AGRISCIENTIA, 2010, VOL. XXVII: 19-26

Estimación bayesiana de componentes de (co)varianza en Brangus argentino para caracteres de res mediante el algoritmo FCG

Cantet, R. J. C. y A. N. Birchmeier

RESUMEN

Se emplearon los datos de 2273 toritos y vaquillonas Brangus para estimar las heredabilidades (h^2) y las correlaciones aditivas y ambientales de caracteres de calidad de carne medidos por ultrasonido. Los registros provenían del programa de evaluación genética de la Asociación Argentina de Brangus. Los caracteres medidos fueron el área del ojo del bife (AOB), el marmoreado (MB), la grasa dorsal (GD) y la grasa de cadera (GC). La edad media de los animales al momento de la medición fue 641 días en machos y 685 días en hembras. Los parámetros genéticos y ambientales fueron estimados mediante un algoritmo bayesiano conjugado. Los valores estimados de h^2 fueron 0,22, 0,16, 0,12 y 0,21, para AOB, GD, CC y MB, respectivamente. En términos generales, las estimaciones de las correlaciones genéticas y ambientales se encontraron cercanas a la cifra media de la literatura. Si bien los valores estimados de h^2 fueron inferiores al promedio de la investigación realizada en vacunos para carne, la variabilidad encontrada es suficiente como para que la respuesta a la selección por estos caracteres – empleando predicciones de los valores de cría calculadas con los parámetros estimados – sea moderadamente efectiva.

Palabras clave: componentes de (co)varianza, Brangus, caracteres de res, algoritmo FCG.

Cantet, R. J. C. and A. N. Birchmeier, 2010. Bayesian estimation of (co) variance components in Argentinian Brangus for carcass traits using the FCG algorithm. Agriscientia XXVII: 19-26

SUMMARY

Data on 2273 Brangus young bulls and heifers were used to estimate heritabilities (h^2) and genetics and environmental correlations for ultrasound carcass measures. Records were from the genetic evaluation program of Asociación Argentina de Brangus. Traits measured were rib-eye area (AOB), marbling (MB), back-fat thickness (GD), and hip-fat thickness (GC). Average ages of measure were 641 days in males and 685 in females. The genetic and environmental dispersion parameters were estimated by a conjugate Bayesian algorithm

Fecha de recepción: 17/11/09; fecha de aceptación: 25/03/10

(FCG). Estimates of h^2 were 0,22, 0,16, 0,12, and 0,21, for AOB, GD, CC, and MB, respectively. In general, estimates of genetic and environmental correlations were close to the average published values. Even tough estimates of h^2 were below the average of published estimates for beef cattle, the additive genetic variation found in the current study would lead to a moderate response to selection – using predictions of breeding value that are calculated with the estimate parameters.

Key words: (co)variance components, Brangus, carcass traits, FCG algorithm.

R. J. C. Cantet y A. N. Birchmeier. Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires, Av. San Martín 4453, 1417 Buenos Aires, Argentina Correspondencia a R. J. C. Cantet: rcantet@agro.uba.ar

INTRODUCCIÓN

La producción de carne vacuna en la Argentina, fundamentalmente la de cortes valiosos destinados al consumo en el mercado europeo (cuota Hilton), ha generado en el criador la necesidad de considerar en la selección caracteres de calidad de producto: carne tierna, magra y con atributos asociados a la salud del consumidor (bajos contenidos de colesterol y ácidos grasos saturados). Dado que las mediciones de caracteres como el área del ojo del bife, la grasa intramuscular, la grasa dorsal o la de cadera, se deben realizar a una edad avanzada en la vida del animal, la eficiencia de la selección empleando mediciones en las reses de la progenie de los reproductores (prueba de progenie) es baja. Alternativamente, el uso de técnicas de ultrasonido ha permitido evaluar los citados caracteres en los propios animales destinados a la selección (Wilson, 1992), lo que permite reducir el intervalo generacional, con su consecuente efecto positivo sobre el cambio genético por selección, y a una reducción de los costos de medición al tener que evaluar animales en la planta de faena. Las mediciones ecográficas en el bovino para carne son utilizadas entonces para la selección a través del cálculo del mérito genético, sean las “diferencias esperadas entre progenies” (DEP) o, su doble numérico, los “valores de cría estimados”.

En la actualidad la evaluación del mérito genético animal en todo el mundo se realiza mediante predicciones BLUP, siglas en inglés de “best linear unbiased prediction”, o predicción lineal insesgada de mínima varianza (Henderson, 1984), calculadas empleando modelos lineales mixtos sobre los registros de caracteres productivos. Esta metodología permite una gran flexibilidad en la especificación del valor esperado y la matriz de (co)varianzas genéticas y ambientales de los datos. La predicción BLUP multicarácter (Henderson & Quaas, 1976), para todos los animales evaluados, se realiza mediante la resolución simultánea del sistema de "ecuaciones de modelo mixto" (Henderson, 1984), que requiere estimaciones de (co)varianzas genéticas y ambientales, o de sus valores estandarizados: heredabilidades y correlaciones. La estimación de estos últimos parámetros es generalmente más imprecisa que la de las heredabilidades (Falconer & Mackay, 1996). En el mejoramiento genético animal, los componentes de (co)varianza se estiman a partir de los datos disponibles, mediante métodos basados en la función de verosimilitud, o estimadores Bayesianos. Así Cantet *et al*. (2004) presentaron el algoritmo FCG (Full Conjugate Gibbs), método basado en la generación de Cadenas de Markov mediante un esquema de Monte Carlo (MCMC). Este algoritmo es eficiente para procesar grandes bases de datos con un porcentaje de ‘información faltante’ (*missing data*).

