & \dots & \mathbf{Z}'_{2(v)}\mathbf{R}^{-1}\mathbf{Z}_{2(v)} + \mathbf{I}\frac{1}{\sigma_{\gamma(v)}^2} \end{bmatrix}^{-1}$$

Sin embargo, si el vector de observaciones se ordena por genotipo, el efecto sobre esta matriz inversa es un efecto de permutar líneas y columnas, por lo que es equivalente a:

puede observarse la matriz  $(\mathbf{W}\mathbf{R}^{-1}\mathbf{W} + \mathbf{D}^{-1})^{-1}$ , puede transformarse en una matriz bloque-diagonal de fácil inversión.

**CONCLUSIONES**

$$\begin{bmatrix} \mathbf{H}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}_j & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{H}_v \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{H}_1^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2^{-1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{H}_j^{-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{H}_v^{-1} \end{bmatrix}$$

donde,

$$\mathbf{H}_j = \begin{bmatrix} \mathbf{I}_j\mathbf{R}_j^{-1}\mathbf{1}_j\beta_j + \frac{1}{\sigma_{\alpha(j)}^2} & \mathbf{I}_j\mathbf{R}_j^{-1}\mathbf{Z}_{1(j)} & \mathbf{I}_j\mathbf{R}_j^{-1}\mathbf{Z}_{2(j)} \\ \mathbf{Z}'_{1(j)}\mathbf{R}_j^{-1}\mathbf{1}_j & \mathbf{Z}'_{1(j)}\mathbf{R}_j^{-1}\mathbf{Z}_{1(j)} + \mathbf{I}\frac{1}{\sigma_{\rho}^2} & \mathbf{Z}'_{1(j)}\mathbf{R}_j^{-1}\mathbf{Z}_{2(j)} \\ \mathbf{Z}'_{2(j)}\mathbf{R}_j^{-1}\mathbf{1}_j & \mathbf{Z}'_{2(j)}\mathbf{R}_j^{-1}\mathbf{Z}_{1(j)} & \mathbf{Z}'_{2(j)}\mathbf{R}_j^{-1}\mathbf{Z}_{2(j)} + \mathbf{I}\frac{1}{\sigma_{\gamma(j)}^2} \end{bmatrix}^{-1}$$

donde  $\mathbf{1}_k$  es un vector de unos de tamaño  $n_k$ ,  $\mathbf{R}_j^{-1}$  es la matriz de residuales  $n_j \times n_j$  debida al genotipo  $j$ , y  $\mathbf{Z}_{1(j)}$  y es la matriz del efecto de bloque  $n_j \times b_j$  dentro de ambiente debida al genotipo  $j$ . Como

Analizar la estabilidad fenotípica basado en un modelo de dos vías con interacción no es adecuado, ya que los efectos de la interacción genotipo por ambiente debida a un determinado genotipo, no representan una valoración per se de la estabilidad del genotipo y esta depende de directamente de los efectos ambientales e indirectamente del desempeño de los demás genotipos. El modelo de varianza ambiental, que es un modelo jerárquico que contempla el desempeño de los genotipos dentro de los ambientes, es un modelo que permite medir la estabilidad fenotípica en cualquier tipo de variable cuantitativa. Se desarrolló este modelo mediante la metodología bayesiana la cual permite la incorporación de información previa que puede estar disponible en algunos genotipos. Se propuso el parámetro  $P_j$  el cual ayuda al mejorador a seleccionar genotipos deseables en el caso de variables con distribución normal.

