

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE VETERINARIA
Departamento de Producción Animal



TESIS DOCTORAL

**Métodos de estimación de parámetros genéticos en el vacuno
frisón español**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Juan Pablo Gutiérrez García

Directores

Marcos Rico Gutiérrez
Francisco Javier Cañón Ferreras

Madrid

R. 7056

172.272-E
GUT

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Veterinaria

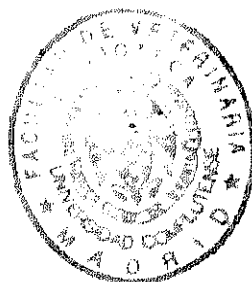
Departamento de Producción Animal

BIBLIOTECA UCM



5300899684

**METODOS DE ESTIMACION DE
PARAMETROS GENETICOS EN EL
VACUNO FRISON ESPAÑOL**



Juan Pablo Gutiérrez García

Madrid, 1992



La Tesis Doctoral de D. Juan Pablo

Gutiérrez García

Titulada Métodos de estimación de parámetros
genéticos en el vacuno fino español.

Director Dr. D. Marcos Rica Gutiérrez y D. Enrique Cañón

fue leída en la Facultad de Veterinaria

de la UNIVERSIDAD COMPLUTENSE DE MADRID, el día 21 ..

de Noviembre .. de 19 91., ante el tribunal

constituido por los siguientes Profesores:

PRESIDENTE Miguel Vallejo Vicente

VOCAL Antonio Muñoz de ma

VOCAL M. Jesús Caraballo

VOCAL Jesús Piedrafita

SECRETARIO Concepción Salgado Morales

.....
habiendo recibido la calificación de

Apto "cum laude" por unanimidad

Madrid, a 21 de Noviembre de 19 91.

EL SECRETARIO DEL TRIBUNAL.

Salgado

UNIVERSIDAD COMPLUTENSE DE MADRID
DEPARTAMENTO DE PRODUCCION ANIMAL
FACULTAD DE VETERINARIA

METODOS DE ESTIMACION DE PARAMETROS
GENETICOS EN EL VACUNO FRISON ESPAÑOL

MEMORIA
QUE PARA OPTAR AL GRADO DE DOCTOR EN VETERINARIA

PRESENTA

Juan Pablo Gutiérrez García

DIRECTORES:

Dr. D. Marcos Rico Gutiérrez
Dr. D. Javier Cañón Ferreras

MADRID, OCTUBRE 1991

A mi padre,
que con la conclusión de este trabajo verá
reflejada una de sus máximas ilusiones

AGRADECIMIENTOS

Es un honor para mí poder expresar mi gratitud a todos los que colaboraron en el desarrollo de este estudio.

En primer lugar, me siento sincera y profundamente agradecido a Marcos Rico y Javier Cañón, mis dos directores de tesis doctoral, no sólo por su constante disposición durante el desarrollo de este trabajo, sino también porque, con sus diferentes enfoques de la mejora genética, fueron los pilares de mi formación profesional en este campo.

Quiero expresar también mi agradecimiento a todos aquellos que, en alguna ocasión, contestaron a todas las dudas que se me plantearon durante el desarrollo de este estudio: Robin Thompson, Miguel Pérez Enciso, M^a Jesús Carabaño, etc. De forma muy especial, mi agradecimiento en este sentido a Keith George Boldman, gran profesional y estupenda persona, no sólo porque fue quien me cedió los programas utilizados para la estimación de componentes de varianza, sino también por su inestimable ayuda posterior, sin la cuál seguramente no hubiera podido concluir este trabajo.

Finalmente quiero agradecer a mis amigos, a mi familia, y a todos los compañeros del Departamento de Producción Animal de la Facultad de Veterinaria, su constante apoyo y estímulo para que este trabajo pudiera llegar a buen fin.

INDICE

I. INTRODUCCION	1
Notación y símbolos utilizados	6
II. REVISION BIBLIOGRAFICA	9
1.- Consanguinidad	9
2.- Componentes de varianza	10
2.1.- Evolución de los métodos de estimación	10
2.2.- Búsqueda de superación de las dificultades de cálculo	16
2.2.1.- Aproximaciones metodológicas	16
2.2.2.- Modelos	18
2.2.3.- Algoritmos	21
2.3.- Comparación de métodos	23
3.- Estimaciones anteriores de parámetros genéticos de interés	25
III. MATERIAL	32
1.- Datos	32
1.1.- Consanguinidad	32
1.2.- Estimación de componentes de varianza	33
IV. METODOLOGIA	44
1.- Consanguinidad	44
1.1.- Estrategia de cálculo	47

2.- Estimación de componentes de varianza	51
2.1.- Modelos	52
2.1.1.- Modelo animal	52
2.1.2.- Modelo padre y vaca jerarquizada	53
2.1.2.1.- Modelo padre y vaca jerarquizada con grupos genéticos	55
2.2.- Procedimientos de estimación	57
2.2.1.- Modelo animal	57
2.2.1.1.- Aspectos computacionales ...	63
2.2.1.1.1.- Factorización de Cholesky	65
2.2.1.1.2.- Aproximación al máximo de la función. El método Simplex .	67
2.2.2.- Modelo padre y vaca jerarquizada	69
2.2.2.1.- Aspectos computacionales ...	78
2.2.2.1.1.- Métodos para acelerar la convergencia ...	79
2.2.2.1.2.- Métodos para reducir cálculos en cada iteración	80
2.3.- Programas utilizados	81

3.- Simulación	81
V. RESULTADOS	86
1.- Consanguinidad	86
2.- Estimación de componentes de varianza	86
2.1.- Datos reales	86
2.1.1.- Tendencia genética	94
2.2.- Primer análisis. Comparación de tres modelos ..	95
2.2.1.- Simulación	95
2.2.2.- Resultados	96
2.3.- Segundo análisis. Comparación de dos modelos	
con y sin efectos fijos	98
2.3.1.- Simulación	98
2.3.2.- Resultados	99
2.4.- Tercer análisis. Comparación con distintas	
heredabilidades en un solo carácter	104
2.4.1.- Simulación	104
2.4.2.- Resultados	105
2.5.- Resultados de computación	106
VI. DISCUSION	107
1.- Consanguinidad	107
2.- Componentes de varianza	108
2.1.- Análisis con datos reales. Modelo padre	108
2.1.1.- Tendencia genética	110

2.2.- Primer análisis. Comparación de tres modelos	113
2.3.- Segundo análisis. Comparación de dos modelos con y sin efectos fijos	121
2.4.- Tercer análisis. Comparación con distintas heredabilidades en un sólo carácter	125
2.5.- Resultados de computación	126
VII. CONCLUSIONES	133
VIII. RESUMEN	134
IX. BIBLIOGRAFIA	137

INDICE DE TABLAS

1.- Revisión de estimaciones de heredabilidades en el ganado vacuno frisón	26
2.- Distribución de los individuos según el número de generaciones	33
3.- Distribución de las vacas según el número de padres conocidos	41
4.- Machos pertenecientes a cada nivel del efecto grupo genético	55
5.- Distribución de las vacas y los registros según el grupo genético del padre	56
6.- Estimaciones de componentes de varianza obtenidos en la población original con el modelo padre	87
7.- Estimaciones de componentes de varianza y parámetros genéticos de interés, obtenidos con el modelo padre a partir de los datos reales	87
8.- Media y varianzas de las poblaciones original y simulada	96

9.- Primer análisis. Resultados obtenidos	97
10.- Medias y varianzas de las poblaciones original y simuladas	98
11.- Comparación de las varianzas obtenidas para cada réplica. <u>CARACTER</u> : Producción de leche	100
12.- Comparación de las varianzas obtenidas para cada réplica. <u>CARACTER</u> : Producción de grasa	101
13.- Variación, absoluta y en tanto por uno, de la estimación de la <u>varianza genética aditiva</u> respecto al valor esperado en las tres réplicas, con los dos modelos utilizados, con y sin la participación de los efectos fijos	102
14.- Variación, absoluta y en tanto por uno, de la estimación de la <u>varianza ambiental permanente</u> respecto al valor esperado en las tres réplicas, con los dos modelos utilizados, con y sin la participación de los efectos fijos	102

- 15.- Variación, absoluta y en tanto por uno, de la estimación de la varianza residual respecto al valor esperado en las tres réplicas, con los dos modelos utilizados, con y sin la participación de los efectos fijos 103
- 16.- Variación, absoluta y en tanto por uno, de la estimación de la varianza total respecto al valor esperado en las tres réplicas, con los dos modelos utilizados, con y sin la participación de los efectos fijos 103
- 17.- Comparación de los resultados obtenidos, cuando se modifica la heredabilidad, manteniendo la varianza ambiental permanente y los valores extraídos aleatoriamente de una distribución normal. Entre paréntesis aparecen las diferencias en tanto por uno respecto al valor esperado CÁRACTER: Producción de grasa 105
- 18.- Comparación de necesidades computacionales en el tercer experimento. Carácter grasa 106
- 19.- Comparación de las heredabilidades obtenidas con otras anteriores de la misma población 108

20.- Primer análisis. Comparación de los resultados obtenidos con los esperados. Las diferencias, entre paréntesis, son expresadas en tanto por uno respecto al valor esperado	114
21.- Variación, absoluta y en porcentaje, de las estimaciones obtenidas respecto a las esperadas, en las tres réplicas, con los dos modelos utilizados al incluir los efectos fijos y al no incluirlos	123
22.- Variación, absoluta y en porcentaje, de las estimaciones obtenidas respecto a las esperadas, en las tres réplicas, con los dos modelos utilizados al no incluir los efectos fijos	124
23.- Estimaciones de componentes de varianza finales en comparación con valores iniciales y con valores intermedios	127
24.- Estimaciones de la heredabilidad y la repetibilidad en comparación con valores iniciales e intermedios ..	128

INDICE DE FIGURAS

1.- Distribución del número de vacas por rebaños	39
2.- Distribución del número de registros por rebaños	40
3.- Distribución de los toros según el número de hijas ...	40
4.- Distribución de los toros según el número de lactaciones	42
5.- Distribución de las vacas según el número de lactaciones	42
6.- Distribución de los años representados en cada rebaño	43
7.- Evolución de los componentes de varianza del modelo a través de las iteraciones. <u>CARACTER: LECHE</u>	88
8.- Evolución de los componentes de varianza de interés a través de las iteraciones. <u>CARACTER: LECHE</u>	89
9.- Evolución de los parámetros genéticos de interés a través de las iteraciones. <u>CARACTER: LECHE</u>	90

10.- Evolución de los componentes de varianza del modelo a través de las iteraciones. <u>CARACTER:</u> GRASA	91
11.- Evolución de los componentes de varianza de interés a través de las iteraciones. <u>CARACTER:</u> GRASA	92
12.- Evolución de los parámetros genéticos de interés a través de las iteraciones. <u>CARACTER:</u> GRASA	93
13.- Tendencia genética observada por el resultado del efecto grupo genético	94
14.- Distribución <u>óptima</u> de las vacas según el número de lactaciones	116
15.- Distribución <u>real</u> de las vacas según el número de lactaciones	116
16.- Distribución de las vacas según el número de lactaciones que <u>impedimenta</u> la estimación de la varianza ambiental permanente	116
17.- Distribución <u>óptima</u> de los toros según el número de hijas	119

18.- Distribución <u>real</u> de los toros según el número de hijas	119
19.- Distribución <u>página</u> de los toros según el número de hijas	119
20.- Comparación (entre modelos) de la evolución de la varianza genética a través de las iteraciones	130
21.- Comparación (entre modelos) de la evolución de la varianza ambiental permanente a través de las iteraciones	131
22.- Comparación (entre modelos) de la evolución de la varianza residual a través de las iteraciones	132

I. INTRODUCCION

La estimación de los parámetros genéticos se presenta, siempre, como el paso previo a la predicción del valor genético de los animales, en los esquemas de mejora genética de cualquier especie ganadera.

Además, las varianzas de los efectos aleatorios son incluidas en la evaluación de reproductores mediante técnicas estadísticas, con el fin de obtener predictores de estos efectos, así como estimadores de los efectos fijos, que posean la propiedad de mínima varianza muestral.

Por otro lado, ciertos parámetros, función de estas varianzas, tales como la heredabilidad, repetibilidad o correlaciones entre caracteres, son necesarios para otras tareas, tales como, comparar esquemas de selección alternativos, al poder predecir la respuesta genética sin necesidad de llevar a cabo el proceso de selección, o calcular los coeficientes de índices de selección combinados.

El desconocimiento de los valores exactos de los efectos aleatorios, como por ejemplo, el valor genético de los animales, impide conocer el verdadero valor de sus varianzas, y nos obliga a obtener estimaciones aproximadas de las mismas.

Por todo ello, dentro del contexto de la mejora animal, la estimación de componentes de varianza, ha sido y seguirá siendo, uno de los puntos donde los investigadores concentrarán sus mayores esfuerzos.

Se podría pensar que no hay muchos problemas para la obtención de estimaciones precisas de componentes de varianza, dado que las bases de datos disponen a menudo de gran número de registros. Sin embargo, mientras que los experimentos de laboratorio son diseñados desde un comienzo para lograr estimaciones eficientes de parámetros genéticos (Hill, 1970, Thompson, 1976, 1977), por desgracia, los datos de campo suelen ser desequilibrados, de estructura muy particular, y normalmente han sido sometidos a algún proceso de selección. Ello hace que sólo sofisticadas técnicas estadísticas nos permitan obtener estimaciones que posean ciertas propiedades deseadas.

Hasta ahora, el REMI (máxima verosimilitud restringida) ha sido el método de elección para la estimación de componentes de varianza en ganado vacuno lechero, normalmente bajo un modelo padre. Mediante el mismo se asume que los registros de la progenia son originados por la mitad del valor de mejora del padre, ignorando así la información aportada por la madre.

El interés sobre el modelo animal, conceptualmente más sencillo, ha aumentado recientemente. Mediante este modelo cada

registro permite conseguir información directa del genotipo del animal sobre el que ha sido hecha la medición, obteniéndose la estimación de su valor de mejora así como las de otros animales con los que está emparentado, aunque éstos no tengan registros de producción. Debido al coste computacional muy superior que supone la estimación de componentes de varianza mediante un modelo animal, debiera estudiarse, en cada situación, si otros modelos más sencillos producen estimaciones suficientemente precisas.

Otro parámetro genético de interés, el coeficiente de consanguinidad, puede interesar para valorar de forma más precisa a los reproductores, así como para evitar el aumento de la misma. Cuando el valor medio de la endogamia es muy elevado, si se incluyese en el análisis, la clasificación de los animales por su mérito genético podría ser diferente de la obtenida si se ignorase. Además, no está clara la influencia que puede tener en los valores obtenidos para las varianzas de los efectos aleatorios cuando su existencia es ignorada. Por todo ello, es preciso calcular los coeficientes de endogamia antes de proceder a un análisis de los datos sin incluir este parámetro, ya que el resultado podría ser discutible.

En el extremo contrario, la información aportada por las relaciones de parentesco entre los machos, en un modelo padre, podría ser tan escasa como para no producir diferencias entre un modelo en el que fueran incluidas y otro en el que se ignorasen.

La validez de la utilización de modelos más sencillos, podría ser aclarada por comparación entre ellos y el considerado más completo. Sin embargo, esta es una de las tareas más confusas en el ámbito de la mejora genética animal, ya que los verdaderos valores genéticos son desconocidos. En consecuencia, no existe un procedimiento general de hacer confrontaciones entre las predicciones o estimaciones obtenidas y los valores auténticos, para ver cuál de los modelos es más preciso. Un procedimiento posible es utilizar todos ellos sobre la misma base de datos, pero la única conclusión que se puede extraer es si difieren mucho o poco al aplicarlos sobre una determinada población. Por otro lado, cabe la posibilidad de que sean probados sobre datos simulados, aunque en este caso, las conclusiones obtenidas no son extrapolables a poblaciones que difieran de la utilizada, bien en la estructura de los datos, en los parámetros, o en ambos a la vez.

Mediante el desarrollo de este trabajo se propone una metodología a seguir para obtener estimaciones lo menos costosas posibles a la vez que suficientemente precisas, de los principales parámetros genéticos de interés en las poblaciones de animales domésticos. Para este estudio se utilizó una muestra española de ganado vacuno de raza Frisóna. En primer lugar se obtuvieron los coeficientes de endogamia de los animales. Posteriormente, para comparar los diferentes modelos utilizables en la estimación de componentes de varianza, se simularon, los efectos considerados

normalmente como aleatorios en la población, manteniendo, sin embargo, su propia estructura. Por último, se contrastó la eficacia de los modelos más utilizados comúnmente aplicándolos sobre los datos simulados.