La Asociación Argentina de Brangus ha recolectado una base de datos de caracteres de res (área de ojo de bife, marmoreado, grasa dorsal y grasa de cadera), para su programa de evaluación genética de reproductores, mediante el cálculo de las DEP que predicen la mitad del valor de cría de un reproductor. Esta presentación tiene dos objetivos: 1) presentar a investigadores de habla castellana el algoritmo bayesiano FCG; 2) estimar las (co)varianzas genéticas y ambientales para caracteres de res medidos por ultrasonido en Brangus argentino, utilizando el algoritmo FCG.

MATERIALES Y MÉTODOS

Base de datos

Se utilizaron los registros de 2273 animales provenientes de seis establecimientos adheridos a la Asociación Argentina de Brangus en el marco del convenio entre esta asociación y la Facultad de Agronomía, Universidad de Buenos Aires (Argentina). Los caracteres medidos fueron el área del ojo del bife (AOB), el marmoreado (MB), la grasa dorsal (GD) y la grasa de cadera (GC). El porcentaje de animales con carácter medido fue 97% para AOB, 98% para GD, 99% para GC y 96% para MB. Los estadísticos descriptivos de los datos se presentan en la Tabla 1.

Tabla 1. Descripción de la base de datos.

Carácter	N° Total	Promedio		Desvío estándar	Mínimo	Máximo
		Machos	Hembras			
AOB (cm²)	2214	62,5	58,9	12,37	31,50	118,80
GD (mm)	2256	3,78	5,02	2,38	1,00	25,00
GC (mm)	2244	3,20	4,04	1,65	0,80	14,20
MB (%)	2200	3,26	4,58	2,04	0,18	9,80

AOB = área del ojo del bife; GD = grasa dorsal; GC = grasa de cadera;
MB = marmoreado.

La edad promedio a la toma de mediciones fue de 641 días en machos y 685 días en hembras. Dado que se incluyeron todas las relaciones de parentesco conocidas de los animales con registros, el número total de animales evaluados fue de 4811.

Modelo de análisis multicarácter para estimar los parámetros de dispersión

Los datos se analizaron con un modelo animal multicarácter. Inicialmente desarrollado por Henderson & Quaas (1976), el modelo animal de evaluación genética multicarácter es aún hoy la metodología esencial de predicción del valor de cría en bovinos de carne en todo el mundo. A los caracteres de ultrasonido se los asoció con el peso del animal a la medición, como un carácter más a evaluar de modo de no remover la variación aditiva asociada con el tamaño del animal en las características de calidad de res. Para el registro del animal i (i = 1, ..., 4811) respecto del carácter j (j = 1, ..., 5), el modelo animal fue el siguiente:

$$y_{ij} = \boldsymbol{X}'_{ij}\,\beta_j + a_{ij} + e_{ij} \qquad (1)$$

siendo y_{ij}, a_{ij}, y e_{ij} las observaciones, los valores de cría y el error, respectivamente. El valor y_{ij} es potencialmente observable directamente, mientras que a_{ij} y e_{ij} no lo son. El vector $\boldsymbol{\beta}_j$ de efectos fijos para el carácter j se relaciona con las observaciones a través de un vector $\boldsymbol{X}'_{ij}$ de constantes conocidas, e incluyó los efectos de sexo y grupo de contemporáneos, como variables clasificatorias y la edad del animal a la medición como covariable ambiental. La suma de todos los niveles de efectos fijos llevó el orden de $\boldsymbol{\beta}$ a 218 × 1.

Para expresar el modelo en notación matricial, sea $\boldsymbol{y}$ el vector de las observaciones ordenado por carácter dentro de animal. Entonces el modelo (1) es igual a:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{e} \qquad (2)$$

La matriz $\boldsymbol{X}$ resulta de "apilar" los vectores $\boldsymbol{X}'_{ij}$, mientras que $\boldsymbol{Z}$ relaciona los datos con los valores de cría en $\boldsymbol{a}$. La matriz de covarianzas en $\boldsymbol{a}$ es igual a:

$$\mathrm{Var}(\boldsymbol{a}) = \begin{bmatrix} g_{1,1} & g_{1,2} & g_{1,3} & g_{1,4} & g_{1,5} \\ g_{2,1} & g_{2,2} & g_{2,3} & g_{2,4} & g_{2,5} \\ g_{3,1} & g_{3,2} & g_{3,3} & g_{3,4} & g_{3,5} \\ g_{4,1} & g_{4,2} & g_{4,3} & g_{4,4} & g_{4,5} \\ g_{5,1} & g_{5,2} & g_{5,3} & g_{5,4} & g_{5,5} \end{bmatrix} \otimes \boldsymbol{A} = \boldsymbol{G}_0 \otimes \boldsymbol{A} \qquad (3)$$

siendo $g_{kk'}$ la covarianza genética aditiva entre los efectos genéticos k y k', cuando k es distinto de k', e igual a la varianza del efecto k en caso que $k = k'$. Los elementos de la matriz $\boldsymbol{A}$ (4811 × 4811) son las relaciones aditivas entre los animales.