**BIBLIOGRAFÍA**

1. Becker H.C. 1981. Correlations among some statistical measures of phenotypic stability. *Euphytica* 30: 835-840.
2. Becker H.C., Leon J. 1988. Stability analysis in plant breeding. *Plant Breeding* 101, 1-23.
3. Besag J., Higdon D. 1999. Bayesian analysis of agricultural field experiments. *Journal Royal Statistical Society. Series B.* 61: 641-746
4. Box G.E.P., Tiao G. C. 1973. *Bayesian Inference in Statistical Analysis.* Wiley Classic Library, New York. 608 p.
5. Burgueño J., Crossa J., Cornelius P.L., Yang R.-C. 2008. Using factor analytic models for joining environments and genotypes without crossover genotype  $\times$  environment interaction. *Crop Science* 48:1291-1305.
6. Cooper M., DeLacy I.H. 1994. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theoretical and Applied Genetics* 88: 561-572.
7. Cornelius P.L., Crossa J. y Seyedsadr M.S. 1996. Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. p. 199-234. En: Kang M.S. y Gauch H.G. (Eds.) *Genotype-by-Environment Interaction.* CRC press, New York, 416 p.
8. Cotes J.M, Zuluaga C. M., González E. P., Morales J.G., Marín, M. A., Ñustez C.E., Gutierrez P.A. 2012. Informe técnico de investigación proyecto "Evaluación fenotípica y genotípica de la colección de *Solanum phureja* por su resistencia a *Spongospora subterranea* f. sp *subterranea*". Universidad Nacional de Colombia y Politécnico Colombiano Jaime Isaza Cadavid. Medellín, 98p.
9. Cotes J.M., Crossa J., Sanches A., Cornelius P.L. 2006. A Bayesian Approach for Assessing the Stability of Genotypes. *Crop Science.* 46:2654-2665.
10. Crossa J., Yang R.-C., Cornelius P.L. 2004. Studying crossover genotype  $\times$  environment interaction using linear-bilinear models and mixed models. *Journal of Agricultural, Biology and Environmental Statistics.* 9: 362-380.
11. DeLacy I. H., Eisemann R.L., Cooper M. 1990. The importance of genotype by environment interaction in regional variety trials. p. 287-300. Kang K.S. (Ed.) En: *Genotype-by-Environment Interaction and Plant Breeding.* Louisiana State University, Baton Rouge, Louisiana.
12. DeLacy I.H., Cooper M., Basford K.E. 1996. Relationships among analytical methods used to study genotype-by-environment interactions and evaluation of their impact on response to selection. p. 51-84. En: Kang M.S. y Gauch H.G. (Eds.) *Genotype-by-Environment Interaction.* CRC press, New York, 416 p.
13. Eberhart S.A., Russell W.A. 1966. Stability parameters for comparing varieties. *Crop Science* 6:36-40.
14. Edwards, J.W., Jannink J.-L. 2006. Bayesian modeling of heterogeneous error and genotype  $\times$  environment interaction variances. *Crop Science* 46:820-833.

15. Falconer D. S., Mackay T. 1996. Introduction to quantitative genetics. 4a edición, Longman Group Ltd, Edinuburgh, 480 p.
16. Finlay K.W., Wilkinson G.N. 1963. The analysis of adaptation in a plant breeding program. *Australian Journal of Agriculture Research* 14: 742-754.
17. Foucteau V. Denis J.B. 2001. Statistical analysis of successive experiments in plant breeding: A Bayesian approach. p.49-56. En: Gallais, A., Dillman C., Goldringer I. (Eds.). *EUCARPIA Quantitative genetics and breeding methods: the way ahead*. INRA Editions, Paris, France. 333 p.
18. Gauch H. G., Zobel R. W. AMMI analysis of yield trials. p. 85-122. En: Kang M.S. y Gauch H.G. (Eds.) *Genotype-by-Environmet Interaction*. CRC press, New York, 416 p.
19. Gauch H.G. 2006. Statistical analysis of yield trials by AMMI and GGE. *Crop Science* 46:1488-1500
20. Gauch H.G., Zobel R.W. 1988. AMMI analysis of yield trials. p. 85-122. En: Kang M.S. y Gauch H.G. (Eds.) *Genotype-by-Environmet Interaction*. CRC press, New York, 416 p.
21. Gauch H.G., Piepho H.P., Annicchiarico P. 2008. Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop Science*. 48:866-889.
22. Gelman A., Carlin J. B., Stern H. S., Rubin D.B. 2004. *Bayesian data analysis*. 2a edición, Chapman & Hall, London, 668p.
23. Gianola D., van Kaam . J.B. C. H. M. 2008 Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* 178: 2289–2303.
24. Hadfield J.D. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* 33: 1-22.
25. Hühn M. 1996. Nonparametric analysis of genotype X environment interactions by ranks. p. 235-271. En: Kang M.S. y Gauch H.G. (Eds.) *Genotype-by-Environmet Interaction*. CRC press, New York, 416 p.
26. Crossa J., Perez-Elizalde S., Jarquin D., Cotes J.M., Viele K., Liu G., Cornelius P.L. 2011. Bayesian Estimation of the Additive Main Effects and Multiplicative Interaction Model. *Crop Science* 51: 1458-1469
27. Kang M.S., Magari, R. 1996. New development in selecting for phenotypic stability in crop breeding. p. 1-14. En: Kang M.S. y Gauch H.G. (Eds.) *Genotype-by-Environmet Interaction*. CRC press, New York, 416 p.
28. Kang M.S., Pham H.N. 1991. Simultaneous selection for high yielding and stable crop genotypes. *Agronomy Journal* 83: 161-165.
29. Kelly A.M., Smith A.B., Eccleston J.A., Cullis B.R. 2007. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. *Crop Science* 47:1063-1070.
30. Lin C.S., Binns M.R., Levkovitch L.P. 1986. Stability analysis: where do we stand?. *Crop Science* 26:894-900
31. Plummer M., Best N., Cowles K., Vines K. 2006. *R News* 6: 7 – 11.
32. Mood A. M., Graybill F.A., Boes D.C. 1974. *Introduction to the theory of statistics*. 3a edición, McGraw Hill, Tokyo, p. 339-350.