Notación y símbolos utilizados

Todos los vectores son representados en minúsculas y todas las matrices en MAYUSCULAS apareciendo ambos en negrita. La numeración de las fórmulas empieza con el número [1] y continúa de forma consecutiva. Otros símbolos y acrónimos se detallan a continuación:

A	Matriz de relaciones aditivas
a	Efecto genético aditivo
A_i	Matriz de relaciones aditivas entre sementales
ANOVA	Análisis de varianza
BLUP	Predicor lineal insesgado de mínima varianza
Cte	Constante
D	Matriz de varianzas y covarianzas de las desviaciones debidas al efecto Mendeliano de los genes
Diag()	LA parte entre llaves es la diagonal de una matriz
EM	Algoritmo de maximización de la esperanza
EMC	Algoritmo EM modificado
†	Vector con todas las desviaciones debidas al efecto Mendeliano de los genes
†_i	Desviación debida al efecto Mendeliano de los genes en el individuo i
h²	Heredabilidad
E()	Esperanza matemática de lo que está entre paréntesis
Ê_i	Estimación mínimo cuadrática del grupo de edad i
e	Efecto aleatorio residual

F_{ij}	Factor de ajuste para la edad i y mes de parto j
I	Matriz identidad
log	Logaritmo
\hat{M}_j	Estimación mínimo cuadrática del mes de parto j
MINQUE	Estimación cuadrática incesgada de norma mínima
MIVQUE	Estimación cuadrática incesgada de varianza mínima
ML	Método de máxima verosimilitud
MME	Ecuaciones del modelo mixto
μ	Media de la población
P	Matriz que define quienes son los padres de cada individuo
p	Efecto aleatorio ambiental permanente
r	Repetibilidad
RA	Efecto fijo combinado del rebaño y el año
REML	Método de máxima verosimilitud restringida
σ^2	Varianza
σ_a^2, σ_u^2	Varianza genética aditiva
σ_e^2	Varianza residual
σ_p^2	Varianza ambiental permanente
σ_s^2	Varianza entre sementales
σ_{ui}^2	Varianza genética aditiva asociada al individuo i
$\sigma_{ui \ uj}$	Covarianza genética aditiva entre los individuos i y j
σ_v^2	Varianza entre vacas
UCP	Unidad Central de Proceso
u	Vector de los valores genéticos aditivos de todos los individuos

- u, Valor genético aditivo del individuo i
- Var Varianza
- x, Valor i extraído aleatoriamente de una distribución normal de media cero y varianza uno.
- ϕ Función de verosimilitud
- |R| Determinante de una matriz. En este caso, de R
- σ Macho
- " Matriz de rango completo (X^a es la submatriz de rango completo de X
- i) Elemento genético de una matriz
- ' Transpuesta de una matriz (p. e. M' es la transpuesta M)
- " Estimador o predictor de un efecto (p. e. \hat{u} es el predictor de u)

II. REVISION BIBLIOGRAFICA

1.- Consanguinidad

Cuando los apareamientos entre animales en una población no se realizan completamente al azar, o cuando aquellos se producen en una población finita, aparece en ella el efecto de la endogamia o consanguinidad. En estas condiciones, los individuos que se cruzan, poseen, en ocasiones, uno o más antecesores comunes, y, en consecuencia, ambos individuos comparten genes que provienen de antepasados comunes. El resultado inmediato es el aumento de la homocigosis en la población. De este modo, genes recesivos, hasta entonces ocultos por segregarse a baja frecuencia en la población, pueden salir a la luz. Normalmente son genes de efectos negativos, que pueden producir una depresión en el tamaño, fertilidad, vigor, salud, y en consecuencia, en el rendimiento.

El cálculo de los coeficientes de consanguinidad en poblaciones de animales domésticos puede interesar, por tanto, para evitar su aumento, pero también para valorar de forma más precisa a los reproductores, ya que su valor debe ser incluido en la matriz de varianzas y covarianzas genéticas en las ecuaciones del modelo mixto.

Dichos coeficientes han sido calculados durante mucho

tiempo mediante la detección de antepasados comunes por comparación de líneas ancestrales (Wright, 1922; Hazel y Lush, 1950; Abe y Nishida, 1971; Schaap y Cohen, 1977) o por detección de antepasados comunes por medio de algoritmos de búsqueda de caminos de parentesco (Rehfeld y col., 1967; Alfonso, 1971; Stevens, 1975; Nadot, 1971; Nadot y Vaysseix, 1973; Alfonso, 1985).

Quass (1976) utiliza un método iterativo de búsqueda que permite calcular los coeficientes de endogamia al tiempo que construye la inversa de la matriz de relaciones aditivas para ser incluida en las ecuaciones del modelo mixto, mediante la búsqueda de los padres de cada individuo repetidas veces. Una modificación de este algoritmo, en el que la búsqueda es de descendientes en lugar de la búsqueda de los padres, fue desarrollado por Gutiérrez y col. (1988). Este procedimiento podría encuadrarse dentro del grupo 3 de la clasificación realizada por Boyce (1983) y permite reducir los tiempos de cálculo.

2.- Componentes de varianza

2.1.- Evolución de los métodos de estimación

Un procedimiento muy sencillo ampliamente utilizado para estimar heredabilidades y repetibilidades en ganado vacuno lechero ha sido el cálculo del coeficiente de regresión de las producciones

de las vacas sobre la de sus madres o sobre la de sus abuelas (Bereskin y Freeman, 1965, Wade y Van Vleck, 1989). La estimación de parámetros genéticos mediante el coeficiente de regresión es un método muy sencillo de cálculo pero restringido a ciertas bases de datos, ya que éstas deben ajustarse a una estructura muy particular. Por otro lado, es un método poco preciso, al utilizar el parecido entre parientes de distintas generaciones como única información. También es un método sesgado al no poder separar fuentes de covarianza ambientales entre distintas generaciones (Falconer, 1981).

El método tradicionalmente más utilizado ha venido siendo el ANOVA (análisis de varianza) (Barker y Robertson, 1966, Van Vleck y Henderson, 1961). Este procedimiento es también muy limitado, ya que sólo es válido para datos equilibrados, aunque modelos muy sencillos permiten el uso de bases de datos con desequilibrio. Cuando la población ha sido sometida a selección es un método sesgado. Sin embargo, las estimaciones que produce son únicas y su cálculo es muy simple, ya que basta igualar cuadrados medios a sus esperanzas, siendo ambos muy sencillos de obtener.

Henderson (1953) extendió las posibilidades del ANOVA a casos desequilibrados con sus métodos I, II y III.

El método I, como todos los métodos aparecidos desde esta fecha, posee la propiedad de ser invariante a la traslación, es

decir, que las estimaciones de los componentes de varianza que produce, no dependen de los valores del vector de efectos fijos. Este método sólo es válido para modelos completamente aleatorios, donde el único efecto fijo es la media general.

El método II ajusta por mínimos cuadrados la variable de trabajo a los efectos fijos, y aplica el método I a la nueva variable ajustada para dichos efectos. Para realizar el primer paso es necesario obtener una inversa generalizada de la matriz de coeficientes. Debido a ello, Searle (1968) mantuvo que el método no estaba bien definido por no producir estimaciones únicas, al depender del vector de soluciones obtenido para los efectos fijos. Sin embargo, se demostró que esta aseveración era incorrecta, ya que se obtenían soluciones únicas siempre que se siguiesen exactamente las directrices de Henderson (Henderson y col., 1974). El método II no puede ser utilizado, por ejemplo, para modelos con interacciones entre efectos fijos y aleatorios (Searle, 1968).

El método III es mucho más complejo pero permite el uso de un gran número de modelos. Consiste en calcular varias formas cuadráticas reducidas e igualarlas a sus esperanzas. Debido a las distintas posibilidades de formas cuadráticas, las soluciones pueden no ser únicas. En este método, el coste de cálculo es más elevado que en los anteriores, aunque no excesivamente, por lo que ha sido incluido en muchos programas de estadística (Harvey, 1960, 1977, 1987). La única propiedad estadística digna de mención de

estos métodos es la falta de sesgo, y aún puede perderse esta propiedad si los datos han sido sometidos a selección, como puede suceder en las poblaciones de animales de interés zootécnico.

La búsqueda de métodos con otras propiedades deseables hizo que LaMotte (1970, 1971) desarrollase el MIVQUE (estimación cuadrática incesgada de varianza mínima) y Rao (1971) el MINQUE (estimación cuadrática incesgada de norma mínima). Estos estimadores incorporan información de las varianzas a priori y tienen mínima varianza si esa información coincide con el verdadero valor de los parámetros. La única diferencia entre ambos es que MINQUE puede ser utilizado sin la necesidad de asumir normalidad de la variable de trabajo. El cálculo de estas estimaciones resulta más bien complicado debido a la necesidad de invertir grandes matrices repetidamente, a pesar de que las formas cuadráticas pueden ser escritas en la forma de las ecuaciones del modelo mixto (Henderson, 1986). Por otra parte, las estimaciones obtenidas por estos métodos no son únicas debido a que distintos valores a priori llevan a distintos resultados. Sin embargo, son métodos incesgados aunque se analicen datos sometidos a selección.

Henderson (1980a, 1980b) presentó un método incesgado e invariante a la traslación que pretendía disminuir las necesidades de cálculo que precisaba su método III. Es conocido generalmente como método IV de Henderson.

Dempfle y col. (1981) compararon sobre una misma base real de datos, los métodos III y IV de Henderson y el procedimiento MINQUE, concluyendo que este último es el más eficiente aunque el método IV de Henderson estimó sorprendentemente mejor la varianza del efecto macho. Este procedimiento aparecía como una útil alternativa al MINQUE.

Hartley y Rao (1967) describieron la aplicación del método de máxima verosimilitud (ML) a la estimación de componentes de varianza. Es un método ampliamente usado en estadística, que proporciona estimaciones únicas al no depender de los valores β iniciales, que además se encuentran siempre dentro del espacio paramétrico.

Patterson y Thompson (1971) describieron formalmente el método de máxima verosimilitud restringida (REML). Consiste en maximizar un conjunto de contrastes de los datos que son independientes de los efectos fijos. Este método, equivalente a la utilización iterativa del MINQUE, tiene en cuenta, por su definición, la pérdida de grados de libertad que se produce por el hecho de estimar los efectos fijos (Patterson y Thompson, 1971), de manera que corrige el sesgo que se producía cuando se utilizaba el ML, cuyo criterio de optimización era la maximización de la verosimilitud del vector de datos. Además utiliza toda la información que producen los datos respecto a los parámetros a estimar (Marville y Callanan, 1990) y las estimaciones no se ven

afectadas por la selección ni la posible diversidad de cruzamientos realizados siempre que se incorpore toda la información existente a priori (Gianola y Fernando, 1986). Es inconveniente que presenta, en comparación a ésta, es que aumenta la dificultad de cálculo.

Gianola y Fernando (1986), han demostrado, mediante una aproximación de tipo Bayesiano, que se obtienen predictores óptimos de variables aleatorias con BLUP (best lineal unbiased predictor), cuando se incorporan en el análisis las estimaciones de componentes de varianza obtenidas utilizando el procedimiento REML, cuando no se tiene completo conocimiento de los parámetros, y siempre que la variable de interés tenga una distribución normal multivariante.

Lin y McAllister (1984) compararon el método III de Henderson, el ML, el MINQUE y el REML, evidenciando que las estimaciones ML son sesgadas en relación a las estimaciones REML, tanto si los datos han sido sometidos a selección como si no, ya que el ML no tiene en cuenta la pérdida de grados de libertad que se produce al estimar los efectos fijos, lo que lleva a la infravaloración de la varianza residual y, en consecuencia, a la sobreestimación de la heredabilidad. El método III de Henderson resultó también sesgado en este sentido al tratar como efecto fijo el efecto macho. El REML produjo estimaciones insesgadas siendo además el más rápido en converger. Por otro lado, se sabe que el REML produce los mismos resultados que ANOVA con datos equilibrados (Corbell y Searle, 1976; Anderson y col., 1984), y son bien

conocidas las propiedades óptimas de MREVA en estas circunstancias. Por estas razones, el método REML se ha convertido, en los últimos años, en el único método aceptado para realizar estimaciones de componentes de varianzas en el ámbito de la producción animal, hasta el punto de generalizarse como el único método usado en la práctica, en detrimento de los otros citados aquí.

1.2.- Búsqueda de superación de las dificultades de cálculo

La estimación de componentes de varianzas mediante el procedimiento REML resulta excesivamente costosa cuando se dispone de un gran volumen de datos y el modelo es complejo, debido a que es necesaria la inversión o diagonalización de la matriz de coeficientes repetidas veces. Esto despertó la preocupación de los investigadores en busca de aproximaciones y algoritmos que hicieran más sencillo su desarrollo práctico.

1.2.1.- Aproximaciones metodológicas

Van Esden y Freeman (1965) desarrollaron un método para obtener los límites superior e inferior de las formas cuadráticas necesarias a partir de sus esperanzas. Es un procedimiento útil ya que, normalmente, se conoce bien en qué grado se aproximan las esperanzas de las formas cuadráticas a su valor real, pero este método no funciona bien cuando la matriz de varianzas y covarianzas

de los efectos aleatorios no tiene una estructura diagonal como sucede en el caso de la matriz de relaciones aditivas.

Una estrategia más utilizada, formalizada por primera vez por Harville (1977), es buscar formas cuadráticas similares a las utilizadas por el REML cuyas esperanzas sean más sencillas de calcular, pudiéndose utilizar de forma iterativa para obtener una buena aproximación de las estimaciones REML. La bibliografía es densa en este sentido. Henderson (1980a, 1980b), en su método IV, desarrolló formas cuadráticas y esperanzas que podrían ser usadas en un esquema iterativo para el REML. Cunningham y Henderson (1968) presentaron una estrategia iterativa como modificación al Henderson III. Thompson (1969) simplificó y corrigió este procedimiento. Este método asume que los valores a priori son los verdaderos valores de las varianzas, obteniéndose las mismas estimaciones finales que si se hubiesen incluido las auténticas esperanzas (Dempster y col., 1977). Schaeffer (1986) hizo una nueva derivación del método extendiéndolo al caso de múltiples varianzas, y lo llamó método de las pseudoesperanzas. Van Raden y Jung (1988) extendieron el método de las pseudoesperanzas siguiendo los pasos desarrollados por Henderson (1980a, 1980b) en su método IV, proponiendo además la inserción de la matriz de relaciones aditivas dentro de las formas cuadráticas.

Smith y Lin (1989) redujeron los tiempos de cálculo de los algoritmos más utilizados calculando los autovectores

implícitamente y realizando transformaciones ortogonales en la diagonalización de la matriz de coeficientes.

Harville y Calanan (1980) mejoraron también los algoritmos más conocidos mediante diversos procesos de diagonalización, tridiagonalización, absorción, linealización y reparametrización.

Mixtal y Du (1990) proponen reducir los cálculos al no invertir la parte de las ecuaciones correspondiente a animales sin registro por intervenir en la traza de forma constante.

1.1.2.- Modelos

Existe una relación inversa entre la información incorporada en cada modelo de análisis posible y su coste computacional. Por ello el modelo elegido será siempre un compromiso entre el modelo ideal y las necesidades de cálculo.

El modelo animal, mencionado por Henderson y Quaas (1976), y llevado a la práctica por primera vez en ganado vacuno lechero por Westell (1984), es el que permite incorporar toda la información disponible de parentesco. Sin embargo, suele ser necesario el uso de modelos aproximados debido a sus exigencias de cálculo.

El modelo padre fue aplicado por primera vez por Everett y Henderson (1972). Mediante este modelo se asume que cada registro de rendimientos es debido a la mitad del valor genético aditivo del padre del animal que lo produjo. Al ser muy inferior el número de sementales que el número de animales, el coste computacional se reduce considerablemente.

Debido a la poca información que utiliza el modelo padre, Henderson (1973) sugiere que se introduzca la matriz de relaciones aditivas entre sementales.

Dado que, como ya se ha dicho, el número de machos es relativamente pequeño, otra aproximación más cercana al modelo animal sería el incluir también el abuelo materno. Este modelo fue utilizado por primera vez por Everett y Quaas (1979), y su desarrollo teórico está descrito por Quaas y col. (1979).

El modelo semental con la vaca jerarquizada a él es otra aproximación al modelo animal que ignora la hembra con la que se cruzó el semental y todas las relaciones existentes a través de las hembras. Henderson (1985) definió como modelos lineales equivalentes aquellos que tenían idénticos momentos de primer y segundo orden del vector de datos. Se puede demostrar fácilmente que este modelo es equivalente a un modelo animal en el que se ignoran las hembras con las que se cruzan los sementales y todas las relaciones que existen a través de las hembras.

Como hemos visto anteriormente, el método REML ha sido el método de elección para la estimación de componentes de variancia, también en ganado vacuno lechero, normalmente bajo un modelo padre. De hecho, en nuestro país (Cebán y col., 1989, Carabaño y col., 1989), y fuera de él (Mill y col., 1983, De Vser y Van Vleck, 1987, Jager, 1987, Beaumont, 1989, Van der Werf y De Boer, 1989a, Boldman y Freeman, 1990, O'Ferrall, 1990, Short y col., 1990, Chauhan y Hayes, 1991), casi todas las estimaciones de componentes de variancia mediante el REML en ganado vacuno lechero han sido realizadas ajustando este modelo. En él se asume que los registros de la progenie son debidos a la mitad del valor de mejora de los padres, ignorando así la información aportada por la madre. El interés por el modelo animal, conceptualmente más sencillo, ha aumentado recientemente. Mediante este modelo cada registro permite obtener información directa del genotipo del animal sobre el que ha sido hecha la medición, obteniéndose la estimación de su valor de mejora así como los de otros animales con los que está emparentado aunque éstos no tengan registros de producción.

La utilización del REML ajustando un modelo animal, cuando se dispone de gran número de datos, necesitaba gran cantidad de memoria de ordenador al tiempo que consumía excesiva cantidad de tiempo de computación, lo que en muchos casos hacía prohibitivos los análisis. Una aproximación que permitió resolver el problema consistía en ignorar las relaciones que existían entre los animales cuando pertenecían a distintos rebaños (Swalve y Van Vleck, 1987,

Dong y Van Vleck, 1989). De este modo, tanto las formas cuadráticas como sus esperanzas, se iban calculando dentro de cada rebaño, después de obtener las soluciones de las ecuaciones del modelo mixto para cada uno de ellos.

2.2.3.- Algoritmos

Otra posibilidad para reducir las necesidades de cálculo del método REML es desarrollar nuevos algoritmos que simplifiquen la localización del máximo de la función de verosimilitud.

Aunque existen multitud de algoritmos diferentes, son pocos los que han venido siendo utilizados para resolver el problema anteriormente citado.

El algoritmo de Fisher-Scoring (Meyer, 1986), originado a partir del de Newton-Raphson, determina la dirección de búsqueda a partir de las derivadas primera y segunda de la función de verosimilitud. Debido a la sencillez de cálculo de las esperanzas de la segunda derivada, utiliza aquellas en lugar de su valor real.

El algoritmo EM (Expectation-Maximization) fue desarrollado por Dempster y col. (1977). Utiliza funciones de los parámetros a estimar, cuyas esperanzas son más sencillas que las de la función verdadera. Realiza la búsqueda del máximo de la función

en dos pasos. En el primero de ellos se calculan las esperanzas de las formas cuadráticas correspondientes y en el segundo se maximiza la función utilizando la primera derivada. Es más lento de converger que otros algoritmos pero se simplifican los cálculos al no ser necesario calcular la segunda derivada de la función.

Si el hecho de no necesitar calcular la segunda derivada de la función supuso un gran avance, éste fue aún mayor con el desarrollo de algoritmos que tampoco necesitaban calcular la primera derivada. El método Simplex es el mejor ejemplo de los métodos libres de derivadas. Fue desarrollado originalmente por Spendley y col. (1962) y modificado por Nelder y Mead (1965) para que pudiera adaptarse a distintas tablas y formas de la superficie buscada. Se clasifica entre los métodos de búsqueda directa por hacer comparaciones entre valores sucesivos sin recurrir a métodos estadísticos. Meyer (1989) probó tres aproximaciones al máximo de la función de verosimilitud concluyendo que el método Simplex resultaba el más robusto y fácil de usar cuando se intentaban estimar varios parámetros a la vez.