Dado que no todos los caracteres fueron observados en todos los individuos, se identificaron 32 (2^5) patrones de "pérdida de información", en función de que los pesos al nacer, al destete y final, hayan sido observados o no. El vector $\boldsymbol{e}$ de errores es ordenado por carácter dentro de cada animal con un determinado patrón de pérdida de información. Así, $\boldsymbol{e}_{(g\,1)}$, $\boldsymbol{e}_{(g\,2)}$,..., $\boldsymbol{e}_{(g\,tg)}$ son vectores de orden t_g × 1 para los diferentes animales con t_g caracteres observados dentro del patrón g. Los errores tienen esperanza igual a cero y, para el animal i con los

tres caracteres observados, la matriz de varianzas y covarianzas es igual a Var($\boldsymbol{e}_{(1i)}$) = $\boldsymbol{R}_0 = [r_{jj'}]$, siendo $r_{jj'}$ la (co)varianza ambiental entre los caracteres *i* y *j'*. Si el animal *i'* tenía registros incompletos dentro del patrón *g*, la varianza es Var($\boldsymbol{e}_{(i'g)}$) = $\boldsymbol{M}_g \boldsymbol{R}_0 \boldsymbol{M}_g'$. El vector $\boldsymbol{e}$ resultante tiene matriz de varianzas y covarianzas igual a:

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{I}_{n_1} \otimes \boldsymbol{M}_1 \boldsymbol{R}_0 \boldsymbol{M}_1' & \boldsymbol{0} & . & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n_2} \otimes \boldsymbol{M}_2 \boldsymbol{R}_0 \boldsymbol{M}_2' & . & \boldsymbol{0} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \boldsymbol{0} & \boldsymbol{0} & . \; . \; . & \boldsymbol{I}_{n_G} \otimes \boldsymbol{M}_G \boldsymbol{R}_0 \boldsymbol{M}_G' \end{bmatrix} = \bigoplus_{g=1}^{G} \boldsymbol{I}_{n_g} \otimes \boldsymbol{M}_g \boldsymbol{R}_0 \boldsymbol{M}_g' \tag{4}$$

Los parámetros a estimar son las matrices $\boldsymbol{G}_0$ y $\boldsymbol{R}_0$.

Estimación de componentes de (co)varianza mediante el algoritmo FCG (*Full conjugate Gibbs*) de Cantet et al. (2004)

Modelo animal con varios caracteres y con información incompleta

Un rasgo esencial del modelo animal multicarácter (1) es que algunos animales pueden tener registros incompletos para uno o más caracteres. Todos los individuos que tienen observaciones para los mismos caracteres ($t_g \leq t$), comparten el mismo "patrón" de caracteres observados y faltantes. Cada patrón puede representarse por una matriz ($\boldsymbol{M}_g$) que posee t_g filas y *t* columnas (Dominici *et al.*, 2001), donde *g* = 1, ... *G* , siendo *G* el número de patrones en un archivo en particular. Sea *n* el número total de animales con registros para al menos un carácter. Todos los elementos en cualquier fila de $\boldsymbol{M}_g$ son 0's excepto por un 1 en la columna donde se ubica el carácter observado. Por lo tanto, $\boldsymbol{M}_g = \boldsymbol{I}_t$ cuando $t_g = 0$. Por ejemplo, si se supone *t* = 6 y un patrón donde se observan los caracteres 1, 2 y 5. Entonces se tiene que:

$$\boldsymbol{M}_g = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Existen $2^t - 1$ diferentes matrices $\boldsymbol{M}_g$ relacionadas con cada uno de los potenciales patrones de datos observados y faltantes. Se referirá al patrón con todos los caracteres observados como *g* = 1, de modo tal que $\boldsymbol{M}_1 = \boldsymbol{I}_t$. Se supone que el patrón completo se observa en al menos *t* animales, y se denota n_g al número de animales que muestran el patrón *g*, con lo cual $n = \sum_{g=1}^{G} n_g$ de modo tal que *n* es el número total de animales con al menos un carácter observado.

Si los valores de cría y los errores siguen una distribución normal, la función de densidad conjunta es la siguiente:

$$p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{R}_0, \boldsymbol{M}_1, ..., \boldsymbol{M}_G) \propto |\boldsymbol{R}|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{a})' \boldsymbol{R}^{-1} (\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{a})\right] \tag{5}$$

Distribuciones *a priori*

Para evitar que la densidad posterior de los efectos fijos sea impropia (es decir que no integre a 1), se debe emplear una distribución previa para el vector (*p* × 1) $\boldsymbol{\beta}$ normal multivariada: $\boldsymbol{\beta} \sim N_p(\boldsymbol{0}, \boldsymbol{K})$. La idea es reflejar escaso o vago conocimiento *a priori*, con lo cual la matriz $\boldsymbol{K}$ será diagonal con elementos diagonales de valor elevado (por ejemplo, $k_{ii} > 10^8$). Esta especificación evita la presencia de distribuciones impropias en el modelo mixto (Hobert & Casella, 1996). La densidad a priori de $\boldsymbol{\beta}$ es entonces:

$$p(\boldsymbol{\beta} \mid \boldsymbol{K}) \propto \left|\prod_{i=1}^{p} k_{ii}\right|^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\sum_{i=1}^{p} \frac{\boldsymbol{\beta}_i^2}{k_{ii}}\right\} \tag{6}$$