33. Nachit M.M., Nachit H., Ketata H., Gauch H.G., Zobel R.W. 1992. Use of AMMI and linear regression model to analyze genotype-environment interaction in durum wheat. *Theoretical and Applied Genetics* 83: 597-601.
34. Nassar R., León J., Hühn M. 1994. Test of significance for combined measures of plant stability and performance. *Biometrical Journal* 36: 109-123.
35. Ñustez C.E. 2011. Estudios fenotípicos y genéticos asociados a la calidad de fritura en *Solanum phureja* Juz et Buk. Tesis de Doctorado en Ciencias Agropecuarias. Facultad de Agronomía. Universidad Nacional de Colombia – Sede Bogotá. Bogotá D.C. 224 p.
36. Perkins J.M., Jinks J.L. 1968. Environmental and genotype-environmental components of variability. III Multiple lines and crosses. *Heredity* 23: 339-356.
37. Piepho H.P. 1998. Empirical best linear unbiased prediction in cultivar trials using factor analytic variance-covariance structures. *Theoretical and Applied Genetics* 97: 195-201.
38. Piepho H.P. 1999. Stability analysis using the SAS system. *Agronomy Journal* 91: 154-160.
39. Plaisted R.L., Peterson L.C. 1959. A technique for evaluating the ability of selections to yield consistently in different locations or seasons. *American Potato Journal* 36: 381-385.
40. Poehlman J.M., Allen D. 2003. *Mejoramiento genético de las cosechas*. 2a edición, Editorial Limusa, México, 506p.
41. R Development Core Team. 2012. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org/>
42. Shukla G.K. 1972. Some statistical aspects of partitioning genotype-environment components of variability. *Heredity* 29: 237-245
43. Smith A.B., Cullis B.R., Thompson R. 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *Journal of Agricultural Science*. 143:1-14.
44. Tai G.C.C. 1971. Genotypic stability and its application to potato regional trials. *Crop Science* 11: 184-190.
45. Theobald C.M., Talbot M., Nabugoomu F. 2002. A Bayesian approach to regional and local-area prediction from crop variety trials. *Journal of Agricultural, Biological and Environmental Statistics* 7: 403-419.
46. Wescott, B. 1987. A method of assessing the yield stability of crop genotypes. *Journal of Agricultural Science* 108: 267-274.
47. Yan W., Kang M.S., Ma B., Woods S., Cornelius P.L. 2007. GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Science* 47: 643-655.
48. Yang R.C., Crossa J., Cornelius P.L., Burgueño J. 2009. Biplot analysis of genotype  $\times$  environment interaction: Proceed with caution. *Crop Science* 49: 1-13
49. Yates F., Cochran W.G. 1938. The analysis of group of experiment. *Journal of Agriculture Science* 28: 556-580.
50. Yau S.K. 1995. Regression and AMMI analyses of genotype  $\times$  environment interactions: An empirical comparison. *Agronomy Journal* 87: 121-126.

## APÉNDICE

Obtención de la distribución condicional de  $\theta$  dado,  $\sigma_\rho^2, \sigma_{\gamma^{(1)}}^2, \dots, \sigma_{\gamma^{(v)}}^2, \sigma_{e^{(1)}}^2, \dots, \sigma_{e^{(a)}}^2$

La distribución *a posteriori* conjunta [5] puede ser escrita como:

$$\begin{aligned} & p(\theta, \sigma_\rho^2, \sigma_{\gamma^{(1)}}^2, \dots, \sigma_{\gamma^{(v)}}^2, \sigma_{e^{(1)}}^2, \dots, \sigma_{e^{(a)}}^2 | \mathbf{y}, \mathbf{h}) \\ & \propto \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{W}\theta)' \mathbf{R}^{-1}(\mathbf{y} - \mathbf{W}\theta)\right] \times \exp\left[-\frac{1}{2}(\theta_0 - \theta)'(\theta_0 - \theta)\right] \\ & \times \left(\sigma_\rho^2\right)^{\frac{\left(\sum_{j=1}^v b_j\right) + v_\rho}{2} - 1} \exp\left[-\frac{1}{2\sigma_\rho^2} \rho' \mathbf{I}_b \rho + v_\rho s_\rho^2\right] \times \left(\prod_{l=1}^v \left(\sigma_{\gamma^{(l)}}^2\right)^{\frac{n_l + v_{\gamma^{(l)}}}{2} - 1} \exp\left[-\frac{1}{2\sigma_{\gamma^{(l)}}^2} \gamma' \mathbf{I}_{n_l} \gamma + v_{\gamma^{(l)}} s_{\gamma^{(l)}}^2\right]\right) \end{aligned}$$