Meyer (1988a, 1988b) escribió un programa en lenguaje FORTRAN 77 para estimar componentes de variancia mediante un algoritmo NEMZ libre de derivadas con un modelo animal y la posibilidad de incluir más de un efecto aleatorio además del residual. Este programa ha tenido amplia difusión a pesar de resultar excesivamente lento cuando la población consta de un

elevado número de individuos. Boldman y Van Vleck (1991) modificaron este programa logrando una velocidad de cálculo aproximadamente 600 veces superior. Esta modificación consistió en sustituir el paso de la absorción por una reordenación de las ecuaciones, descomposición de tipo Cholesky, y la utilización de un paquete informático en lenguaje FORTRAN77 que aprovecha las propiedades de las matrices ralas.

2.3.- Comparación de métodos

Uno de los problemas de mayor impacto en mejora genética animal es la dificultad para comparar distintos métodos. El problema fundamental estriba en el desconocimiento de los verdaderos valores genéticos, y, en consecuencia, no hay forma de hacer comparaciones entre las predicciones o estimaciones obtenidas y los valores auténticos, para ver cuál de los métodos es más preciso.

Henderson (1975) sugiere tres procedimientos posibles para comparar métodos alternativos de análisis. En primer lugar, los métodos analíticos, que determinan las propiedades de cada uno de los procedimientos posibles según los parámetros asumidos en el modelo y, según se trate, el diseño del experimento o la estructura de la población. Sería deseable que las propiedades examinadas fueran determinantes de la calidad del método, a la vez que

sencillas de calcular, como por ejemplo, la minimización del error de predicción. Por desgracia, los métodos analíticos conllevan normalmente manipulaciones matemáticas extremadamente complejas. Una segunda posibilidad es probar los distintos procedimientos sobre datos simulados cuyos verdaderos valores son conocidos (Lin y McAllister, 1984, Van Der Werf y De Boer, 1989b). El problema que presenta esta opción es, que en la mayoría de los problemas prácticos, el elevado número de diferentes estructuras de los datos, así como los diferentes valores que puedan tomar los parámetros, hace que sólo se puedan obtener resultados muy restringidos a las características de la población simulada. Por último, el método más extendido es aplicar los diferentes procedimientos a la misma base de datos real (Dempfle y col., 1983, Van Raden y Jung, 1988). Como ya ha sido comentado, si los resultados son diferentes, nunca podremos concluir cuál de ellos es el mejor. Por otro lado, habiéndose obtenido una conclusión, ésta no puede ser extrapolable a poblaciones con diferente estructura o experimentos de diferente diseño. Además, en cuanto a los métodos analíticos y los basados en la simulación necesitan definir inicialmente un diseño y unos valores para los parámetros, lo que hace imposible prever un resultado en caso de que aquellos fueran diferentes.

3.- Estimaciones anteriores de parámetros genéticos de interés.

El coeficiente de consanguinidad debe ser vigilado atentamente en poblaciones de tamaño pequeño para evitar su aumento, ya que existe una alta probabilidad de cruzar individuos emparentados. Sin embargo, el censo, en poblaciones de ganado vacuno lechero suele ser muy elevado, ya que, en todos los países, la mejora de este tipo de ganado suele estar en manos de organizaciones nacionales. Por este motivo no ha existido una preocupación entre los investigadores por el cálculo del coeficiente de endogamia en este tipo de ganado. No obstante, el uso masivo de algunos sementales podría provocar su aumento. Pogrebnyak (1990) detectó consanguinidades medias en distintas muestras de ganado vacuno lechero en Rusia del orden de 0,039.

Por el contrario, existen en la bibliografía muchas estimaciones de la heredabilidad en ganado vacuno lechero calculadas por diferentes métodos, aunque muy pocas han sido realizadas ajustando un modelo animal a las observaciones (tabla 1).

Tabla 1.- Revisión de estimaciones de heredabilidades en el ganado vacuno frisón.

Autor	Año	País	Método	h ² leche	h ² grasa	h ² % grasa
Van Vleck y Henderson	1961	EEUU	Regresión	0,23	0,17	
Bereskin y Freeman	1965	EEUU	Regresión	0,22	0,18	
Barker y Robertsons	1966	G. Bretaña	ANOVA			
			Toros propios:			
			1 lactación	0,34		0,43
			media de 2 lact.	0,36		0,47
			media de 3 lact.	0,41		0,48
			Toros import:			
			1 lact	0,24		0,44
			media de 2 lact.	0,37		0,41
			media de 3 lact.	0,41		0,43
Mokhtar	1979	Alemania	ANOVA			
			nivel bajo	0,22		
			nivel medio	0,28		
			nivel alto	0,23		
Rothschild y Henderson	1979	EEUU	ML mod. padre			
			1ª lactación	0,41		
			1ª y 2ª lact.	0,35		
Tong y col.	1979	Canadá	ML mod. padre			
			1ª lactación	0,26	0,26	0,35
			1ª y 2ª lact.	0,19	0,17	0,43
			1ª, 2ª y 3ª lact	0,17	0,15	0,51

(continúa)

Autor	Año	País	Método	h ² leche	h ² grasa	h ² t grasa
Danell	1981	Suecia	ANOVA			
			nivel bajo	0,21		
			nivel medio	0,25		
			nivel alto	0,28		
Hill y col.	1983	G.Bretaña	REML mod. padre			
			nivel bajo	0,24	0,29	0,41
			nivel alto	0,30	0,24	0,46
			población	0,25	0,24	0,43
Meyer	1983	G.Bretaña	REML mod. padre			
			1ª lactación	0,33	0,32	0,36
			1ª y 2ª lact.	0,35	0,32	0,44
			1ª, 2ª y 3ª lact	0,28	0,23	0,43
Manfredi y col.	1984	EEUU	REML mod. padre	0,21	0,24	
Meyer	1984	G.Bretaña	REML mod. padre			
			1ª lactación	0,28	0,27	
			1ª y 2ª lact.	0,19	0,21	
			1ª, 2ª y 3ª lact	0,24	0,25	
Lawlor y col.	1984	EEUU	REML mod. padre	0,19	0,26	
Meyer	1985a	Australia	REML mod. padre			
			1ª lactación	0,17	0,13	0,49
			1ª y 2ª lact.	0,15	0,17	0,45
			1ª, 2ª y 3ª lact	0,18		
Meyer	1985b	G.Bretaña	REML mod. padre	0,21	0,16	

Autor	Año	País	Método	h^2 leche	h^2 grasa	h^2 t grasa
Simianer	1986	Alemania	REM. mod. padre 1ª lactación 1ª y 2ª lact. 1ª, 2ª y 3ª lact	0,29 0,21 0,22		
Celaco y col.	1987	Portugal	REM. mod. padre	0,35		0,54
Cue y col.	1987	Canadá	REM. mod. padre	0,36	0,38	0,56
Jager y Kennedy	1987	Canadá	REM. mod. padre	0,28	0,31	0,61
Swalve y Van Vleck	1987	EEUU	REM. mod. animal ignora relaciones a través de reb. 1ª lactación 1ª y 2ª lact. 1ª, 2ª y 3ª lact	0,33 0,33 0,34		
Robinson	1988	EEUU	REM. mod. padre 1ª lactación 1ª y 2ª lact. 1ª, 2ª y 3ª lact	0,32 0,29 0,30	0,34 0,31 0,32	
Van Vleck y Dong	1988	EEUU	REM. mod. animal ignora relaciones a través de reb.	0,34	0,30	

Autor	Año	País	Método	h^2 leche	h^2 grasa	h^2 % grasa
Beaumont	1989	Francia	REML mod. padre			
			1ª lactación	0,27	0,26	0,55
			1ª y 2ª lact.	0,24	0,20	0,44
			1ª, 2ª y 3ª lact	0,27	0,24	0,44
Carabaño y col.	1989	España	REML mod. padre	0,16	0,14	
Carabaño y col.	1989	EEUU	REML mod. padre			
			población 1	0,29	0,28	
			población 2	0,37	0,33	
Cañón y col.	1989	España	REML mod. padre			
			1ª lactación	0,20	0,15	
			todas lactaciones	0,17	0,13	
			nivel bajo	0,21	0,15	
			nivel alto	0,17	0,13	
Gómez y col. (datos no publicados)	1989	España	Henderson III	0,12	0,10	0,11
			REML mod. padre	0,12	0,11	0,13
Dong y Van Vleck	1989	EEUU	REML mod. animal			
			ignora relaciones			
			a través de reb.			
			población 1	0,33		
			población 2	0,28		
Lee y col.	1989	Corea	Henderson III	0,26		
Mainert y col.	1989	Alemania	REML mod. padre	0,28	0,40	0,74

2
5

(continúa)

Autor	Año	País	Método	h ² leche	h ² grasa	h ² t grasa
Menéndez y col.	1989	Cuba	REML mod. padre	0,17	0,16	
			nivel bajo	0,22	0,17	
			nivel medio			
			nivel alto	0,24	0,22	
Wade y Van Wicck	1989	EEUU	Regresión			
			California	0,36	0,34	0,55
			N. York	0,34	0,13	0,64
			Wisconsin	0,38	0,13	0,68
Carabaño y col.	1990	EEUU	REML mod. padre			
			California	0,21	0,21	0,34
			N. York	0,27	0,26	0,48
			Wisconsin	0,23	0,22	0,41
O'Ferrall	1990	Irlanda	REML mod. padre	0,38	0,31	0,43
Shert y col.	1990	EEUU	REML mod. padre			
			nivel bajo	0,15		
			nivel medio	0,26		
			nivel alto	0,23		
Shutz y col.	1990	EEUU	Henderson III			
			1ª lactación	0,20	0,17	0,46
			1ª y 2ª lact.	0,17	0,20	0,50
			1ª, 2ª y 3ª lact	0,12	0,18	0,44
			REML mod. padre			
			1ª lactación	0,16	0,16	0,38
			1ª y 2ª lact.	0,16	0,18	0,45
1ª, 2ª y 3ª lact	0,11	0,17	0,41			

(continúa)

Autor	Año	País	Método	h^2 leche	h^2 grasa	h^2 v grasa
Ben Gara y col.	1991	España	REML mod. padre	0,24	0,18	
Chauhan y Nayes	1991	Canadá	REML mod. padre	0,29	0,31	0,65

III. MATERIAL

1.- DATOS

1.1.- Consanguinidad.

El archivo de pedigrís fue proporcionado por la Asociación Nacional de Fricosa Española (A.N.F.E.), siendo su estructura la siguiente:

La población sobre la que se calcula la consanguinidad está constituida por 22.100 individuos de ambos sexos, de los cuales, se consideran población base, es decir, que no se conocen sus padres, 8.946.

Entre los 816 toros que aparecen en el análisis, sólo de 115 se conocen sus ascendientes.

De las 21.464 hembras existentes, únicamente 12.769 tienen al menos un padre conocido.

El número de generaciones se calculó buscando el antecesor más lejano por cualquiera de las vías, paterna o materna, en cada uno de los individuos. No se pudo estimar el número de generaciones mediante la edad de los animales por carecer un gran

numero de ellos de la fecha de nacimiento.

El número máximo de generaciones existentes en esta población, calculado de la forma descrita, fue de 6, y se distribuyeron de la siguiente manera:

Tabla 2.- Distribución de los individuos según el número de generaciones.

	N° Total	%
Individuos con 1 generación	8.996	40,71
Individuos con 2 generaciones	933	4,22
Individuos con 3 generaciones	8.308	37,59
Individuos con 4 generaciones	1.598	16,28
Individuos con 5 generaciones	263	1,19
Individuos con 6 generaciones	2	0,01

El número medio de generaciones encontradas por la vía más larga, sea ésta la vía materna o la paterna, fue de 3,24443.

1.2.- Estimación de componentes de varianza.

El archivo de datos fue proporcionado por la Dirección General de la Producción Agraria (D.G.P.A.) del Ministerio de Agricultura, Pesca y Alimentación (M.A.P.A.).

El archivo original contenía 97.776 lactaciones, registradas entre los años 1980 y 1988, y correspondientes a vacas de raza Frisona.

Se realizó una estricta depuración de los datos en tres pasos:

a) En el primer paso se intentaron detectar errores comunes, debidos, normalmente, a fallos en el control de rendimientos y de tipo mecanográfico. Para ello se eliminaron aquellos registros:

- que tuvieran el número de identificación de la vaca inferior a 50.000, ya que por debajo de este número existían animales que poseían la misma identificación, por no haber sido llevado un control más riguroso hasta 1980.

- en los que la edad de la vaca fuera inferior a 18 meses o superior a 144.

- con una longitud de lactación mayor de 425 días, ya que, probablemente, eran debidas a la unión de otras dos lactaciones consecutivas.

- que presentaran el año de parto de la vaca anterior a 1980 o posterior a 1988.

- cuyos registros de producción fueran nulos.

Durante esta depuración se redujo el número de registros a 61.570, que corresponde, tan sólo, a un 62% del archivo de datos original.

b) En un segundo paso se eliminaron aquellos registros en los que la genealogía de la vaca fuera desconocida. Solamente 32.610 lactaciones superaron esta condición, es decir, un 53% de los datos que habían superado la criba anterior y un 33,3% del archivo original. Las dos depuraciones anteriores son completamente necesarias antes de pasar a hacer los análisis planteados en este trabajo.

La base de datos correspondiente a esta raza era, probablemente, y en líneas generales, la mejor que poseía nuestro país en aquel momento. A pesar de ello, sólo la tercera parte del archivo original, parece tener utilidad para realizar trabajos de interés genético.

c) En el último paso se eliminaron todos los registros que provocaban desconexión de la información en la población.

Los grupos de contemporáneas se definen como grupos de comparación para evitar sesgos debidos a asociaciones entre animales y ellos mismos. Por ejemplo, si ciertos sementales no son

utilizados en diversos rebados y a lo largo de distintos años, jamás podrá tomarse en cuenta la influencia de estos efectos sobre la evaluación de los machos. La definición de muchos grupos de comparación puede originar incrementos en la varianza del error de predicción al quedar algunos grupos compuestos de pocas observaciones. Una consecuencia directa de esto es la dificultad para conectar toda la información disponible.

Muchos programas de uso extendido, han sido escritos sin tener en cuenta este punto. La ausencia de conexión produce, entre otras cosas, falta de estimabilidad. A pesar de ello, estos programas suelen alcanzar un resultado, que, desde este punto de vista, podría no ser del todo válido.

Para llevar a cabo esta última tarea, se utilizaron tres programas diferentes aplicados de forma secuencial e iterativa, de manera que el primer programa se volvía a aplicar después de acabar el último, repitiéndose la operación hasta que no se produjese pérdida de registros:

- En el primero se exigía que en cada rebado hubiese, al menos, hijas de dos sementales diferentes.
- En el segundo se eliminaban las vacas que no tuviesen, al menos, la primera lactación y se obligaba a que el resto de sus registros fueran lactaciones consecutivas. De este modo se corrige el sesgo

debido a la selección que se produce al eliminar los registros de los animales que producen demasiado poco (Boldman y Freeman, 1990).

- El tercer programa obligaba a cada semental a tener al menos cinco hijas.

El número de lactaciones que resultó fue de 5.496, lo que supone, tan sólo, un 5,6% de las existentes en el archivo original. El número de registros que se obtienen, cuando sólo se exigen dos hijas a cada macho, es de 11.636, es decir, el 11,9% de las presentes en el archivo de partida. En cualquier caso, el archivo de datos resultante, parece ser excesivamente pequeño si se compara con el que se introdujo al principio. La eliminación de un elevado número de registros en esta última tarea muestra una enorme desconexión, debida, en parte, a la eliminación de registros durante la depuración efectuada en los pasos anteriores. No obstante, resulta imprescindible realizar esta criba, más aún, después de haber observado el resultado obtenido.

Existen técnicas que permiten ajustar algunos factores cuando no van a ser definidos en los modelos de análisis (Lush y Shrode, 1950). En nuestro caso se calcularon factores de ajuste multiplicativos para la edad y mes de parto de la vaca, los cuales habían sido calculados por Cañón (1989) mediante el siguiente cociente:

$$F_{ij} = (\hat{A}_i + \hat{A}_j) / (\hat{E}_i + \hat{E}_j)$$

siendo:

F_{ij} , el factor de ajuste para la edad i y mes de parto j

\hat{A}_i , la estimación mínimo cuadrática del grupo de edad i

\hat{A}_j , la estimación mínimo cuadrática del mes de parto j

\hat{A} , y \hat{A} , las estimaciones mínimo cuadráticas del grupo de edad y mes de parto de referencia respectivamente.

El archivo de datos resultante fue reenumerado completamente para que los animales tuvieran un número consecutivo, procurando seguir las indicaciones sugeridas por Meyer (1989) para reducir el coste computacional durante la absorción, manteniendo los animales emparentados lo más cerca posible unos de otros. Se mejoró, asimismo, el tiempo de cálculo, cuando se ordenan los rebases de forma que los que tienen menor número de datos sean los primeros en ser absorbidos durante el proceso de eliminación Gaussiana. Esto mismo se tuvo en cuenta con los otros efectos del modelo.

Existen animales en la genealogía que sólo aparecen una vez, de manera que no aportan ningún tipo de información, ocupando memoria de forma innecesaria, al tiempo que pueden provocar dificultades durante el proceso de estimación. Estos animales, normalmente madres de vacas con registro, fueron ignorados

considerando a sus descendientes como animales de los que se desconoce uno de sus padres.

El archivo definitivo presentó un gran desequilibrio como queda reflejado en las figuras 1, 2, 3, 4, 5 y 6. En ellas se puede observar, por ejemplo, que más de la mitad de los rebaños poseen cinco o menos vacas (figura 1) y menos del 5% presentan un número

DISTRIBUCIÓN DE LOS REBAÑOS SEGUN EL NUMERO DE VACAS

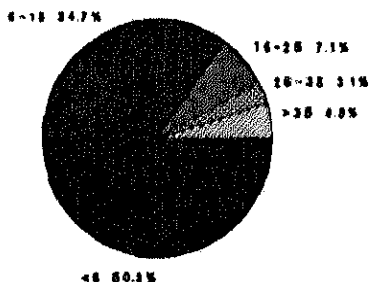


Figura 1.- Distribución del número de vacas por rebaños

superior a 15. El 90% de ellos, al mismo tiempo, poseen menos de ocho registros de producción (Figura 2).