Los valores de cría para los *t* caracteres en los *q* animales se distribuyen *a priori* como $\boldsymbol{a} \sim N_{rq}(\boldsymbol{0}, \boldsymbol{G}_0 \otimes \boldsymbol{A})$, de modo tal que:

$$p(\boldsymbol{a} \mid \boldsymbol{G}_0, \boldsymbol{A}) \propto |\boldsymbol{G}_0|^{-\frac{q}{2}} |\boldsymbol{A}|^{-\frac{r}{2}} \exp\left\{-\tfrac{1}{2}\boldsymbol{a}'(\boldsymbol{G}_0^{-1} \otimes \boldsymbol{A}^{-1})\boldsymbol{a}\right\} \tag{7}$$

La matriz de (co)varianzas aditivas $\boldsymbol{G}_0$ sigue *a priori* una distribución Wishart invertida (IW): $\boldsymbol{G}_0 \sim$ IW ($\boldsymbol{G}_0^*$, n_A), siendo $\boldsymbol{G}_0^*$ la matriz de (co)varianzas y n_A los "grados de credibilidad". En el enfoque bayesiano los grados de credibilidad representan el nivel de certeza o confianza que posee el analista en las distribuciones del modelo de trabajo. Por lo tanto:

$$p(\boldsymbol{G}_0 \mid \boldsymbol{G}_0^*, n_A) \propto |\boldsymbol{G}_0^*|^{\frac{n_A}{2}} |\boldsymbol{A}|^{-\frac{(n_A+r+1)}{2}} \exp\left\{-\tfrac{1}{2}\operatorname{tr}(\boldsymbol{G}_0^* \boldsymbol{G}_0^{-1})\right\} \tag{8}$$

Se presenta ahora la densidad *a priori* de $\boldsymbol{R}_0$. Si se hubieran observados todos los datos, en todos los caracteres de cada animal, la densidad previa de $\boldsymbol{R}_0$ hubiese sido W ($\boldsymbol{R}_0^*$, ν), siendo los "hiperparámetros" la matriz de (co)varianzas *a priori* $\boldsymbol{R}_0^*$ y los grados de credibilidad ν. De modo de acomodar todos los patrones de datos faltantes, Cantet *et al.* (2004) consideraron la siguiente distribución *conjugada* (es decir, cuando las densidades *a priori* y *a posteriori* tienen

una estructura similar) para $\boldsymbol{R}_0$, considerando la propuesta de Kadane & Trader (1987) y de Dominici *et al.* (2001):

$$p\left(\boldsymbol{R}_0 \mid \boldsymbol{R}_0^*, \boldsymbol{M}_1, \ldots, \boldsymbol{M}_G, \nu_g\right) \propto \prod_{g=1}^{G} \left|\boldsymbol{M}_g \boldsymbol{R}_0 \boldsymbol{M}_g'\right|^{-\frac{(\nu_g+2t_g+1)}{2}}$$

$$\times \quad \exp\left\{-\tfrac{1}{2}\operatorname{tr}\left[\boldsymbol{M}_g \boldsymbol{R}_0^* \boldsymbol{M}_g' \left(\boldsymbol{M}_g \boldsymbol{R}_0 \boldsymbol{M}_g'\right)^{-1}\right]\right\} \tag{9}$$

En palabras de Dominici *et al.* (2001), la especificación (9) imita la densidad natural conjugada *a priori* para el problema de inferencia en t_g dimensiones de las variables dentro del patrón *g*.

Distribución conjunta *a posteriori*

Multiplicando (5) con (6), (7), (8) y (9), se obtiene la densidad posterior conjunta de todos los parámetros, la cuál es proporcional a:

$$p\left(\boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{G}_0, \boldsymbol{R}_0 \mid \boldsymbol{y}, \boldsymbol{M}_1, \ldots, \boldsymbol{M}_G\right) \propto$$

$$|\boldsymbol{R}|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{a})'\boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{Z}\boldsymbol{a})\right] \exp\left\{-\tfrac{1}{2}\sum_{i=1}^{p}\frac{\boldsymbol{\beta}_i^2}{k_{ii}}\right\}$$

$$\exp\left\{-\tfrac{1}{2}\boldsymbol{a}'(\boldsymbol{G}_0^{-1}\otimes \boldsymbol{A}^{-1})\boldsymbol{a}\right\} \quad |\boldsymbol{G}_0|^{-\frac{(n_A+r+q+1)}{2}} \quad \exp\left\{-\tfrac{1}{2}\operatorname{tr}(\boldsymbol{G}_0^*\boldsymbol{G}_0^{-1})\right\}$$

$$\prod_{g=1}^{G}\left|\boldsymbol{M}_g \boldsymbol{R}_0 \boldsymbol{M}_g'\right|^{-\frac{(\nu_g+2t_g+1)}{2}} \quad \exp\left\{-\tfrac{1}{2}\operatorname{tr}\left[\boldsymbol{M}_g \boldsymbol{R}_0^* \boldsymbol{M}_g' \left(\boldsymbol{M}_g \boldsymbol{R}_0 \boldsymbol{M}_g'\right)^{-1}\right]\right\} \tag{10}$$

Para muestrear de la densidad (10) se puede utilizar un procedimiento de Monte Carlo con Cadenas de Markov (MCMC) que combina el algoritmo clásico DA (*data augmentation*) utilizado para caracteres múltiples normalmente distribuidos propuesto por Van Tassell & Van Vleck (1996), para $\boldsymbol{\beta}$, $\boldsymbol{a}$ y $\boldsymbol{G}_0$, con la especificación conjugada (9) para $\boldsymbol{R}_0$ propuesta por Cantet *et al.* (2004) a instancias del trabajo de Dominici *et al.* (2001).