Por lo tanto la distribución condicional de  $\theta$  dado,  $\sigma_\rho^2, \sigma_{\gamma^{(1)}}^2, \dots, \sigma_{\gamma^{(v)}}^2, \sigma_{e^{(1)}}^2, \dots, \sigma_{e^{(a)}}^2$  puede ser escrita como:

$$\begin{aligned} & p(\theta, \sigma_\rho^2, \sigma_{\gamma^{(1)}}^2, \dots, \sigma_{\gamma^{(v)}}^2, \sigma_{e^{(1)}}^2, \dots, \sigma_{e^{(a)}}^2 | \mathbf{y}, \mathbf{h}) \\ & \propto \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{W}\theta)' \mathbf{R}^{-1}(\mathbf{y} - \mathbf{W}\theta)\right] \times \exp\left[-\frac{1}{2}(\theta_0 - \theta)'(\theta_0 - \theta)\right] \end{aligned}$$

Utilizando solamente las formas cuadráticas del exponente, se obtiene:

$$\begin{aligned} & (\mathbf{y} - \mathbf{W}\theta)' \mathbf{R}^{-1}(\mathbf{y} - \mathbf{W}\theta) + (\theta_0 - \theta)' \mathbf{D}^{-1}(\theta_0 - \theta) \\ & = (\theta - \hat{\theta})' \mathbf{W}' \mathbf{R}^{-1} \mathbf{W} (\theta - \hat{\theta}) + (\mathbf{y} - \mathbf{W}\hat{\theta})' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{W}\hat{\theta}) + (\theta_0 - \theta)' \mathbf{D}^{-1}(\theta_0 - \theta) \end{aligned}$$

donde  $\hat{\theta} = (\mathbf{W}' \mathbf{R}^{-1} \mathbf{W})^{-1} \mathbf{W}' \mathbf{R}^{-1} \mathbf{y}$ . Porque  $(\mathbf{y} - \mathbf{W}\hat{\theta})' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{W}\hat{\theta})$  no depende de  $\theta$ , puede ser absorbido por el signo de proporcionalidad. Así, quedándose solamente con la parte variable, se obtiene que:

$$\begin{aligned} & (\theta - \hat{\theta})' \mathbf{W}' \mathbf{R}^{-1} \mathbf{W} (\theta - \hat{\theta}) + (\theta_0 - \theta)' \mathbf{D}^{-1}(\theta_0 - \theta) \\ & = (\theta - \hat{\theta})' \mathbf{W}' \mathbf{R}^{-1} \mathbf{W} (\theta - \hat{\theta}) + (\theta - \theta_0)' \mathbf{D}^{-1}(\theta - \theta_0) \\ & = (\theta - \hat{\theta})' (\mathbf{W}' \mathbf{R}^{-1} \mathbf{W} + \mathbf{D}^{-1}) (\theta - \hat{\theta}) + (\hat{\theta} - \theta_0)' (\mathbf{W}' \mathbf{R}^{-1} \mathbf{W}) (\mathbf{W}' \mathbf{R}^{-1} \mathbf{W} + \mathbf{D}^{-1}) \mathbf{D}^{-1} (\hat{\theta} - \theta_0) \end{aligned}$$

donde  $\hat{\theta} = (\mathbf{W}' \mathbf{R}^{-1} \mathbf{W} + \mathbf{D}^{-1})^{-1} (\mathbf{W}' \mathbf{R}^{-1} \mathbf{W} \hat{\theta} + \mathbf{D}^{-1} \theta_0)$ . Substituyendo  $\hat{\theta}$ , la expresión queda como:

$$\hat{\theta} = (\mathbf{W}' \mathbf{R}^{-1} \mathbf{W} + \mathbf{D}^{-1})^{-1} (\mathbf{W}' \mathbf{R}^{-1} \mathbf{y} + \mathbf{D}^{-1} \theta_0)$$

Considerando que  $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'(\mathbf{W}'\mathbf{R}^{-1}\mathbf{W})(\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{D}^{-1})\mathbf{D}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$  no depende de  $\boldsymbol{\theta}$ , entonces la distribución condicional de  $\boldsymbol{\theta}$  dado  $\sigma_{\rho}^2, \sigma_{\gamma^{(1)}}^2, \dots, \sigma_{\gamma^{(v)}}^2, \sigma_{e^{(1)}}^2, \dots, \sigma_{e^{(a)}}^2$  puede ser escrita como:

$$p(\boldsymbol{\theta} | \sigma_{\rho}^2, \sigma_{\gamma^{(1)}}^2, \dots, \sigma_{\gamma^{(v)}}^2, \sigma_{e^{(1)}}^2, \dots, \sigma_{e^{(a)}}^2, \mathbf{y}, \mathbf{h}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'(\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{D}^{-1})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right]$$