De forma similar se presenta el desequilibrio en los toros, ya que el 66,5% tienen entre cinco y diez hijas y sólo un 1% tiene más de 50 hijas (Figura 3). Los mismos resultados aparecen

**DISTRIBUCIÓN DE LOS REBAÑOS SEGUN
EL NÚMERO DE REGISTROS**

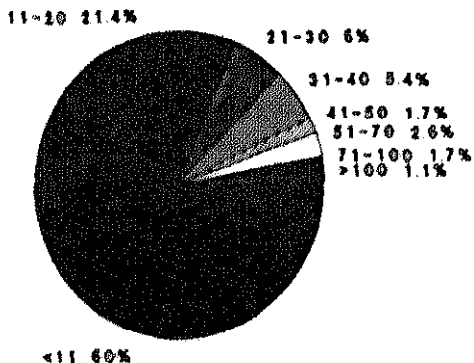


Figura 2.- Distribución del número de registros por rebaños

**Distribución de los toros según
el número de hijas**

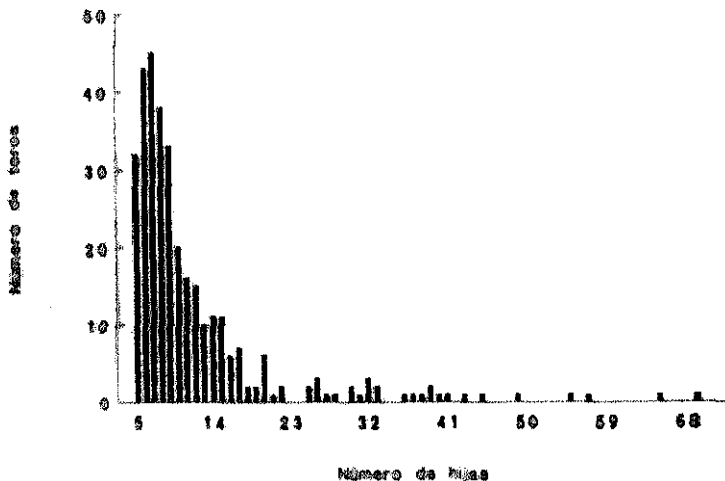


Figura 3.- Distribución de los toros según el número de hijas

para las lactaciones, ya que el 76% de los sementales se corresponden con menos de 20 registros de sus hijas mientras que con más de 100 sólo se observan un 0,6% de ellos (Figura 4).

El hecho de que madres fueran eliminadas por no tener más que una hija provocó que el 85% de las vacas y el 86% de los registros presentasen solamente el padre en su información genealógica:

Tabla 3.- Distribución de las vacas según el número de padres conocidos.

Vacas con un sólo padre conocido	3157	85%
Vacas con ambos padres conocidos	568	15%
Registros con un sólo padre conocido	4624	84%
Registros con ambos padres conocidos	872	16%

La restricción impuesta referente a las lactaciones en la primera depuración, causó asimismo un fuerte desequilibrio en el número de lactaciones registradas de cada vaca, apareciendo muchas vacas que sólo poseen una lactación.

En resumen, la muestra resultante se presenta completamente desequilibrada y, por ello, parece necesario realizar un estudio previo de los métodos a seguir para estimar parámetros genéticos en ella, antes de introducir ciegamente nuevas técnicas que podrían elevar el coste de computación innecesariamente.

**DISTRIBUCION DE LOS TOROS SEGUN EL
NUMERO DE REGISTROS**

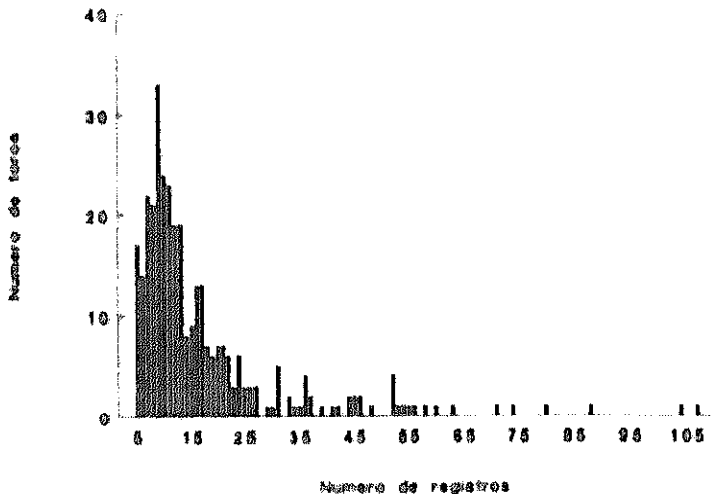


Figura 4.- Distribución de los toros según el número de lactaciones

**Distribución del número de vacas
según el número de lactaciones**

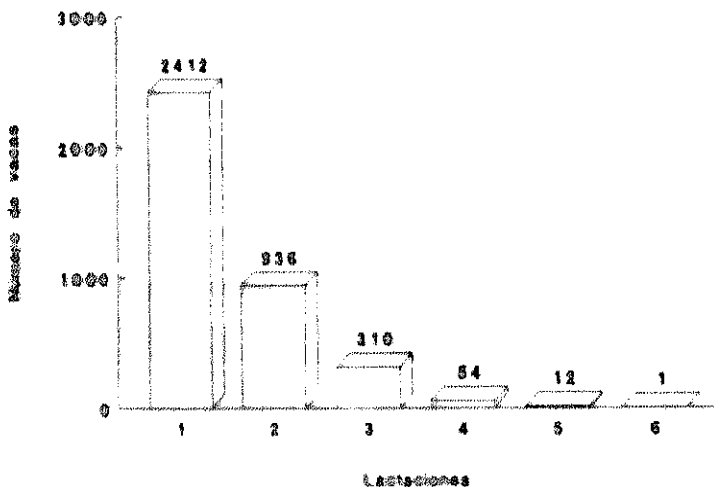


Figura 5.- Distribución de las vacas según el número de lactaciones

Distribución del número de años representados en cada rebaño

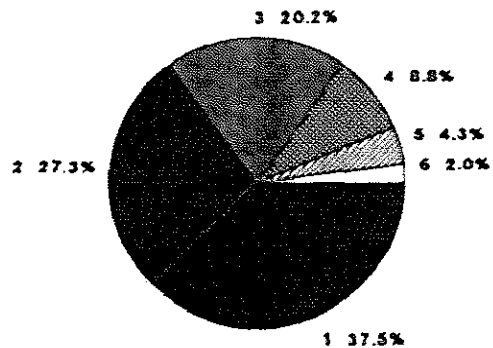


Figura 6.- Distribución de los rebaños representados en cada rebaño

IV. METODOLOGIA

1.- Coancestrinidad

El método que se utilizó en este trabajo fue el desarrollado por Gutiérrez y col. (1988).

Si definimos u como el vector de valores genéticos aditivos de cada uno de los individuos de una población, es evidente que:

$$u_i = u_j + u_k + \delta_i \quad [1]$$

siendo:

j y k los padres de i

u_j , u_k y u_i los valores aditivos de los individuos j , k y i

δ_i la desviación debida al efecto Mendeliano de los genes en el individuo i

Si por otra parte, llamamos P a la matriz de padres, definida como aquella en la que aparece un uno en la columna donde se encuentra cada padre del individuo i (en la fila i), podemos generalizar la expresión [1] de la forma siguiente:

$$u = \frac{1}{2} P u + \epsilon \quad (2)$$

o:

$$u = (I - \frac{1}{2} P)^{-1} \epsilon$$

La variancia de u resulta ser igual a:

$$\text{Var}(u) = \lambda \sigma_u^2$$

siendo: σ_u^2 la variancia genética aditiva entre los individuos suponiendo a sus padres sin relación de parentesco entre ellos.

Por tanto:

$$\lambda = (I - \frac{1}{2} P)^{-1} \frac{\text{Var } \epsilon}{\sigma_u^2} (I - \frac{1}{2} P)^{-1} \quad (3)$$

Si suponemos que las desviaciones ϵ_i son independientes, se puede definir D como una matriz diagonal:

$$D = \left(\frac{\text{Var } \epsilon}{\sigma_u^2} \right)$$

Y llamando a T:

$$T = (I - \frac{1}{2} P)^{-1}$$

La expresión [3] se puede escribir de la forma:

$$A = T D^2 D^2 T'$$

Si a $T D^2$ lo llamamos L la expresión [3] queda:

$$A = L L' \quad [4]$$

La inversa de la matriz A se puede expresar, por tanto, de la siguiente forma:

$$A^{-1} = (I - \frac{1}{2} P') D^{-2} (I - \frac{1}{2} P) \quad [5]$$

La única dificultad que puede aparecer al intentar construir la matriz A^{-1} es el cálculo de los valores d_{ii} , para construir la matriz diagonal D

Tomando varianzas en [1] y dividiendo por σ_u^2 se obtiene:

$$\frac{\sigma_{y_i}^2}{\sigma_u^2} = \frac{1}{2} \frac{\sigma_{y_i}^2}{\sigma_u^2} + \frac{1}{2} \frac{\sigma_{u_i}^2}{\sigma_u^2} + \frac{1}{2} \frac{\sigma_{y_i}^2}{\sigma_u^2} + d_i \quad [6]$$

Sabiendo que:

$$a_{ii} = 1 + F_i = \frac{\sigma_{v_i}^2}{\sigma_v^2}$$

$$F_i = \frac{1}{2} a_{jk} = \frac{\sigma_{v_i}^2}{\sigma_v^2}$$

Sustituyendo en [6] se obtiene:

$$a_{ii} = 1 + F_i = \frac{1}{2} a_{jj} + \frac{1}{2} a_{kk} + \frac{1}{2} a_{jk} + d_i$$

$$d_i = 1 - \frac{1}{2} a_{jj} - \frac{1}{2} a_{kk}$$

Por tanto, conociendo a_{jj} y a_{kk} , tendremos todo lo necesario para calcular d_i . Para obtener el valor de a_{jj} y a_{kk} debemos conocer d_j y d_k que a su vez necesitan ser calculados. De este modo y ascendiendo en la genealogía llegará un momento en que se desconozcan los padres, en cuyo caso se asume que son población base, y por tanto a_{jj} y a_{kk} son igual a cero.

1.1.- Estrategia de cálculo

El archivo de datos debe estar organizado con los individuos numerados secuencialmente e incluidos en una matriz de dos filas, correspondiendo la primera al número del padre y la segunda al de la madre, y el número de la columna al del individuo.

El algoritmo de cálculo que se presenta, puede ser dividido en dos partes. En la primera se calcula la matriz diagonal D para, en la segunda parte, calcular la matriz A^{-1} mediante las semillas regias de Henderson (Henderson y Quass, 1976).

Si comenzamos por el primer animal (ordenados por fecha de nacimiento), para calcular a_{11} , recurrimos a la siguiente expresión, deducida fácilmente de [4]:

$$a_{11} = k_{p-1} \cdot l_p^2 \quad [7]$$

por tratarse del primer individuo se calcula sencillamente por:

$$a_{11} = l_p^2$$

y como sabemos que:

$$L = (I - \lambda P)^{-1} D^{\lambda}$$

$$d_{11} = 1 - \lambda a_{11} - \lambda a_{1k}$$

entonces,

$$d_{11} = 1 \quad \forall$$

$$a_{11} = 1$$

Conocido el valor del primer animal, podemos rellenar la primera columna de la matriz triangular inferior L en un vector de trabajo de la siguiente manera: buscamos el primer hijo según el orden de nacimiento. Cuando lo encontramos le asignamos a l_{1m} (siendo m el número de identificación del hijo encontrado) el valor de a_{11} y pasamos a buscar un hijo del individuo m, para asignarle el valor a_{11} , y así hasta llegar al último individuo de la población para entonces retroceder nuevamente una generación y buscar un hermano del último individuo que se buscó. Así se continúa el proceso hasta acabar con el último hijo del individuo primero.

Quass (1976), calcula este vector, animal por animal, mediante la expresión:

$$l_{1n} = \frac{1}{2} l_{1n} + \frac{1}{2} l_{1n}$$

para todos los individuos de la población, lo que supone un tiempo de cálculo muy superior al que resulta con el algoritmo utilizado.

Para comenzar nuevamente con otro individuo primero se eleva cada uno de los elementos del vector de trabajo ya mencionado al cuadrado para acumular el resultado en un segundo vector de trabajo con el fin de ir calculando la expresión [7].

Una vez calculados todos los elementos de la diagonal de la matriz B se obtiene el inverso de cada uno de ellos, y pasamos a la segunda parte de la subrutina en la cual se calcula la inversa de la matriz A con ayuda de las conocidas reglas de Henderson:

- En el cruce del individuo consigo mismo se incluye d_i .
- En el cruce del individuo con cada uno de sus padres se añade $(-1/2)d_i$.
- En los lugares donde se cruzan los padres $(1/4)d_i$.

De esta modo realizamos los cálculos pertinentes con cada uno de los individuos y la matriz A^{-1} queda construida.

La subrutina devuelve únicamente los elementos de A^{-1} que son distintos de cero, con su valor, y, la fila y columna que ocupan.

Para poder realizar una búsqueda más rápida se puede incluir otro vector que codifica el sexo del individuo.

Conviene resaltar la importancia que tiene el hecho de construir la matriz A^{-1} ya que el hecho de asumir ausencia de consanguinidad puede afectar de forma importante a la clasificación de los reproductores por su mérito genético.

2.- Estimación de componentes de varianzas

Se utilizaron modelos lineales en todos los casos, cuya expresión más general, en forma matricial, es la siguiente:

$$y = Kb + Zu + e$$

siendo:

y el vector de observaciones

K la matriz de incidencia de los efectos fijos

b el vector de efectos fijos

Z la matriz de incidencia de los efectos aleatorios

u el vector de efectos aleatorios

e el vector de términos residuales

Se considera que ambos vectores de efectos aleatorios tienen una distribución normal conjunta.

Las esperanzas de los distintos efectos, así como las varianzas y covarianzas de los mismos, son descritas al especificar cada uno de los modelos de análisis por separado.

2.1.- Modelos

En este trabajo se compararon tres de los modelos más utilizados:

- Modelo animal
- Modelo padre ignorando las relaciones de parentesco existentes entre los individuos.
- Modelo padre incluyendo las relaciones de parentesco existentes entre los machos.

2.1.1.- Modelo animal

Los componentes del modelo animal ajustado a los datos son: una media general fija (μ), un efecto fijo rebaño-año (RA), un efecto genético aditivo debido al animal (a), un efecto ambiental permanente debido a la lactación (p) y un efecto residual (e), siendo los tres últimos aleatorios:

$$Y_{ijkl} = \mu + RA_i + a_j + p_{jk} + e_{ijkl}$$

con los momentos siguientes:

$$E(Y_{ijkl}) = \mu + RA_i$$

$$E(a_j) = E(p_{jk}) = E(e_{ijkl}) = 0$$

$$\text{Var} \begin{bmatrix} a \\ p \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & 0 & 0 \\ 0 & I\sigma_p^2 & 0 \\ 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

siendo:

$Y_{i,jk}$ el registro de producción l , de la lactación k , de la vaca j en el rebaño-año i

λ el numerador de la matriz de relaciones aditivas entre animales
 σ_a^2 , σ_p^2 y σ_e^2 , las varianzas de los efectos aleatorios genético, ambiental permanente y residual respectivamente

2.1.2.- Modelo padre y vaca jerarquizada

El modelo ajustado incluye una media general fija (μ), un efecto fijo rebaño-año (RA), un efecto aleatorio debido al semental (s), un efecto aleatorio debido a la vaca (v) y un efecto aleatorio residual (e):

$$Y_{i,jk} = \mu + RA_i + s_j + v_{j,k} + e_{i,jk}$$

con los momentos siguientes:

$$E(Y_{i,jk}) = \mu + RA_i$$

$$E(s_j) = E(v_{j,k}) = E(e_{i,jk}) = 0$$

$$\text{Var} \begin{bmatrix} u \\ v \\ e \end{bmatrix} = \begin{bmatrix} A_k \sigma_u^2 & 0 & 0 \\ 0 & I \sigma_v^2 & 0 \\ 0 & 0 & I \sigma_e^2 \end{bmatrix}$$

siendo:

y_{ijk} el registro de producción l , de la vaca k , hija del semental j en el rebando-año i

A_k el generador de la matriz de relaciones aditivas entre sementales

σ_u^2 , σ_v^2 y σ_e^2 las varianzas de los efectos aleatorios semental, vaca y residual respectivamente

Este modelo también se aplicó sin tener en cuenta las relaciones de parentesco existentes entre los machos. De este modo, el modelo difiere del anterior, tan sólo, en la matriz de varianzas y covarianzas de los efectos aleatorios:

$$\text{Var} \begin{bmatrix} u \\ v \\ e \end{bmatrix} = \begin{bmatrix} I \sigma_u^2 & 0 & 0 \\ 0 & I \sigma_v^2 & 0 \\ 0 & 0 & I \sigma_e^2 \end{bmatrix}$$

2.1.2.1.- Modelo padre y vasa jerarquitada, con grupos genéticos

Cuando la genealogía de que se dispone, como es nuestro caso, resulta incompleta, deben crearse grupos genéticos para la información de parentesco que se ignora (Westell y col., 1966). De este modo se corrige el sesgo que se produciría al asumir que todos los animales de ascendencia desconocida procedan de la misma población base, cuando en esta existe cualquier tipo de tendencia genética.

Este modelo fue el que se usó inicialmente para calcular los componentes de varianza que luego habrían de ser usados para simular los datos. El efecto grupo genético no se incluyó después sobre los datos simulados, ya que en este archivo, todos los animales de la población base pertenecían a un mismo origen.

En nuestro caso se crearon cuatro niveles del efecto grupo genético según su año de nacimiento:

Tabla 4.- Machos pertenecientes a cada nivel del efecto grupo genético.

	Grupo genético	Número σ
Toros nacidos antes de 1970	1	252
Toros nacidos entre 1970 y 1972	2	42
Toros nacidos entre 1973 y 1975	3	19
Toros nacidos después de 1975	4	15

Segun esta particion los datos quedan distribuidos de la siguiente manera:

Tabla 5.- Distribucion de las vacas y los registros segun el grupo genético del padre.