Distribuciones posteriores condicionales para $\boldsymbol{\beta}$, $\boldsymbol{a}$ **y** $\boldsymbol{G}_0$

El algoritmo empleado por Van Tassell & Van Vleck (1996) involucra, en primer lugar, el muestreo de los efectos fijos y los valores de cría. Para ello se considera el siguiente sistema de ecuaciones lineales:

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}+\boldsymbol{K}^{-1} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}+\boldsymbol{G}_0\otimes\boldsymbol{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{a}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix} \tag{11}$$

La expresión (11) es una función de $\boldsymbol{K}$, $\boldsymbol{G}_0$ y $\boldsymbol{R}_0$, y permite escribir la distribución condicional conjunta posterior de $\boldsymbol{\beta}$ y de $\boldsymbol{a}$ como sigue:

$$\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{a} \end{bmatrix} \mid \boldsymbol{G}_0, \boldsymbol{R}_0, \boldsymbol{y} \sim N_{p+qr}\left(\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{a}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}+\boldsymbol{K}^{-1} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}+\boldsymbol{G}_0\otimes\boldsymbol{A}^{-1} \end{bmatrix}^{-1}\right) \tag{12}$$

La densidad normal (12) puede muestrearse por parámetro o por grupo de parámetros (Van Tassell & Van Vleck, 1996).

Para muestrear de la distribución posterior condicional de $\boldsymbol{G}_0$, se define la siguiente matriz

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{A}^{-1}\boldsymbol{a}_1 & \boldsymbol{a}_1'\boldsymbol{A}^{-1}\boldsymbol{a}_2 & . & . & \boldsymbol{a}_1'\boldsymbol{A}^{-1}\boldsymbol{a}_r \\ \boldsymbol{a}_2'\boldsymbol{A}^{-1}\boldsymbol{a}_1 & \boldsymbol{a}_2'\boldsymbol{A}^{-1}\boldsymbol{a}_2 & . & . & \boldsymbol{a}_2'\boldsymbol{A}^{-1}\boldsymbol{a}_r \\ . & . & \boldsymbol{a}_i'\boldsymbol{A}^{-1}\boldsymbol{a}_j & . & . \\ . & . & . & . & . \\ \boldsymbol{a}_r'\boldsymbol{A}^{-1}\boldsymbol{a}_1 & \boldsymbol{a}_r'\boldsymbol{A}^{-1}\boldsymbol{a}_2 & . & . & \boldsymbol{a}_r'\boldsymbol{A}^{-1}\boldsymbol{a}_r \end{bmatrix} \tag{13}$$

Van Tassell & Van Vleck (1996) observaron que la distribución posterior condicional de $\boldsymbol{G}_0$ es Wishart invertida con matriz de ponderación (*scaling*) $\boldsymbol{G}_0^*$ + $\boldsymbol{S}$ y grados de credibilidad $n_A + q$, de modo tal que:

$$p\left(\boldsymbol{G}_0 \mid \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{R}_0\right) \propto \left|\boldsymbol{G}_0\right|^{-\frac{(n_A+q+r+1)}{2}} \exp\left\{-\tfrac{1}{2}\operatorname{tr}\left[\left(\boldsymbol{G}_0^* + \boldsymbol{S}\right)\boldsymbol{G}_0^{-1}\right]\right\} \tag{14}$$

Distribución posterior condicional de $\boldsymbol{R}_0$

El procedimiento de muestreo de la matriz de (co)varianzas ambientales $\boldsymbol{R}_0$ difiere del algoritmo de "aumento de los datos", en el muestreo de las matrices de varianzas y covarianzas: por patrón en vez de muestrearse por errores individuales. Cantet *et al.* (2004) demostraron que la distribución posterior condicional de $\boldsymbol{R}_0$ es la siguiente densidad:

$$p\left(\boldsymbol{R}_0 \mid \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{G}_0\right) \propto \prod_{g=1}^{G}\left|\boldsymbol{M}_g \boldsymbol{R}_0 \boldsymbol{M}_g'\right|^{-\frac{(\nu_g+n_g+2t_g+1)}{2}}$$

$$\times \quad \exp\left\{-\tfrac{1}{2}\operatorname{tr}\left[\left(\boldsymbol{M}_g \boldsymbol{R}_0^* \boldsymbol{M}_g' + \boldsymbol{E}_g\right)\left(\boldsymbol{M}_g \boldsymbol{R}_0 \boldsymbol{M}_g'\right)^{-1}\right]\right\} \tag{15}$$