Grupo genético	Sementales	Vacas	Lactaciones
1	131	2648	3877
2	42	642	972
3	19	314	483
4	15	121	164

Para calcular los componentes de varianza, necesarios para realizar la simulación, recurrimos a este modelo cuya expresion es la siguiente:

$$Y_{i,j,k} = \mu + BA_i + G_j + B_{jk} + v_{jki} + e_{i,j,k} \quad [8]$$

con momentos:

$$E(Y_{i,j,k}) = \mu + BA_i + G_j$$

$$E(B_{jk}) = E(v_{jki}) = E(e_{i,j,k}) = 0$$

$$\text{Var} \begin{bmatrix} \mu \\ v \\ e \end{bmatrix} = \begin{bmatrix} A_i \sigma_A^2 & 0 & 0 \\ 0 & I \sigma_v^2 & 0 \\ 0 & 0 & I \sigma_e^2 \end{bmatrix}$$

siendo:

$Y_{i,j,k}$ el registro de producción m_i de la vaca i , hija del semental k , perteneciente éste al grupo genético j y localizada aquella en el rebano-año i .

λ_j el numerador de la matriz de relaciones aditivas entre sementales.

σ_s^2 , σ_v^2 y σ_e^2 , las varianzas de los efectos aleatorios semental, vaca y residual respectivamente

2.1.- Procedimientos de estimación

2.1.1.- Modelo animal

Un modelo lineal en general tiene la siguiente expresión:

$$y = Xb + Zu + e \quad [9]$$

siendo:

y el vector de n observaciones

X la matriz de incidencia de los efectos fijos, de tamaño $n \times nf$, con rango por columnas nf'

b el vector de los nf efectos fijos

Z la matriz de incidencia de los efectos aleatorios de tamaño $n \times na$

u el vector de los na efectos aleatorios

• el vector de n términos residuales

con momentos siguientes:

$$E(y) = XB$$

$$E(ue) = E(e) = 0$$

$$\text{Var} \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & R \end{bmatrix}$$

y por tanto:

$$\text{Var}(y) = XB' + R$$

Las ecuaciones del modelo mixto pertenecientes a [9] son entonces (Henderson, 1973):

$$\begin{bmatrix} X'R'X & X'R'z \\ Z'R'X & Z'R'z+0' \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ u \end{bmatrix} = \begin{bmatrix} X'R'y \\ Z'R'y \end{bmatrix} \quad (10)$$

El método REML trabaja sobre la parte de la verosimilitud del vector de datos que es independiente de los efectos fijos, lo cual, produce una pérdida de grados de libertad que es tenida en cuenta por el método (Patterson y Thompson, 1971).

Si el vector de datos y se distribuye normalmente con

media $\mathbf{X}\beta$ y varianza \mathbf{V} , el logaritmo de la verosimilitud es (Harville, 1977):

$$\log \phi = -\frac{1}{2}[\text{Cte} + \log|\mathbf{V}| + \log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + (\mathbf{y} - \mathbf{X}\beta)' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)] \quad [11]$$

siendo:

ϕ la función de verosimilitud que será evaluada

\mathbf{X}' , de orden $n \times n'$, una submatriz de rango completo de \mathbf{X}

Siguiendo las igualdades de matrices descritas por Searle (1979), [11] puede ser escrita como:

$$-\frac{1}{2} \log \phi = \text{Cte} + \log|\mathbf{R}| + \log|\mathbf{G}| + \log|\mathbf{C}'| + \mathbf{y}'\mathbf{P}\mathbf{y} \quad [12]$$

siendo:

\mathbf{C}' la matriz de coeficientes en [10] cambiando \mathbf{X} por \mathbf{X}'

\mathbf{P} la matriz siguiente:

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{X}'\mathbf{V}^{-1} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$$

De todos los términos que aparecen en [12], los dos primeros, exceptuando la parte constante, dependen de la estructura de las matrices \mathbf{R} y \mathbf{G} pero normalmente son fáciles de calcular, como veremos después. Los otros dos pueden ser determinados

mediante la absorción, una a una, de todas las filas de la matriz de coeficientes del modelo mixto (Smith y Graser, 1986) ampliada, es decir, la matriz de coeficientes en [10] aumentada con el vector del lado derecho y una forma cuadrática del vector de datos (Graser y col., 1987):

$$M = \begin{bmatrix} y'R'y & y'R'x & y'R'z \\ x'R'y & x'R'x & x'R'z \\ z'R'y & z'R'x & z'R'z + \sigma^{-1} \end{bmatrix} \quad [13]$$

Si se absorben en esta matriz todas las filas y columnas pertenecientes a los efectos aleatorios obtenemos:

$$\begin{bmatrix} y'v'y & y'v'x \\ x'v'y & x'v'x \end{bmatrix}$$

Haciendo lo propio con las filas y columnas referentes a los efectos fijos se obtiene $y'Py$, suma ponderada de los errores al cuadrado. Efectuando la absorción de una fila y la columna correspondiente cada vez, podemos también determinar de forma simultánea el $\log |C'|$.

Si ahora subdividimos la matriz M de [13] de tamaño k , siendo $k = n_f + n_a + 1$, de la siguiente manera:

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

El determinante de M puede escribirse:

$$|M| = |m_{11}| |M_{11} - m_{12}m_{22}^{-1}m_{21}'|$$

Y su logaritmo:

$$\log |M| = \log m_{11} + \log |M_{11}^*|$$

siendo:

$$M_{11}^* = M_{11} - m_{12} m_{22}^{-1} m_{21}' = (m_{11} - m_{12} m_{22}^{-1} m_{21}) = (m_{11})$$

M_{11}^* es la matriz resultante de la absorción de una fila y su columna respectiva simultáneamente. El uso repetido de esta técnica muestra que el determinante requerido es simplemente la suma del logaritmo de los términos diagonales:

$$\log |C^*| = \sum_{i=1}^n \log m_{ii}^*$$

Este término puede ir calculándose, por tanto, mientras se va realizando la absorción descrita anteriormente. En la matriz de coeficientes, X debe estar sustituido por X^* , o bien, omitir en todos los pasos aquellas filas y columnas cuyos elementos

diagonales sean nulos.

El cálculo de los dos primeros términos de la expresión (11) es sencillo en nuestro caso. La matriz R tiene la siguiente estructura:

$$R = I\sigma^2$$

y por tanto:

$$\log |R| = n \log \sigma^2$$

La matriz Q tiene una estructura ligeramente más complicada. En ella van a intervenir dos efectos aleatorios, el efecto genético aditivo del animal y un segundo efecto, conocido con el nombre de efecto ambiental permanente, al poseer más de un registro muchos de los animales de nuestra población:

$$u' = (a' \ m')$$

$$Q = \text{Diag} \{ I\sigma_a^2 ; I\sigma_m^2 \}$$

siendo:

a el vector que contiene los niveles del segundo efecto aleatorio, cuyo tamaño es na .

De este modo, el término que falta se calcula de la siguiente manera:

$$\log |G| = na (\log \sigma_1^2 + \log \sigma_n^2) + \log |A|$$

El $\log |A|$ es constante y por tanto no es necesario calcularlo para evaluar la función de verosimilitud.

2.2.1.1.- Aspectos computacionales

Se tuvieron en cuenta diversas técnicas para aberrar los tiempos de cálculo.

Para valores iniciales de los otros componentes de varianza, la varianza residual puede ser estimada directamente, en este caso, a partir de la suma de cuadrados del error, de la siguiente manera (Marville, 1977):

$$\hat{\sigma}_e^2 = Y'Py / (n-nf) \quad (14)$$

De este modo, los otros parámetros a estimar ($\hat{\sigma}_1^2$ y $\hat{\sigma}_p^2$) pueden ser reparametrizados en función de $\hat{\sigma}_e^2$. Marville y Callanan (1990) mostraron que una función de estimaciones REML de un grupo de parámetros es también la estimación REML de esta función. Una

elección obvia es expresar los otros dos parámetros como una proporción de σ^2 (Mayer, 1989):

$$k_1 = \sigma_D^2 / \sigma^2$$

$$k_2 = \sigma_A^2 / \sigma^2$$

Mediante esta transformación, el procedimiento de estimación conllevará dos pasos. En primer lugar se procede a maximizar una función de verosimilitud en k_1 y k_2 , obteniéndose \hat{k}_1 y \hat{k}_2 ; posteriormente, y a partir de ellos, se obtiene $\hat{\sigma}_D^2$ de (14). Debido a que el número de parámetros a estimar incrementa considerablemente, tanto el número de iteraciones como el número de evaluaciones de la función de verosimilitud, esta reparametrización disminuye sustancialmente las necesidades de cálculo.

Por otro lado, trabajando con poblaciones de animales domésticos la matriz M de (13) suele ser una matriz muy grande, pero al mismo tiempo, con muchos elementos nulos. Consecuentemente, las técnicas para matrices raras, descritas por Knuth (1973), resultan muy útiles en este tipo de análisis. Mediante estas técnicas, conocidas con el nombre de listas ligadas, la matriz M es almacenada en forma de listas de elementos no nulos, una por cada fila a las que se asocia un vector para situar el primer valor de cada fila. De este modo, sólo los elementos no nulos son guardados en memoria, pudiendo manejarse matrices cuyo orden es de muchos miles. Además, la matriz M es siempre simétrica y, por tanto, sólo

es necesario guardar en memoria algo más de la mitad de la misma. Existen técnicas para trabajar con matrices que posean ambas propiedades descritas aquí (Tier y Smith, 1989).

En la construcción de M , el orden de las ecuaciones afecta de forma importante al proceso de llenado de huecos de la matriz durante el proceso de absorción, de manera que el número de elementos distintos de cero va aumentando. Evidentemente, a efectos de cálculo, se pretende que este número de elementos sea lo más pequeño posible. Una estrategia interesante (Meyer, 1989) es comenzar la absorción por aquellas filas y columnas cuyo número de elementos no nulos fuera de la diagonal sea lo menor posible, por ejemplo, comenzando por los animales más jóvenes (Graser y col., 1987). Para nuestro modelo, con dos efectos aleatorios a estimar además del error, y por la misma razón, resulta práctico agrupar los dos efectos correspondientes a cada animal.

2.2.1.1.1.- Factorización de Cholesky

Bolduan y Van Vleck (1991) lograron una velocidad 600 veces superior sustituyendo el proceso de absorción por una reordenación de las ecuaciones, una descomposición de tipo Cholesky de la matriz M , y la utilización de un paquete informático especial para operar con matrices reales denominado SPANSPAK (George y col., 1980).

La reordenación se hizo utilizando el algoritmo de mínimo grado (Duff y col., 1986) adaptado para tratar matrices simétricas. Este procedimiento normalmente necesita más tiempo para realizar la reordenación pero menos tiempo para hacer la factorización de la matriz de coeficientes que los otros algoritmos disponibles en el SPARSPAK. Este algoritmo es especialmente útil para resolver muchos sistemas de ecuaciones con la misma estructura pero valores numéricos diferentes, como sucede en el caso de las ecuaciones del modelo mixto, que trabaja con la misma matriz de coeficientes pero distintos valores *a priori* de los componentes de varianza en cada iteración. En estas condiciones el sistema sólo necesita ser ordenado una vez.

Si la matriz de coeficientes C^* de [10] es simétrica y definida positiva, puede ser descompuesta utilizando el procedimiento de Cholesky de la siguiente manera:

$$C^* = LL^*$$

siendo L una matriz triangular inferior. El vector de soluciones de las ecuaciones del modelo mixto puede calcularse entonces a partir de dos sistemas triangulares realizando sustituciones alternas en uno y otro.

y^*Py de [11] puede obtenerse de este modo:

$$y'Py = y'Y - a'r = y'y - b'X'y - d'H'y$$

Si siguiendo los pasos desarrollados por Press y col. (1989), el $\log |C'|$ en [12] puede ser calculado fácilmente del siguiente modo:

$$\log |C'| = \sum [\log(l_{ii}^2)] = 2 \sum [\log(l_{ii})]$$

siendo l_{ii} el elemento diagonal i del factor de Cholesky L .

El uso de la factorización de tipo Cholesky, seguido de una reordenación apropiada de las ecuaciones antes de utilizar los programas pertinentes para el uso de matrices esparcidas, producen una drástica reducción en los tiempos de cálculo (Goldman y Van Vleet 1991).

2.2.1.1.2.- Aproximación al máximo de la función. El método Simplex

La elección de la estrategia a utilizar para localizar el máximo de la función de verosimilitud depende de varios factores. Interesa, en primer lugar, que el método realice el menor número posible de evaluaciones de la función de verosimilitud, ya que este paso es, con mucho, el que posee más exigencias de cálculo. En segundo lugar, debería alcanzar el máximo de la función aunque los valores de partida distaran mucho del mismo. Otra propiedad

deseable es que no se vea afectado por problemas de precisión numérica, es decir, que tenga poco error, incluso con funciones muy planas. Por último, la inclusión de restricciones en el espacio paramétrico debe poder hacerse de forma sencilla y sin afectar seriamente al tiempo de cálculo. Mayer (1989) probó tres aproximaciones al máximo de la función de verosimilitud concluyendo que el método Simplex resultaba el más robusto y fácil de usar cuando se intentaba estimar varios parámetros a la vez.

El método Simplex es generalmente el método indicado cuando se pretende minimizar una función discontinua de más de un parámetro, aunque en desarrollo no estaba enfocado a minimizar una función, sino a maximizar una verosimilitud. Suele clasificarse entre los métodos de búsqueda directa por hacer comparaciones entre valores sucesivos sin recurrir a métodos estadísticos. Fue modificado por Nelder y Mead (1965) para que pudiera adaptarse a distintos tamaños y formas de la superficie buscada.

La función de verosimilitud se evalúa en cada iteración para una serie de puntos, reemplazándose el peor de ellos. Los nuevos puntos son generados como combinación lineal de los anteriores. La dirección de búsqueda se define por simple comparación entre los valores de la función. El primer punto de prueba se genera reflejando el peor punto sobre el centro de la figura, es decir, regresando cada coordenada del peor punto hacia la apropiada coordenada del punto central. Si este punto es mejor

que el mejor punto anterior, se asume que la dirección de búsqueda es correcta, por lo que se genera un nuevo punto expandiendo el anterior punto de prueba en esta dirección. El mejor de los dos puntos se utiliza entonces para completar la iteración. Similarmente, si el punto reflejado es peor que el peor punto anterior, el nuevo punto de prueba se obtiene regresando más aún, o contrayendo hacia el centro. Si la contracción es inútil, es decir, si el segundo punto es todavía peor que cualquier otro en el Simplex, todos los puntos sufren una contracción, reduciendo el tamaño de la figura. Si el primer punto de prueba no es ni mejor ni peor, el peor punto es reemplazado por él y se comienza una nueva iteración. La combinación de estas operaciones de reflexión, expansión y contracción explican la adaptabilidad del procedimiento.

2.2.2.- Modelo padre y vasa jerarquizada

El modelo más completo de todos los descritos dentro de este grupo, puede presentarse de la siguiente manera:

$$Y_{i,j,k} = \tau_i + \theta_j + \beta_{jk} + \gamma_{ij} + \epsilon_{i,j,k} \quad [15]$$

con los siguientes momentos:

$$E(Y_{i,j,k}) = \tau_i + \theta_j$$

$$E(\mathbf{y}_{i,j,k}) = E(\mathbf{v}_{i,j,k}) = E(\mathbf{e}_{i,j,k}) = \mathbf{0}$$

$$\text{Var} \begin{bmatrix} \mathbf{r} \\ \mathbf{v} \\ \mathbf{s} \end{bmatrix} = \begin{bmatrix} \Lambda_1 \sigma_r^2 & 0 & 0 \\ 0 & I \sigma_v^2 & 0 \\ 0 & 0 & I \sigma_e^2 \end{bmatrix}$$

siendo:

$\mathbf{y}_{i,j,k}$ el registro de producción m , de la vaca l , hija del semental k , perteneciente éste al grupo genético j y localizada aquella en el rebano-año i .

Λ_1 el numerador de la matriz de relaciones aditivas entre sementales.

σ_r^2 , σ_v^2 y σ_e^2 , las varianzas de los efectos aleatorios semental, vaca y residual respectivamente.

En forma matricial, [15] pueda expresarse como:

$$\mathbf{y} = \mathbf{I}\mathbf{r} + \mathbf{Q}\mathbf{q} + \mathbf{3}\mathbf{s} + \mathbf{W}\mathbf{v} + \mathbf{e} \quad [16]$$

con los momentos:

$$E \begin{bmatrix} \mathbf{y} \\ \mathbf{s} \\ \mathbf{v} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{I}\mathbf{r} + \mathbf{Q}\mathbf{q} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\text{Var} \begin{bmatrix} y \\ s \\ v \\ e \end{bmatrix} = \begin{bmatrix} ZA_1s^1 + WW^1 + I\sigma_e^2 & ZA_1\sigma_s^2 & W\sigma_v^2 & I\sigma_e^2 \\ ZA_1\sigma_s^2 & A_1\sigma_s^2 & 0 & 0 \\ W\sigma_v^2 & 0 & I\sigma_v^2 & 0 \\ I\sigma_e^2 & 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

siendo:

y el vector de producciones

r el vector del efecto fijo rebaño-año

g el vector del efecto fijo grupo genético

s el vector del efecto aleatorio semental

v el vector del efecto aleatorio vaca

e el vector del efecto aleatorio residual

X, Q, Z y W las matrices diseño de los efectos r, q, s y v respectivamente.

A, la matriz de relaciones aditivas entre sementales

σ_s^2 , σ_v^2 y σ_e^2 las varianzas de los efectos aleatorios s, v y e respectivamente.

Como ya se dijo, en este modelo se asumió que las vacas no estaban relacionadas para simplificar los cálculos.

Las ecuaciones del modelo mixto (MME) para [16] son:

$$\begin{bmatrix}
 X'X & X'Q & X'Z & X'W \\
 Q'X & Q'Q & Q'Z & Q'W \\
 Z'X & Z'Q & Z'Z + \lambda_1^{-1} \alpha_1 & Z'W \\
 W'X & W'Q & W'Z & W'W + I\alpha_v
 \end{bmatrix}
 \begin{bmatrix}
 \hat{\beta} \\
 \hat{g} \\
 \hat{b} \\
 \hat{v}
 \end{bmatrix}
 =
 \begin{bmatrix}
 X'y \\
 Q'y \\
 Z'y \\
 W'y
 \end{bmatrix}
 \quad [17]$$

siendo:

$$\alpha_1 = \hat{\sigma}_v^2 / \hat{\sigma}_1^2$$

$$\alpha_v = \hat{\sigma}_v^2 / \hat{\sigma}_v^2$$

$\hat{\sigma}_1^2$, $\hat{\sigma}_v^2$ y $\hat{\sigma}_e^2$ las estimaciones respectivas de σ_1^2 , σ_v^2 y σ_e^2 .

Se ha demostrado que la estimación de componentes de varianza por máxima verosimilitud restringida (REML), mediante un modelo semental y vaca jerarquizada reduce sustancialmente el sesgo debido a eliminación de hembras (Ouweltjes y col., 1988).

Las estimaciones REML de componentes de varianza se obtuvieron, para este tipo de modelos, mediante un algoritmo del tipo EM (expectation-maximitation), descrito por Dempster y col. (1977), usando un proceso iterativo sobre las ecuaciones del modelo mixto (MME).

Definimos una inversa generalizada de la matriz de coeficientes escrita en [17] como:

$$\begin{bmatrix}
 X'X & X'Q & X'Z & X'W \\
 Q'X & Q'Q & Q'Z & Q'W \\
 Z'X & Z'Q & Z'Z + A_1^{-1} \Sigma_1 & Z'W \\
 W'X & W'Q & W'Z & W'W + I_{n_v}
 \end{bmatrix}^{-1} = \begin{bmatrix}
 C_{rr} & C_{rs} & C_{rv} & C_{rw} \\
 C_{sr} & C_{ss} & C_{sv} & C_{sw} \\
 C_{vr} & C_{vs} & C_{vv} & C_{vw} \\
 C_{wr} & C_{ws} & C_{wv} & C_{ww}
 \end{bmatrix} \quad [18]$$

Los componentes de varianza se estimaron en cada iteración según las siguientes expresiones (Meyer, 1987):

$$\hat{\sigma}_s^2 = \hat{s}' A_1^{-1} \hat{s} / [n_s - \alpha, \text{tr}(A_1^{-1} C_{ss})] \quad [19]$$

$$\hat{\sigma}_v^2 = \hat{v}' \hat{v} / [n_v - \alpha, \text{tr}(C_{vv})] \quad [20]$$

$$\hat{\sigma}_e^2 = \hat{s}' \hat{s} / [ngle + \alpha, \text{tr}(A_1^{-1} C_{ss}) + \alpha, \text{tr}(C_{ww})] \quad [21]$$

donde:

n , n_s y n_v son el número de observaciones, sementales y vacas respectivamente

$ngle = n - n_s - n_v - \text{rango}(X:Q)$, es el número de grados de libertad del error

$$\hat{s}' \hat{s} = y'(y - X\hat{r} - Q\hat{v} - Z\hat{s} - W\hat{v}) - \alpha, \hat{s}' A_1^{-1} \hat{s} - \alpha, \hat{v}' \hat{v}$$

Se utilizó un algoritmo de computación para estimar componentes de varianza por REML en un modelo con dos efectos aleatorios y un gran número de niveles de efectos fijos descrito por Meyer (1987). Las vacas y los efectos fijos son absorbidos

dentro de las ecuaciones de los sementales. La traza de la inversa correspondiente a las vacas se determina de forma indirecta, como veremos después.

Como puede observarse en las expresiones [19], [20] y [21], para estimar los componentes de varianza, el algoritmo EM necesita las soluciones de las ecuaciones del modelo mixto, la traza de C_w y la traza del producto de C_{**} por la inversa de la matriz de relaciones aditivas.

Una estrategia eficaz es absorber las ecuaciones de las vacas, y después, las del efecto rebaño-año, dentro de las correspondientes a los sementales y grupos genéticos. Si las vacas permanecen siempre en el mismo rebaño, los distintos registros de la misma vaca están jerarquizados dentro de los rebaños y así, la inversa que se necesita para absorber el efecto rebaño, puede ser calculada después de leer todos los registros producidos en cada uno de ellos.

Debido a que los sementales están jerarquizados al grupo genético al que pertenecen, las ecuaciones para aquel efecto fijo pueden ser construidas a partir de las ecuaciones de los sementales, una vez que los otros efectos han sido ya absorbidos. Por tanto, las MME omitiendo los grupos serían:

$$\begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + \lambda_0^{-1}a_0 & Z'W \\ W'X & W'Z & W'W + I_{a_0} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\alpha} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix} \quad (22)$$

Después de absorber las ecuaciones correspondientes a las vacas, [22] queda:

$$\begin{bmatrix} X'BX & X'BY \\ Z'BX & Z'BY + \lambda_0^{-1}a_0 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} X'By \\ Z'By \end{bmatrix} \quad (23)$$

siendo:

$$B = I - W(W'W + I_{a_0})^{-1}W'$$

El siguiente paso es absorber las ecuaciones del efecto fijo rebaño-año en las correspondientes a los sementailes, resultando de [23]:

$$\begin{bmatrix} Z'PX & Z'PY + \lambda_0^{-1}a_0 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} Z'Py \end{bmatrix} \quad (24)$$

siendo:

$$P = B - BX(X'BX)^{-1}X'B$$

Suponiendo que el modelo incluye el efecto grupo genético, éste se puede extraer del propio semental, descomponiéndose la expresión [24] de la siguiente manera:

$$\begin{bmatrix} Q'PQ & Q'PI \\ I'PQ & I'PI + \lambda_1' a_1 \end{bmatrix} \begin{bmatrix} \hat{g} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} Q'Py \\ I'Py \end{bmatrix}$$

Como puede verse, se necesita calcular una inversa generalizada de orden igual a la suma de sementales y grupos genéticos:

$$\begin{bmatrix} Q'PQ & Q'PI \\ I'PQ & I'PI + \lambda_1' a_1 \end{bmatrix}^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

De este modo se calculan ya las soluciones para ambos efectos:

$$\begin{bmatrix} \hat{g} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} Q'Py \\ I'Py \end{bmatrix}$$

Hasta aquí hemos obtenido las soluciones de los sementales y los grupos genéticos, así como la inversa de la matriz de coeficientes correspondiente al bloque de sementales, C_{11} . El algoritmo necesita también las soluciones para los otros dos

efectos del modelo. Estas pueden ser obtenidas a partir de las matrices usadas anteriormente durante la absorción:

$$\begin{aligned}\hat{\beta} &= (X'BX)^{-1} [X'By - X'DZ(\hat{\alpha} + \hat{q})] \\ \hat{v} &= (W'W - I_{n_2})^{-1} [W'y - W'X\hat{\beta} - W'Z(\hat{\alpha} + \hat{q})]\end{aligned}$$

Lo único que queda por calcular para poder obtener las estimaciones de los componentes de variancia, es la traza de la matriz C_{vv} , es decir, la parte de la inversa de la matriz de coeficientes correspondiente a las ecuaciones de las vacas. Esta puede ser derivada usando técnicas de matrices particionadas, con la ayuda de inversas y productos de matrices que se obtienen de los pasos de las absorciones anteriores.

Definamos:

$$\begin{aligned}H_v &= (W'W - I_{n_2})^{-1} \\ H_x &= (X'BX)^{-1} \\ L_{vv} &= X'WH_v \\ L_{1,2v} &= (Q:Z)'WH_v \\ L_{1,2x} &= (Q:Z)'BXH_x \\ T &= [L_{1,2v} - (L_{1,2v}L_{vv})] [L_{1,2v} - (L_{1,2v}L_{vv})]'\end{aligned}$$

La traza de la parte de la inversa correspondiente a las vacas es entonces:

$$\text{tr}(C_{m+1}) = \text{tr}(H_m) + \text{tr}(H_m L_{m+1} L'_{m+1}) > \text{tr}(C_m P)$$

De esta manera obtenemos las estimaciones de los componentes de varianza en cada iteración. El siguiente paso es utilizarlos como valores λ_{EM} para la siguiente vuelta. Se asumió que se había alcanzado la convergencia cuando el cambio en las estimaciones respecto a la iteración precedente era inferior a un 1%.

2.2.2.1.- Aspectos computacionales

El algoritmo EM produce estimaciones no negativas (Merrill, 1977) pero a menudo su convergencia es lenta, especialmente cuando la varianza residual es grande en comparación con los otros componentes (Laird y col., 1983), por ejemplo, cuando la heredabilidad y la repetibilidad son pequeñas. Se han sugerido muchas aproximaciones para reducir las necesidades computacionales del algoritmo EM. La mayoría de estos métodos aproximados pueden ser clasificadas en dos categorías generales: 1) métodos para reducir el número de iteraciones para alcanzar la convergencia, o 2) métodos para reducir el coste por iteración. En este estudio se utilizaron varios métodos para reducir el coste computacional.

2.2.2.1.1.- Metodos para acelerar la convergencia

Thompson y Meyer (1988) describieron una reparametrización para acelerar la convergencia del algoritmo EM. Se derivó considerando las esperanzas de los cuadrados medios en un análisis de varianza equilibrado. Al aplicar esta reparametrización sobre datos reales se obtuvieron, sin embargo, estimaciones negativas, especialmente cuando los verdaderos valores de las varianzas eran cercanos a cero. Ésta no parece ser la razón en este análisis, ya que todos los componentes de varianza eran mayores que cero. Una posible explicación de valores negativos estimados podría ser el modelo de análisis. La reparametrización se derivó, en principio, para un modelo en el que el segundo efecto aleatorio, por ejemplo la vaca, estaba jerarquizado al primero, semental. Sin embargo, en nuestro modelo, el primer efecto aleatorio, el semental, estaba jerarquizado a un efecto fijo, el grupo genético. De este modo, las esperanzas de los cuadrados medios en el análisis de varianza, serían distintas, con lo cual podría ser necesaria una reparametrización diferente.

Van Raden y Freeman (1987) presentaron un algoritmo EMMI que llamaron EMC. Mediante este algoritmo, a cada forma cuadrática se le resta un término que aproxima la esperanza inicial de la varianza residual. Fue desarrollado inicialmente para un modelo semental, pero ha sido aplicado por Bateman (1987) en un modelo jerárquico con dos efectos aleatorios observándose una disminución

del 25% en el número de iteraciones necesarias para alcanzar la convergencia. El uso del algoritmo EMC, sin embargo, no aumentó la velocidad de convergencia para los datos y el modelo usado en este estudio.

Ninguno de estos métodos disminuyó el número de iteraciones necesarias, las cuales fueron entre 22 y 60, aunque a partir de la mitad de las iteraciones, las estimaciones de heredabilidad y repetibilidad apenas se modificaron.

2.2.2.1.2.- Métodos para reducir cálculos en cada iteración

El paso que demandó mas necesidades de cálculo por iteración fue la inversión de la matriz de coeficientes en la que todos los efectos, excepto el semental, habían sido ya absorbidos. La vectorización es un proceso por el cual un grupo de números es manejado simultáneamente mediante una sencilla instrucción. Los bucles de la subrutina de inversión fueron convertidos en operaciones con matrices (VAST-E User's Guide, 1988) para incrementar la velocidad de proceso. El tiempo necesario de unidad central de proceso (UCP) para invertir una matriz está en función del cubo del orden de la misma. Se encontraron 27 padres y abuelos maternos comunes de los 328 machos con hijas, lo que resulta en una matriz de orden 355. Para reducir el coste de inversión se absorbieron en la matriz de relaciones aditivas todos aquellos

nachos sin hijas, de forma que el orden de la matriz resultante fue 328. La reducción en el tiempo de UCP fue superior así al 20%:

$$\frac{A_1' \text{ con absorción } (328)^1}{A_1' \text{ completa } (355)^1} = 0.79$$

2.3.- Programas utilizados

Los programas (FORTRAN77) utilizados para la estimación de componentes de varianza con ambos modelos fueron amablemente cedidos por Keith G. Seldman y constituyen una modificación a los programas escritos por Meyer (1987, 1988a, 1988b).

3.- Simulación

Con el fin de probar los distintos modelos de análisis, se simuló una población de tal manera que se asemejara lo más posible a la real. Para ello, se mantuvo la estructura de la población de partida, de forma que la única diferencia existente entre el archivo de datos simulados y el de datos reales, radicara, precisamente, en los parámetros a estimar.

El método de simulación consistió en una modificación al expuesto por Rönningén (1974), que permitiese incluir efectos fijos

en la población resultante. Cada registro de producción fue simulado según el siguiente modelo:

$$Y_{i,j,k} = RA_i + a_j + p_{j,k} + e_{i,j,k} \quad [25]$$

siendo:

RA_i , el nivel i del efecto fijo rebaño-año

a_j , el valor genético aditivo de la vaca j

$p_{j,k}$, el efecto ambiental permanente \times ligado a la vaca j

$e_{i,j,k}$, el efecto residual correspondiente a la lactación i de la vaca j en el rebaño-año i

El valor del efecto fijo que se introdujo en la ecuación [25] del modelo, consistió en la estimación mínima cuadrática de cada nivel de este efecto a partir de los datos originales. De este modo se intentó reflejar, lo más fielmente posible, el valor real del rebaño-año en la población de partida.

Los métodos en uso de estimación de componentes de varianzas poseen la propiedad de ser invariantes a la traslación, es decir, que los resultados obtenidos son independientes del valor de los efectos fijos incluidos en el modelo. Por la complejidad que conlleva la estimación de varianzas cuando se incluyen muchos efectos fijos, ni los datos simulados, ni en su análisis posterior, se incluyó el efecto del grupo genético en el modelo.

Para simular la parte aleatoria del modelo se obtuvieron las varianzas de cada uno de estos efectos en la población real, con el modelo descrito en [8], es decir, el modelo ancestral con la vaca jerarquizada a él incluyendo grupos genéticos. Los valores genéticos de los animales de la población base, es decir, aquellos cuyos dos padres son desconocidos, se obtuvieron como el producto de la desviación típica genética aditiva por un valor z_i extraído aleatoriamente de una distribución normal de media cero y varianzas uno:

$$a_i = z_i \sigma_a$$

El valor genético aditivo de los animales con ascendientes conocidos fue simulado según la siguiente expresión:

$$a_i = \frac{1}{2} a_j + \frac{1}{2} a_k + \epsilon_i \quad [16]$$

siendo:

j y k los padres de i

a_j , a_j y a_k los valores genéticos aditivos de los individuos j , j y k

k

ϵ_i la desviación debida al efecto Mendeliano de los genes en el individuo i

Si se asume que todos los valores genéticos de los

ambos son independientes entre sí e independientes de las desviaciones debidas al efecto Mendeliano de los genes, tomando varianzas en [16] se obtiene:

$$\sigma_y^2 = 4\sigma_g^2 + 4\sigma_d^2 + \sigma_e^2 \quad [27]$$

y por tanto:

$$\sigma_g^2 = \frac{1}{4}\sigma_y^2$$

De esta modo, y una vez obtenida la varianza de y , el mismo momento de la expresion [16], se calcula de la siguiente manera:

$$g_i = z_i \sigma_g \quad [28]$$

siendo z_i un valor tomado al azar de una distribución normal de media cero y varianza uno, distinto del citado anteriormente.

Cuando solo uno de los padres es conocido, y asumiendo independencia entre los valores genéticos y las desviaciones debidas al efecto Mendeliano de los genes, la expresion [16] difiere ligeramente:

$$a_i = 2g_i + d_i \quad [29]$$

Y [17] es ahora de la siguiente manera:

$$\sigma_1^2 = \frac{1}{2}\sigma_1^2 + \sigma_0^2$$

Y por tanto:

$$\sigma_0^2 = \frac{1}{2}\sigma_1^2$$

procediéndose en este momento igual que en [20] y pudiendo completar así la expresión [22].

El efecto ambiental permanente p_{jk} , y el efecto residual $e_{1(jk)}$, se obtuvieron de la siguiente manera:

$$p_{jk} = z_0 \sigma_0$$

$$e_{1(jk)} = z_1 \sigma_0$$

siendo z_0 y z_1 valores obtenidos al azar de una distribución normal de media cero y varianza uno, distintos a su vez de todos los anteriores.

La misma operación fue realizada para los dos caracteres que se presentaban en el archivo de datos original, es decir producción de leche y producción de grasa.

V. RESULTADOS

1.- Consanguinidad

La endogamia media obtenida fue de $4,03 \times 10^{-4} \pm 7,72 \times 10^{-5}$, siendo 0,25 y 0,0625 las consanguinidades máxima y mínima encontradas, respectivamente.

Este resultado muestra un coeficiente de endogamia medio despreciable; por otro lado, los valores máximo y mínimo obtenidos indican que este parámetro puede presentar valores superiores a los calculados. Este punto será comentado detalladamente en el capítulo correspondiente.

2.- Estimación de componentes de varianza

2.1.- Datos reales

Para realizar la simulación lo más parecida posible a la población original, se obtuvo una primera estimación del valor de las varianzas sobre la muestra de datos original ajustando un modelo padre con la vaca jerarquizada al mismo, incluyendo el grupo genético como efecto fijo así como la matriz de relaciones aditivas entre toros, tal como queda descrito en el apartado correspondiente

de la metodología.

Los valores de los parámetros a estimar, introducidos como valores a priori para la ejecución de los programas correspondientes, fueron los obtenidos por Cañón y col. (1989).

Los resultados obtenidos para los caracteres producción de leche y producción de grasa en leche se presentan en el siguiente cuadro:

Tabla 6.- Estimaciones de componentes de varianza obtenidos en la población original con el modelo padre.

Carácter	σ^2 machos	σ^2 vacas	σ^2 residual	σ^2 total
LECHE	83.920,8	544.625,6	940.190,7	1.568.737
GRASA	93,7929	629,6101	1.265,0586	1.988,4616

La transformación de las varianzas entre machos y entre vacas a varianzas genética y ambiental permanente, permite el cálculo de otros parámetros de interés como son la heredabilidad y la repetibilidad:

Tabla 7.- Estimaciones de componentes de varianza y parámetros genéticos de interés, obtenidos con el modelo padre a partir de los datos reales.