Dominici *et al.* (2001) propusieron muestrear la matriz de (co)varianzas de una distribución normal multivariada, mediante un esquema MCMC basado en un análisis recursivo a partir de la distribución Wishart invertida descripta por Bauwens *et al.* (1999) en su teorema A.17. Este enfoque es

equivalente a caracterizar la expresión (15) como una densidad Wishart invertida generalizada (Brown *et al.*, 1994). El algoritmo para muestrear $\boldsymbol{R}_0$ consiste de los siguientes pasos. En primer lugar, se completan las matrices de hiperparámetros para cada patrón, respecto de sus filas y columnas faltantes. Se indica dicha matriz para el patrón *g* como $\boldsymbol{R}_g^*$, la que está compuesta por las submatrices $\boldsymbol{R}_{oo}^{*g}$ $\boldsymbol{R}_{om}^{*g}$ y $\boldsymbol{R}_{mm}^{*g}$ para los caracteres observados, la covarianza entre observados y faltantes y la (co)varianza entre los caracteres faltantes, respectivamente. Una vez que las matrices $\boldsymbol{R}_g^*$ han sido calculadas, se muestrea $\boldsymbol{R}_0$ de una densidad Wishart invertida con matriz de ponderación igual a la suma de las matrices $\boldsymbol{R}_g^*$ obtenidas en el paso anterior. El muestreo de $\boldsymbol{R}_0$ consta de los siguientes pasos:

1. *Muestreo de las hipercovarianzas entre los caracteres observados y faltantes en el patrón g.*

En este paso se requiere calcular el producto de $\boldsymbol{R}_{oo}^{*g}$ con una matriz muestreada en forma aleatoria:

$$\left(\boldsymbol{R}_{oo}^{*g}\right)^{-1}\boldsymbol{R}_{om}^{*g}\,|\,\boldsymbol{R}_0,\boldsymbol{R}_{mm.o} \sim \boldsymbol{R}_{oo}^{*g} \times N_{t_g \times (t-t_g)}\left[\boldsymbol{R}_{oo}^{-1}\boldsymbol{R}_{om},\ \boldsymbol{R}_{oo}^{-1} \otimes \boldsymbol{R}_{mm.o}\right] \tag{16}$$

En la práctica, el muestreo de la matriz que sigue una distribución normal "matriz-variada" (Bauwens *et al.*, 1999) en (16) se puede lograr mediante las siguientes operaciones. Primero, se multiplican: 1) la descomposición de Cholesky de la matriz de (co)varianzas $\left(\boldsymbol{R}_{oo}^{*g}\right)^{-1}$ ¼ $\boldsymbol{R}_{mm.o}$ por un vector de variables normales estándar de orden $[t_g(t-t_g)]$, con 2) la matriz $\boldsymbol{R}_{mm.o} = \boldsymbol{R}_{mm} - \boldsymbol{R}_{mo}\boldsymbol{R}_{oo}^{-1}\boldsymbol{R}_{om}$. El vector aleatorio resultante se transforma en una matriz de orden $t_g \times (t-t_g)$ mediante la inversa de la operación vec: la primera columna de la matriz se forma con los primeros t_g elementos del vector, la segunda columna con los t_g elementos siguientes, y así sucesivamente hasta terminar con la columna $(t-t_g)$. Las matrices $\boldsymbol{R}_{oo}$, $\boldsymbol{R}_{mm}$ y $\boldsymbol{R}_{mo}$ son submatrices de $\boldsymbol{R}_0$ obtenidas en la iteración previa del muestreo FCG. Posteriormente $\boldsymbol{R}_{oo}^{-1}\boldsymbol{R}_{om}$, que es la media de la distribución matriz-variada, se debe sumar a la matriz aleatoria de orden $t_g \times (t - t_g)$. Finalmente, la matriz resultante debe premultiplicarse por $\boldsymbol{R}^*_{oo} = \boldsymbol{R}_{0oo}^* + \boldsymbol{E}_g$, que es la matriz de orden $t_g \times t_g$, de los caracteres observados en el patrón *g*. La matriz $\boldsymbol{R}_{0oo}^*$ contiene las hipervarianzas y covarianzas de $\boldsymbol{R}_0^*$. La matriz con las sumas de los errores al cuadrado en el patrón *g* ($\boldsymbol{E}_g$) fue definida por Cantet *et al.* (2004).

2. *Muestreo de la matriz de hipercovarianzas entre los caracteres faltantes en el patrón g.*

Este paso se realiza muestreando de la siguiente distribución:

$$\boldsymbol{R}_{mm.o}^{*g}\,|\,\boldsymbol{R}_0 \sim \mathrm{W}_m\left(\nu_g + 2t, \boldsymbol{R}_{mm.o}\right) \tag{17}$$

donde W_m indica la distribución Wishart, $\boldsymbol{R}_{mm.o}$ es la matriz de ponderación y ν_g son los grados de credibilidad más dos veces el número de caracteres.

3. *Cálculo de la matriz de varianzas incondicional entre los caracteres faltantes en el patrón g.*

Este cálculo se realiza mediante la siguiente fórmula:

$$\boldsymbol{R}_{mm}^{*g} = \boldsymbol{R}_{mm.o}^{*g} + \boldsymbol{R}_{mo}^{*g}\left(\boldsymbol{R}_{oo}^{*g}\right)^{-1}\boldsymbol{R}_{om}^{*g} \tag{18}$$

4. *Cálculo de la matriz de hipercovarianzas.*

Sea $\boldsymbol{P}_g$ una matriz de permutación que recupera las posiciones de los caracteres observados y faltantes del patrón *g*, dentro de la matriz de (co) varianzas completa. Por lo tanto, la contribución del patrón *g* a la matriz de hipercovarianzas para muestrear $\boldsymbol{R}_0$ es igual a:

$$\boldsymbol{P}_g \begin{bmatrix} \boldsymbol{R}_{oo}^{*g} & \boldsymbol{R}_{om}^{*g} \\ \boldsymbol{R}_{mo}^{*g} & \boldsymbol{R}_{mm}^{*g} \end{bmatrix} \boldsymbol{P}_g' = \boldsymbol{R}_g^* \tag{19}$$

con lo cual la matriz de hipercovarianzas completa es igual a $\sum_{g=1}^{G} \boldsymbol{R}_g^*$.