Carácter	σ^2 genét.	σ^2 amb. p.	h^2	r
LECHE	335.683,0	292.863,3	0,2140	0,4007
GRASA	375,1716	348,2314	0,1887	0,3638

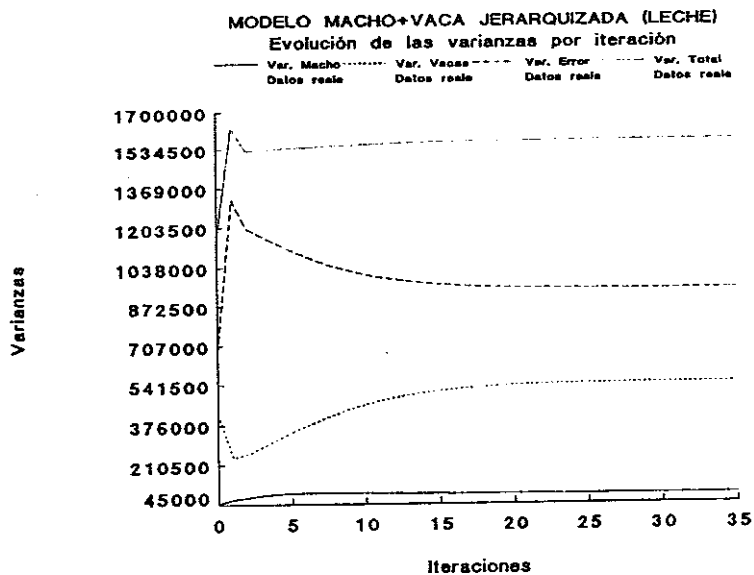


Figura 7.- Evolución de los componentes de varianza del modelo a través de las iteraciones.
CARACTER: LECHE

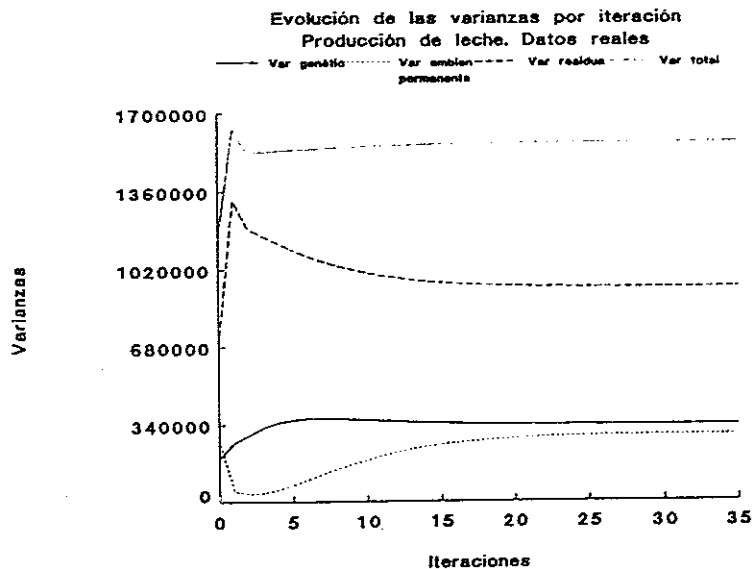


Figura 8.- Evolución de los componentes de varianza de interés a través de las iteraciones.
CARACTÉR: LECHE

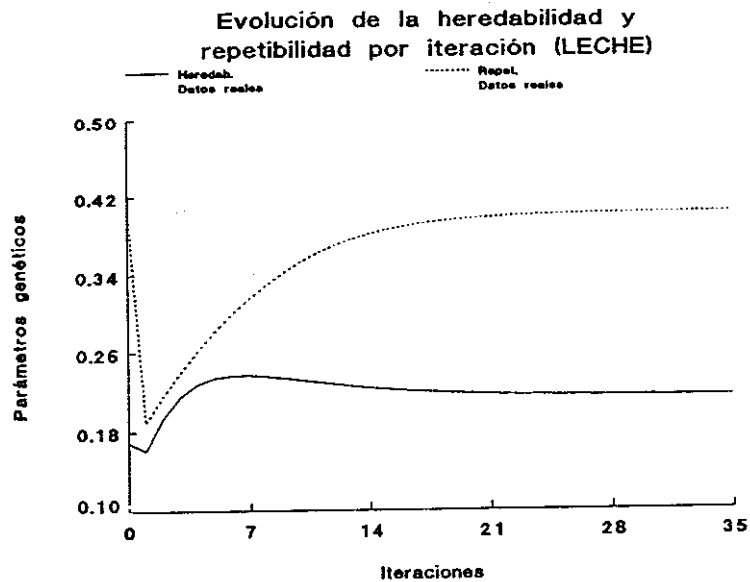


Figura 9.- Evolución de los parámetros genéticos de interés a través de las iteraciones.
CARACTER: LECHE

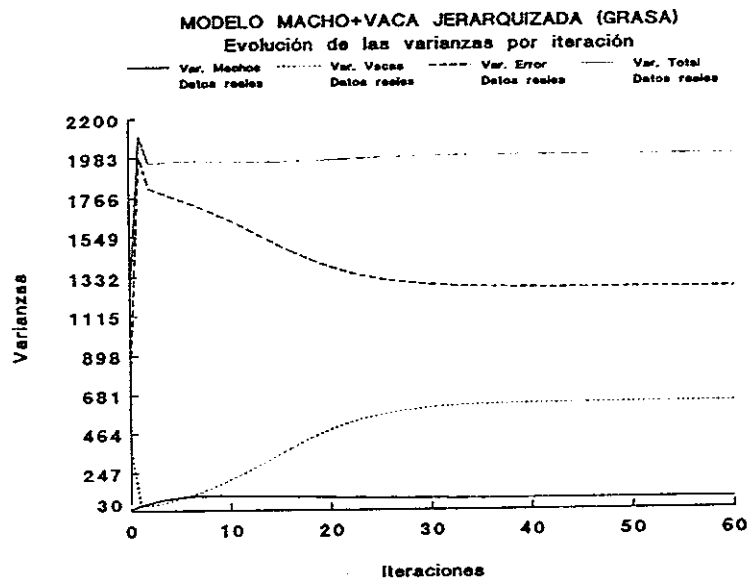


Figura 10.- Evolución de los componentes de varianza del modelo a través de las iteraciones.
CARACTER: GRASA

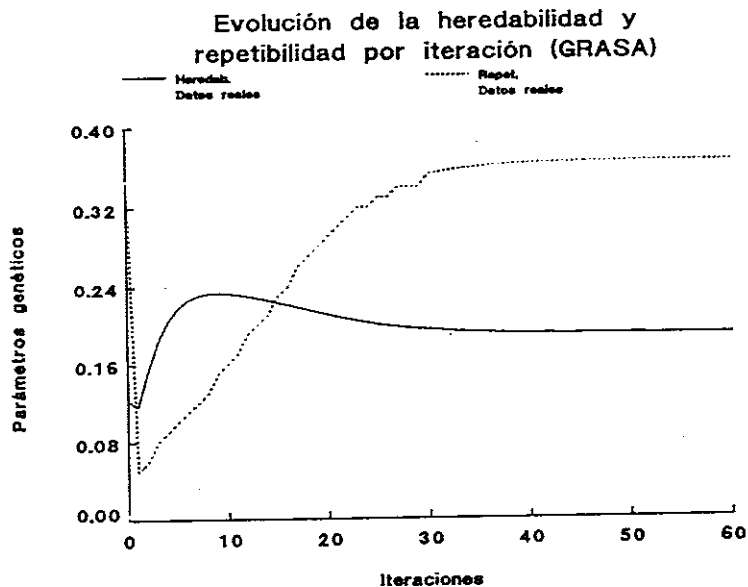


Figura 12.- Evolución de los parámetros genéticos de interés a través de las iteraciones.
CARACTER: GRASA

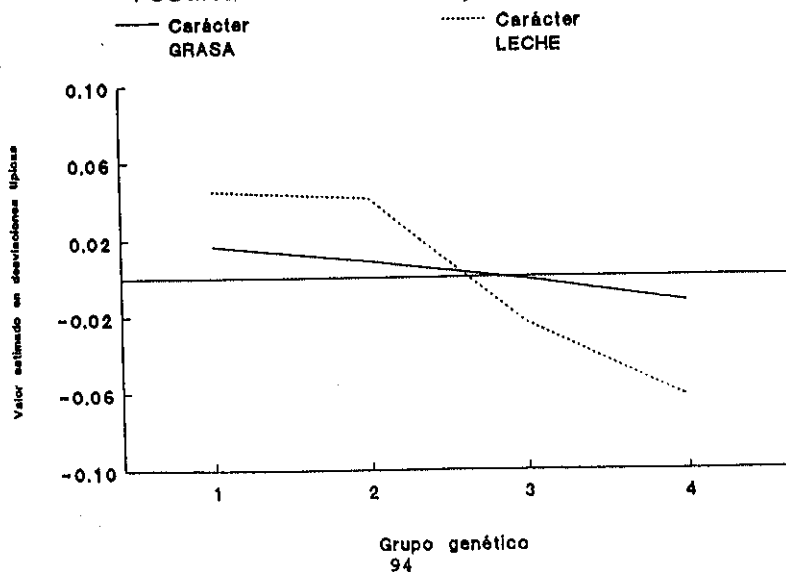
Las evoluciones de todos estos parámetros a través de las iteraciones quedan reflejadas en las figuras 7, 8, 9, 10, 11 y 12. Se observa que, una vez transcurridas aproximadamente la mitad de las iteraciones, los valores variaron muy ligeramente hasta alcanzar la convergencia.

2.1.1.- Tendencia genética

Puesto que los niveles del efecto fijo correspondiente al grupo genético fueron definidos según el año de nacimiento del padre de la vaca que produjo el registro, la estimación de este efecto, resultado de la ejecución del programa, permite observar la

Figura 13.- Tendencia genética observada por el resultado del efecto grupo genético

Tendencia genética observada por el resultado del efecto grupo genético.



tendencia genética existente en ambos caracteres. Se apreció una tendencia negativa en la población, resultando superior para el carácter producción de leche que para el carácter producción de grasa. En la figura 13 se puede observar esta tendencia con valores normalizados, es decir, desviados con respecto a la media y divididos por la desviación típica.

2.2.- Primer análisis. Comparación de tres modelos

2.2.1.- Simulación

Se simuló una población manteniendo la misma estructura que la original, es decir, que las únicas diferencias entre la población simulada y la real fueran los valores de los efectos aleatorios, manteniendo los toros con el mismo número de hijas, distribuidas entre los mismos rebaños, en los mismos años y con el mismo número de lactaciones. Las varianzas que se introdujeron en la simulación fueron los obtenidos para el fichero de datos original y representados en las tablas 6 y 7 anteriormente expuestas.

Para comprobar la validez de la simulación, para cada uno de los caracteres se obtuvo la media aritmética, así como las varianzas de los tres efectos aleatorios incluidos, antes de ser combinados para formar la lactación completa. Estos valores se

presentan en la tabla 8.

Tabla 8.- Media y varianzas de las poblaciones original y simulada.

	Media	σ^2 gen.	σ^2 per.	σ^2 res.	σ^2 total	h^2	r
LECHE							
Orig.	6.501,4	335.683	292.863	940.191	1.568.737	0,21	0,40
Simul.	6.628,3	337.352	294.719	959.087	1.591.158	0,21	0,40
GRASA							
Orig.	228,53	375,2	348,2	1.265,1	1.988,5	0,19	0,36
Simul.	246,68	374,0	350,8	1.227,1	1.951,9	0,19	0,37

2.2.2.- Resultados

En este primer análisis se compararon los siguientes modelos:

Modelo 1. Modelo padre con la vaca jerarquizada a él, incluyendo la matriz de relaciones aditivas entre toros.

Modelo 2. Modelo padre con la vaca jerarquizada a él, ignorando las relaciones existentes entre los toros.

Modelo 3. Modelo animal.

Los resultados obtenidos para tres modelos diferentes se presentan en la tabla 9. Para poder establecer comparaciones con valores esperados se incluyen en esta tabla también los valores obtenidos a partir de los datos reales con los que posteriormente se realizó la simulación. También se presentan las varianzas de los

efectos considerados independientemente, es decir, calculadas directamente sobre los efectos antes de que fueran introducidos junto con el resto para formar el dato simulado.

Tabla 9.- Primer análisis. Resultados obtenidos.

LECHE						
	σ^2 gen.	σ^2 perm	σ^2 res.	σ^2 total	h^2	r
I	418.270	338.795	982.102	1.739.167	0,24	0,44
II	426.347	333.911	982.381	1.742.640	0,24	0,44
III	431.073	329.350	982.649	1.743.072	0,25	0,44
IV	335.683	292.863	940.191	1.568.737	0,21	0,40
V	337.352	294.719	959.087	1.591.158	0,21	0,40

GRASA						
	σ^2 gen.	σ^2 perm	σ^2 res.	σ^2 total	h^2	r
I	265,7	644,0	1.269,4	2.179,0	0,12	0,42
II	263,1	646,4	1.269,2	2.178,7	0,12	0,42
III	215,6	692,4	1.269,3	2.177,3	0,10	0,42
IV	375,2	348,2	1.265,1	1.988,5	0,19	0,36
V	374,0	350,8	1.227,1	1.951,9	0,19	0,37

- I Modelo padre con relaciones entre machos
- II Modelo padre ignorando relaciones
- III Modelo animal
- IV Datos reales
- V Varianzas de los efectos simulados

Las diferencias observadas entre los tres modelos probados fueron pequeñas en general. Sin embargo, entre el modelo animal y el modelo padre fueron superiores a las encontradas entre los dos modelos padre, no pareciendo influir la inclusión de la matriz de relaciones aditivas entre sementales.

Por otro lado, las estimaciones obtenidas con todos los modelos utilizados no parecen ser muy precisas cuando se comparan con los valores esperados.

2.3.- Segundo análisis. Comparación de dos modelos con y sin efectos fijos

2.3.1.- Simulación

Se simularon dos réplicas más con los mismos valores y estructura que en el experimento anterior. Simultáneamente se hicieron réplicas en las que los efectos fijos no fueron incluidos para poder valorar su influencia en los resultados.

TABLA 10.- Medias y varianzas de las poblaciones original y simuladas.

	Media	σ^2 gen.	σ^2 perm	σ^2 res.	σ^2 total	h^2	r
LECHE							
Orig	6.501,4	335.683	292.863	940.191	1.568.737	0,21	0,40
Rép I	6.628,3	337.352	294.719	959.087	1.591.158	0,21	0,40
Rép II	6.613,7	340.867	294.663	946.009	1.581.539	0,22	0,40
Rép III	6.603,0	336.695	296.179	907.029	1.539.903	0,22	0,41
GRASA							
Orig	228,53	375,2	348,2	1.265,1	1.988,5	0,19	0,36
Rép I	246,68	374,0	350,8	1.227,1	1.951,9	0,19	0,37
Rép II	234,05	376,7	350,9	1.292,2	2.019,8	0,19	0,36
Rép III	232,20	376,9	349,1	1.268,1	1.994,1	0,19	0,36

Como en el caso anterior se obtuvieron las medias

aritméticas de ambos caracteres así como las varianzas de los efectos independientemente, es decir, antes de ser combinados para formar la lactación completa, para comprobar la validez de la simulación. Estos valores se muestran en la tabla 10.

2.3.2.- Resultados

Al no haber encontrado diferencias importantes entre los dos modelos padre analizados en el apartado anterior, en este punto sólo se realizó la comparación de dos modelos:

Modelo 1. Modelo padre con la vaca jerarquizada a él, ignorando las relaciones existentes entre los toros.

Modelo 2. Modelo animal.

Los resultados obtenidos en este análisis se presentan en las tablas 11 y 12.

TABLA 11.- Comparación de las varianzas obtenidas para cada réplica.
CARACTER: Producción de leche.

100

VARIANZAS		Aditiva	Amb. per.	Residual	Total	h^2	r
REP I	Var. de los efectos simul.	337.352	294.719	959.087	1.591.158	0,21	0,40
	Con efec. fijos Modelo padre	426.347	333.911	982.391	1.742.640	0,24	0,44
	Modelo animal	431.073	329.350	982.649	1.743.072	0,25	0,44
	Sin efec. fijos Modelo padre	332.744	240.405	958.941	1.532.090	0,22	0,37
	Modelo animal	341.021	231.240	959.250	1.531.512	0,22	0,37
REP II	Var. de los efectos simul.	340.867	294.663	946.009	1.581.539	0,22	0,40
	Con efec. fijos Modelo padre	374.996	390.394	963.988	1.729.379	0,22	0,44
	Modelo animal	402.963	367.923	964.691	1.735.576	0,23	0,44
	Sin efec. fijos Modelo padre	349.264	245.053	989.968	1.584.285	0,22	0,38
	Modelo animal	380.560	214.192	990.004	1.584.756	0,24	0,38
REP III	Var. de los efectos simul.	336.695	296.179	907.029	1.539.903	0,22	0,41
	Con efec. fijos Modelo padre	301.211	470.165	956.865	1.728.242	0,17	0,45
	Modelo animal	361.326	415.039	957.305	1.733.670	0,21	0,45
	Sin efec. fijos Modelo padre	328.572	272.774	940.626	1.541.972	0,21	0,39
	Modelo animal	342.329	261.506	941.883	1.545.717	0,22	0,39

TABLA 12.- Comparación de las varianzas obtenidas para cada réplica.
 CARACTER: Producción de grasa.

VARIANZAS		Aditiva	Amb. per.	Residual	Total	h ²	r	
REP I	Var. de los efectos simul.	374,0	350,8	1.227,1	1.951,9	0,19	0,37	
	Con efec. fijos	Modelo padre	263,1	646,4	1.269,2	2.178,7	0,12	0,42
		Modelo animal	215,6	692,4	1.269,3	2.177,3	0,10	0,42
	Sin efec. fijos	Modelo padre	258,0	439,7	1.286,9	1.984,5	0,13	0,35
Modelo animal		263,6	434,0	1.286,3	1.983,9	0,13	0,35	
REP II	Var. de los efectos simul.	376,7	350,9	1.292,2	2.019,8	0,19	0,36	
	Con efec. fijos	Modelo padre	473,5	486,1	1.327,8	2.287,5	0,21	0,42
		Modelo animal	448,4	503,0	1.328,9	2.280,4	0,20	0,42
	Sin efec. fijos	Modelo padre	261,2	399,4	1.336,7	1.997,2	0,13	0,33
Modelo animal		267,8	387,4	1.337,3	1.992,4	0,13	0,33	
REP III	Var. de los efectos simul.	376,9	349,1	1.268,1	1.994,1	0,19	0,36	
	Con efec. fijos	Modelo padre	620,1	353,9	1.277,7	2.251,7	0,28	0,43
		Modelo animal	500,5	458,1	1.280,9	2.239,4	0,22	0,43
	Sin efec. fijos	Modelo padre	450,3	314,2	1.243,9	2.008,4	0,22	0,38
Modelo animal		439,9	319,8	1.245,0	2.004,7	0,22	0,38	

En las tablas 13, 14, 15 y 16 se presentan las diferencias, absolutas y en tanto por uno, entre los valores obtenidos y los esperados en las tres réplicas, con los dos modelos, incluyendo y sin incluir efectos fijos. La participación o no de estos efectos en el modelo resultó decisiva en las estimaciones finales.