5. *Muestreo de* $\boldsymbol{R}_0$

$$p\left(\boldsymbol{R}_0 \mid \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{G}_0\right) \sim \mathrm{IW}\left(\sum_{g=1}^{G} \nu_g + n + (G-1)(t+1), \left(\sum_{g=1}^{G} \boldsymbol{R}_g^*\right)\right) \tag{20}$$

Resumen del algoritmo FCG

1. Armar y resolver las ecuaciones (11).
2. Muestrear $\boldsymbol{\beta}$ y $\boldsymbol{a}$ de (12).
3. Calcular los residuales: $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{a}$
4. Para cada patrón hacer lo siguiente:

4.1. Muestrear las hipercovarianzas entre los

caracteres observados y faltantes en el patrón correspondiente, mediante (16);

4.2. Muestrear la matriz de hipercovarianzas entre los caracteres faltantes en el patrón correspondiente, usando la distribución Wishart invertida en (17);

4.3. Calcular la matriz de varianzas incondicionales entre los caracteres faltantes en el patrón correspondiente, usando (18);

5. Calcular la matriz de hipercovarianzas para $\boldsymbol{R}_0$ sumando todas las matrices $\boldsymbol{R}_g^*$ de modo de obtener $\sum_{g=1}^{G} \boldsymbol{R}_g^*$;

6. Muestrear $\boldsymbol{R}_0$ a partir de (20);

7. Calcular $\boldsymbol{S}$;

8. Muestrear $\boldsymbol{G}_0$ de (8), y volver al paso 1.

El algoritmo FCG fue calculado en una cadena de largo 200000. Luego de controlar convergencia como sugirieron Cantet *et al.* (2004), se obtuvieron las medias de las distribuciones posteriores de cada parámetro en $\boldsymbol{G}_0$ y $\boldsymbol{R}_0$.

RESULTADOS

Las estimaciones centrales de los parámetros de dispersión, los valores esperados estimados de las distribuciones *a posteriori* de $\boldsymbol{G}_0$ y $\boldsymbol{R}_0$, calculados mediante FCG, fueron respectivamente iguales a

$$\hat{\boldsymbol{G}}_0 = \begin{bmatrix} 633{,}60 & 58{,}95 & 9{,}92 & 6{,}04 & 6{,}82 \\ 58{,}95 & 23{,}56 & -0{,}48 & -0{,}04 & 0{,}08 \\ 9{,}92 & -0{,}48 & 0{,}82 & 0{,}25 & 0{,}07 \\ 6{,}04 & -0{,}04 & 0{,}25 & 0{,}32 & 0{,}04 \\ 6{,}82 & 0{,}08 & 0{,}07 & 0{,}04 & 0{,}84 \end{bmatrix}$$

$$\hat{\boldsymbol{R}}_0 = \begin{bmatrix} 2733{,}44 & 251{,}06 & 14{,}98 & 11{,}22 & -11{,}75 \\ 251{,}06 & 82{,}34 & 4{,}49 & 3{,}00 & 0{,}54 \\ 14{,}98 & 4{,}49 & 4{,}31 & 1{,}82 & 0{,}02 \\ 11{,}22 & 3{,}00 & 1{,}82 & 2{,}22 & 0{,}03 \\ -11{,}75 & 0{,}54 & 0{,}02 & 0{,}03 & 308 \end{bmatrix}$$

La Tabla 2 muestra los parámetros de dispersión en términos de heredabilidades y correlaciones.

Tabla 2: Estimaciones de las heredabilidades (en la diagonal y subrayada), correlaciones genético aditivas (sobre la diagonal) y correlaciones ambientales (bajo la diagonal) para caracteres de res en animales Brangus argentino.

	Peso	AOB	GD	GC	MB
Peso	0,19	0,48	0,43	0,42	0,29
AOB	0,59	0,22	-0,10	-0,01	0,02
GD	0,13	0,23	0,16	0,48	0,08
GC	0,14	0,22	0,58	0,12	0,09
MB	-0,12	0,03	0,008	0,01	0,21

AOB = área del ojo del bife; GD = grasa dorsal; GC = grasa de cadera; MB = marmoreado.

DISCUSIÓN

En el mejoramiento genético animal metodológico, el empleo del enfoque bayesiano para la estimación paramétrica data de más de dos décadas (Gianola y Fernando, 1986). En el texto de Sorensen y Gianola (2002) se compilan los desarrollos metodológicos bayesianos realizados por los mejoradores animales en detalle. Frente al proverbial método de "máxima verosimilitud residual" (o restringida, REML), para la estimación de componentes de (co)variancia en modelos mixtos normales, el enfoque bayesiano tiene la ventaja de contar con la distribución posterior marginal de los parámetros. Con dicha densidad se puede calcular la varianza posterior de los estimadores, u obtener un intervalo de alta densidad posterior de los parámetros, a fin de realizar aseveraciones probabilísticas sobre el valor de éstos. Estos estadísticos sólo pueden ser aproximados en REML usando teoría asintótica. Dado que ambos estimadores utilizan la función de verosimilitud marginalizada con respecto a los efectos fijos, sus performances tienden a igualarse a medida que la información en los datos aumenta. Cuando la información en los datos es escasa, los métodos bayesianos cuentan con la ventaja del uso de la información *a priori*. Asimismo, el desarrollo de métodos MCMC (por ejemplo, Sorensen y Gianola, 2002) han posibilitado el cálculo de la distribución posterior de los parámetros de dispersión, evitando tener que invertir matrices, como ocurre en ciertos algoritmos para obtener los estimadores REML. Esto permite aumentar el tamaño de las bases de datos a analizar.

La estimación de heredabilidad (h^2) para AOB (0,22) resultó inferior al valor promedio de la literatura e igual a 0,40 (AAABG, 2009), en particular para animales de raza Brangus: 0,29 (Moser *et al.*, 1998) y 0,31 (Stelzleni *et al.*, 2002). Si bien, la h^2 estimada para MB (0,21) es inferior al promedio de la literatura, que es igual a 0,45 (AAABG, 2009), éste valor es superior a la estimación hecha con

toros y vaquillonas Brangus realizada por Stelzleni *et al.* (2002) e igual a 0,16. Asimismo, la h^2 de GD (0,16) es intermedia a las reportadas por Moser *et al.* (1998) de 0,11 y por Stelzleni *et al.* (2002) de 0,26. No se han encontrado estimaciones de h^2 en Brangus para GC, aunque el promedio de la literatura, 0.41 (AAABG, 2009), es superior al valor aquí obtenido (0,12). En términos generales, las estimaciones de las correlaciones genéticas y ambientales se encontraron cercanas a la cifra media de la literatura. Si bien los valores estimados de h^2 fueron inferiores al promedio de la investigación realizada en vacunos para carne, la variabilidad encontrada es suficiente como para que la selección por estos caracteres –empleando predicciones del valor de cría calculadas con los parámetros estimados– sea moderadamente efectiva. El análisis de los parámetros estimados indica que las (co) varianzas aditivas fueron similares a las encontradas en la bibliografía, mientras que fueron mayores las magnitudes de las (co)varianzas ambientales. En consecuencia, sería deseable re-estimar las heredabilidades y correlaciones aditivas y ambientales cuando la base de datos de caracteres de res calculados por ultrasonido en el programa de evaluación genética, tenga al menos el doble de datos que la base actual.

AGRADECIMIENTOS

La financiación para esta investigación proviene de subsidios de la Secretaría de Ciencia y Técnica, Universidad de Buenos Aires (UBACyT G042/08), de la Agencia Nacional de Ciencia y Tecnología (PICT 1863/06) y de la Asociación Argentina de Brangus mediante el Convenio Facultad de Agronomía UBA – AAB.

BIBLIOGRAFÍA

AAABG, 2009. Genetic parameters. *In*: http://www.gparm.csiro.au/index.html. Fecha de consulta: 15 de octubre de 2009.

Bauwens, L.; M. Lubrano and J.F. Richard, 1999. Bayesian inference in dynamic econometric models. Oxford University Press.

Brown, P.J.; N.D. Le and J.V. Zidek, 1994. Inference for a covariance matrix, in: Freeman P.R. and A.F.M. Smith (Eds.), Aspects of uncertainty: a tribute to D. V. Lindley, Wiley, New York, 1994, pp. 77-92.

Cantet, R.J.C.; A.N. Birchmeier and J.P. Steibel, 2004. Full conjugate analysis of normal multiple traits with missing records using a generalized inverted Wishart distribution. Genet. Sel. Evol. 36:49-64.

Dominici, F.; G. Parmigiani and M. Clyde, 2001. Conjugate analysis of multivariate normal data with incomplete observations. Can. J. Stat. 28:533-550.

Falconer D.S. and T.F.C. Mackay, 1996. Introduction to Quantitative Genetics. Fourth Edition. Longman Group Ltd.

Gianola, D. and R.L. Fernando, 1986. Bayesian methods in animal breeding theory. J. Anim. Sci. 63: 217–244.

Henderson, C.R., 1984. Application of linear models in animal breeding. University of Guelph, Guelph.

Henderson, C.R. and R.L. Quaas, 1976. Multiple trait evaluation using relatives' records. J. Anim. Sci. 43:1188–1197.

Hobert, J.P. and G. Casella, 1996. The effects of improper priors on Gibbs sampling in hierarchical linear models. J. Amer. Statist. Assoc. 91:1461-1473.

Kadane, J.B. and R.L. Trader, 1988. A Bayesian treatment of multivariate normal data with observations missing at random. Statistical decision theory and related topics IV, 1:225-234.

Moser, D.W.; J.K. Bertrand, I. Misztal, L.A. Kriese and L.L. Benyshek, 1998. Genetic parameter estimates for carcass and yearling ultrasound measurements in Brangus cattle. J. Anim. Sci. 76:2542–2548.

Sorensen, D. and D. Gianola, 2002. Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics. First Edition. Springer-Verlag, New York.

Stelzleni, A.M.; T.L. Perkins, A.H. Brown Jr., F.W. Pohlman, Z.B. Johnson and B.A Sandelin, 2002. Genetic parameter estimates of yearling live animal ultrasonic measurements in Brangus cattle. J. Anim. Sci. 80:3150-3153.

Van Tassell, C.P. and L.D. Van Vleck, 1996. Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. J. Anim. Sci. 74:2586-2597.

Wilson, D. E., 1992. Application of ultrasound for genetic improvement. J. Anim. Sci. 70: 973-983.