TABLA 13.- Variación, absoluta y en tanto por uno, de la estimación de la varianza genética aditiva respecto al valor esperado en las tres réplicas, con los dos modelos utilizados, con y sin la participación de los efectos fijos.

	LECHE				GRASA			
	FIJOS		NO FIJOS		FIJOS		NO FIJOS	
PADRE	88.995	0,26	-4.608	-0,01	-110,9	-0,30	-116,0	-0,31
	34.129	0,10	8.397	0,02	96,8	0,26	-115,5	-0,31
	-35.484	-0,11	-8.123	-0,02	243,2	0,65	73,4	0,19
ANIMAL	93.721	0,28	3.669	0,01	-158,4	-0,42	-110,4	-0,30
	62.096	0,18	39.693	0,12	71,7	0,19	-108,9	-0,29
	24.631	0,07	5.634	0,02	123,6	0,33	63,0	0,17

TABLA 14.- Variación, absoluta y en tanto por uno, de la estimación de la varianza ambiental permanente respecto al valor esperado en las tres réplicas, con los dos modelos utilizados, con y sin la participación de los efectos fijos.

	LECHE				GRASA			
	FIJOS		NO FIJOS		FIJOS		NO FIJOS	
PADRE	39.192	0,13	-54.314	-0,18	295,6	0,84	88,9	0,25
	95.731	0,32	-49.610	-0,17	133,2	0,38	48,5	0,14
	173.986	0,59	-23.405	-0,08	4,8	0,01	-34,9	-0,10
ANIMAL	34.631	0,12	-63.479	-0,22	341,6	0,97	83,2	0,24
	73.260	0,25	-80.471	-0,27	152,1	0,43	36,5	0,10
	118.860	0,40	-34.673	-0,12	109,0	0,31	-29,3	-0,08

Las estimaciones de los componentes de varianza genética y ambiental permanente se alejaron más de los valores esperados

cuando se incluían los efectos fijos que cuando eran omitidos.

La estimación del componente de varianza residual se acercó en todos los casos al valor esperado. La varianza total, sin embargo, fue sobreestimada cuando se incluían los efectos fijos en el modelo, reafirmando así su influencia sobre el resultado final.

TABLA 15.- Variación, absoluta y en tanto por uno, de la estimación de la varianza residual respecto al valor esperado en las tres réplicas, con los dos modelos utilizados, con y sin la participación de los efectos fijos.

	LECHE		GRASA	
	FIJOS	NO FIJOS	FIJOS	NO FIJOS
PADRE	23.304 0,02	-146 -0,00	42,1 0,03	59,8 0,05
	17.976 0,02	43.959 0,05	35,6 0,03	44,5 0,03
	49.836 0,05	33.597 0,04	9,6 0,01	-24,2 -0,02
ANIMAL	23.562 0,02	163 0,00	42,2 0,03	59,2 0,05
	18.862 0,02	43.995 0,05	36,7 0,03	45,1 0,03
	50.276 0,06	34.854 0,04	12,8 0,01	-23,1 -0,02

TABLA 16.- Variación, absoluta y en tanto por uno, de la estimación de la varianza total respecto al valor esperado en las tres réplicas, con los dos modelos utilizados, con y sin la participación de los efectos fijos.

	LECHE		GRASA	
	FIJOS	NO FIJOS	FIJOS	NO FIJOS
PADRE	151.482 0,10	-59.068 -0,04	226,8 0,12	32,6 0,02
	147.840 0,09	2.746 0,00	267,7 0,13	-22,6 -0,01
	188.339 0,12	2.069 0,00	257,6 0,13	14,3 0,01
ANIMAL	151.914 0,10	9.645 -0,04	225,4 0,12	32,0 0,02
	154.037 0,10	3.217 0,00	260,6 0,13	-27,4 -0,01
	193.767 0,13	5.814 0,00	245,3 0,12	10,6 0,01

Respecto a la comparación de modelos, los resultados

obtenidos para las nuevas réplicas reflejan la misma situación que en el primer caso, es decir, que las diferencias observadas entre ambos son prácticamente inapreciables dentro de cada una de las bases de datos creadas, más aún si se comparan con las que existen entre los valores obtenidos y los esperados.

2.4.- Tercer análisis. Comparación con distintas heredabilidades en un solo carácter

2.4.1.- Simulación

Se simularon dos nuevas réplicas, sin incluir efectos fijos, para el carácter producción de grasa en leche de forma que sólo se diferenciase de la tercera réplica del experimento anterior en la magnitud de las varianzas genética aditiva y residual, y consecuentemente en la heredabilidad. Las dos nuevas repeticiones incluyeron heredabilidades de 0,10 y 0,28.

Para ello se mantuvieron constantes los valores obtenidos aleatoriamente de una distribución normal de media cero y varianza uno y se modificó sólo la magnitud de las varianzas antes mencionadas con el fin de obtener los valores de los efectos aleatorios que intervienen en el modelo.

2.4.2.- Resultados

El resultado para las dos nuevas réplicas produjo un sesgo de la misma magnitud, dirección y sentido que la anterior demostrando que las diferencias observadas entre caracteres no se deben a que posean distinta heredabilidad, ni a la similar magnitud de las varianzas genética y ambiental permanente (Tabla 17).

TABLA 17.- Comparación de los resultados obtenidos, cuando se modifica la heredabilidad, manteniendo la varianza ambiental permanente y los valores extraídos aleatoriamente de una distribución normal. Entre paréntesis aparecen las diferencias en tanto por uno respecto al valor esperado.

CARACTER: Producción de grasa.

	Aditiva	Amb. per.	Residual	Total	h^2	r
<hr/>						
h^2 0,10						
Orig.	199,8	349,1	1.427,5	1.976,4	0,10	0,28
M.padre	230,8 (0,16)	346,9 (-0,01)	1.421,0 (-0,00)	1.998,8 (0,01)	0,12	0,29
M.animal	236,6 (0,18)	338,3 (-0,03)	1.421,8 (-0,00)	1.996,8 (0,01)	0,12	0,29
<hr/>						
h^2 0,19						
Orig.	376,9	349,1	1.268,1	1.994,1	0,19	0,36
M.padre	450,3 (0,19)	314,2 (-0,10)	1.243,9 (-0,02)	2.008,4 (0,01)	0,22	0,38
M.animal	439,9 (0,17)	319,8 (-0,08)	1.245,0 (-0,02)	2.004,7 (0,01)	0,22	0,38
<hr/>						
h^2 0,28						
Orig.	559,4	349,1	1.083,0	1.991,5	0,28	0,46
M.padre	665,2 (0,19)	286,7 (-0,18)	1.072,1 (-0,01)	2.024,0 (0,02)	0,33	0,47
M.animal	640,9 (0,15)	304,6 (-0,13)	1.073,2 (-0,01)	2.018,7 (0,01)	0,32	0,47
<hr/>						

2.5.- Resultados de computación

En el tercer experimento se midió el tiempo de UCP consumido por cada uno de los programas, así como el número de iteraciones realizadas por ellos. Los resultados obtenidos muestran que el modelo animal alcanzó la convergencia en menos tiempo y en menos iteraciones que el modelo padre (Tabla 18).

Tabla 18.- Comparación de necesidades computacionales en el tercer experimento. Carácter grasa.

h^2		Iteraciones	Tiempo UCP
0,10	Modelo animal	17	4429
	Modelo padre	40	5305
0,28	Modelo animal	13	2593
	Modelo padre	26	3432

Aunque tanto el número de iteraciones como el tiempo total necesario para alcanzar la convergencia fueron inferiores en el modelo animal, en éste el tiempo invertido por iteración fue superior.

La interpretación de las diferencias de tiempo debe hacerse con precaución, ya que los algoritmos de estimación son diferentes en cada modelo. Además, el programa utilizado para el modelo animal, a diferencia del empleado para el modelo padre, se encuentra altamente optimizado.

VI. DISCUSION

1.- Consanguinidad

De acuerdo con los resultados presentados, el nivel de consanguinidad medio es despreciable, lo cuál, desde un punto de vista general, es positivo. No obstante, mediante un análisis más detallado de los datos se puede llegar a la conclusión de que la información de pedigrí presenta parece muy escasa como para considerar estos resultados como definitivos.

En primer lugar, el número de individuos cuyos padres son desconocidos, y que por lo tanto se consideran población base, supone el 40% del total, lo cuál indica la escasa información de parentesco de que se dispone. Esto provoca que la consanguinidad calculada pueda ser más baja que la real.

Un dato que permite intuir la existencia de una endogamia superior a la calculada en este análisis es que la consanguinidad máxima obtenida (0,25) corresponde a aquella que se produce cuando los abuelos paternos son los mismos que los maternos, lo que supone la máxima endogamia posible cuando sólo tres generaciones son consideradas. Dado que el número medio de generaciones encontradas es de 3,2 y que además sólo el 17,5% de los animales superan este número de generaciones, este dato debería ser cuando menos preocupante.

Por todo ello, parece posible que, de haber tenido mayor cantidad de información de pedigrís, los resultados tal vez hubieran sido otros muy diferentes.

Sería necesario un fichero de genealogías más completo para obtener conclusiones de mayor peso que las expuestas.

2.- Componentes de varianza

2.1.- Análisis con datos reales. Modelo padre

Los parámetros genéticos obtenidos, para ambos caracteres, no difieren sustancialmente de los obtenidos hasta la fecha por Cañón y col. (1989) y Ben Gara y col. (1991), aunque fueron superiores a los encontrados por Carabaño y col. (1989) y Gómez y col. (datos no publicados) (Tabla 19).

Tabla 19.- Comparación de las heredabilidades obtenidas con otras anteriores de la misma población.

	Heredabilidad	
	Kg de leche	Kg de grasa
Carabaño y col. 1989	0,16	0,14
Cañón y col. 1989	0,17	0,13
Gómez y col. 1989	0,12	0,10
Ben Gara y col. 1991	0,24	0,18
Obtenidos	0,21	0,19

Parece lógico pensar que dos estimaciones de los mismos parámetros en una misma población deberían ser aproximadamente iguales. Sin embargo, Chauhan y Hayes (1991), trabajando con datos pertenecientes a la raza Holstein-Friesian canadiense, obtuvieron discrepancias de magnitud similar a las nuestras respecto a las estimaciones obtenidas por Cue y col. (1987) en la misma población, atribuyendo las diferencias al uso de distintas bases de datos bajo modelos diferentes.

En las figuras 7, 8, 9, 10, 11 y 12 se presenta la evolución de las estimaciones de las varianzas a través de las iteraciones, así como la evolución de la heredabilidad y repetibilidad, observándose que el criterio de convergencia utilizado podría haber sido menos estricto, ya que los resultados variaron muy ligeramente a partir de aproximadamente la mitad de las iteraciones.

La heredabilidad del carácter producción de leche fue ligeramente superior a la del carácter producción de grasa en leche, tal como ya había sido detectado por otros investigadores (Cañón y col., 1989, Carabaño y col., 1989, Ben Gara y col. 1991). Cuando el carácter a analizar es el porcentaje de grasa en leche, en lugar de kilogramos de grasa la heredabilidad de aquel carácter suele aparecer superior a la del otro (Barker y Robertson 1966, Tong y col. 1979, Hill y col. 1983, Meyer 1985a, Carabaño y col. 1990).

La validez de estos resultados depende fundamentalmente de la información disponible. La estructura de los datos tiene un claro efecto sobre la precisión de las estimaciones (Juga y Thompson, 1989). En nuestro caso, el análisis realizado inicialmente sobre la estructura y equilibrio de los datos parece indicar que el resultado puede caracer de precisión, como veremos después.

2.1.1.- Tendencia genética

En un modelo lineal en el que no se incluye el efecto grupo genético, todos los animales de genealogía desconocida son considerados como individuos pertenecientes a una misma población base. Si estos animales, cuyo pedigrí se ignora, han nacido en distintos años, no parece correcto suponer que tengan un origen común, ya que son el resultado de decisiones de selección tomadas en generaciones anteriores y, por tanto, es lógico pensar que difieran en el valor genético medio (Cañón y Gutiérrez, 1989).

Para poder corregir el efecto de la posible tendencia genética en nuestra población, se incluyó en el modelo el efecto fijo grupo genético, cuyos niveles fueron definidos según el año de nacimiento del padre de la vaca que produce el registro.

El programa utilizado para la estimación de componentes

de varianza permitió, a su vez, obtener estimaciones para este efecto, lo cual permite dibujar una tendencia genética en la población de trabajo.

La tendencia genética observada por los valores estimados para cada nivel del efecto grupo genético resultó ser negativa, coincidiendo con la que había sido observada por Cañón y Muñoz, (1991) y Alenda y col. (1986) para esta misma población.

Aunque en otras poblaciones de ganado vacuno lechero ya habían sido detectadas tendencias negativas (Van Vleck y col. 1986, Powell y col. 1977), éstas se producían sólo en un cierto intervalo de tiempo y eran localizadas en regiones muy concretas. En nuestro caso, sin embargo, se producen a nivel nacional a lo largo de todo el periodo de estudio haciendo muy difícil la posibilidad de encontrar una explicación racional de la tendencia genética observada.

Si el objetivo de selección es mejorar las producciones de los animales y éstas se pueden medir, las decisiones a tomar deberían estar basadas en estas mismas producciones. Sin embargo, las decisiones de selección de los sementales en nuestro país ha estado fuertemente influido por las valoraciones morfológicas de los animales (Cañón y Muñoz, 1991). Como consecuencia, algunos machos con buen índice morfológico y desconocido mérito genético para caracteres de producción, han sido utilizados de forma masiva

por todo el país. El progreso genético que se puede esperar en estas circunstancias para los caracteres de producción, depende de la correlación existente entre los dos tipos de caracteres.

Otra posible causa puede encontrarse en el intercambio de material genético con otros países. La población española de ganado vacuno de raza Frisona no puede considerarse como un núcleo cerrado, ya que está sujeta a importaciones de animales procedentes de otros países, generalmente de Estados Unidos y Canadá. La política seguida en este sentido no ha estado en manos de profesionales del mundo de la genética y, como consecuencia, las acciones realizadas podrían haber carecido de apoyo científico.

Otra razón para justificar una tendencia genética negativa podría encontrarse en el coeficiente de endogamia medio, discutido en un apartado anterior. La consanguinidad produce disminución de las producciones como consecuencia de la expresión de genes que segregan a baja frecuencia. Tal y como fue expuesto previamente, algunos datos indicaron la existencia de una consanguinidad posiblemente superior a la obtenida.

La elección de los objetivos de selección correctos, los esquemas de selección a seguir, los métodos de valoración de reproductores y las importaciones de material genético, deberían ser vigilados con atención en nuestro país. La existencia de una tendencia genética negativa en nuestra población nos sugiere que,

uno o varios de estos capítulos, han estado muy desatendidos o mal gestionados.

2.2.- Primer análisis. Comparación de tres modelos

Se compararon en esta fase los modelos siguientes:

- Modelo 1. Modelo padre con la vaca jerarquizada al mismo incluyendo la matriz de relaciones aditivas entre machos.
- Modelo 2. Modelo padre con la vaca jerarquizada al mismo ignorando las relaciones existentes entre los machos.
- Modelo 3. Modelo animal.

Los resultados (tabla 20) muestran estimaciones muy parecidas para todos los modelos puestos a prueba, especialmente para el carácter producción de leche. Sin embargo, los valores obtenidos para los modelos 1 y 2 son muy similares y distintos del modelo 3, más aún en el segundo carácter analizado. Consecuentemente, la inclusión de la matriz de relaciones aditivas existentes entre los machos no afecta de forma significativa al resultado, y, por tanto, se puede considerar que la información genealógica existente en el archivo de pedigrís de los sementales puede ser ignorada.

En la misma tabla se puede observar, por otro lado, que las estimaciones obtenidas para los componentes de varianza

genética y ambiental permanente se alejaron mucho de los esperados en todos los casos.

Tabla 20.- Primer análisis. Comparación de los resultados obtenidos con los esperados. Las diferencias, entre paréntesis, son expresadas en tanto por uno respecto al valor esperado.

LECHE						
	σ^2 gen.	σ^2 per.	σ^2 resid.	σ^2 total	h^2	r
Esp	337.352	294.719	959.087	1.591.158	0,21	0,40
I	418.270 (0,24)	338.795 (0,15)	982.102 (0,02)	1.739.167	0,24	0,44
II	426.347 (0,26)	333.911 (0,13)	982.381 (0,02)	1.742.640	0,24	0,44
III	431.073 (0,28)	329.350 (0,12)	982.649 (0,02)	1.743.072	0,25	0,44
GRASA						
	σ^2 gen.	σ^2 per.	σ^2 resid.	σ^2 total	h^2	r
Esp	374,0	350,8	1.227,1	1.951,9	0,19	0,37
I	265,7 (-0,29)	644,0 (0,84)	1.269,4 (0,03)	2.179,0	0,12	0,42
II	263,1 (-0,30)	646,4 (0,84)	1.269,2 (0,03)	2.178,7	0,12	0,42
III	215,6 (-0,42)	692,4 (0,97)	1.269,3 (0,03)	2.177,3	0,10	0,42

Esper. Varianzas de los efectos simulados
 I Modelo padre con relaciones entre machos
 II Modelo padre ignorando relaciones de parentesco
 III Modelo animal

Los experimentos de laboratorio son diseñados optimizando la estructura en busca de lograr el fin que se propone (Hill, 1970). En los datos de campo, por el contrario, la estructura de la población puede tener una fuerte influencia sobre la calidad de las estimaciones obtenidas. En nuestro caso, una vez vistos los

resultados y dado el fuerte desequilibrio expuesto, este punto parece haber intervenido de forma concluyente, tanto para la estimación de la varianza genética aditiva, como para la de la varianza ambiental permanente.

La varianza total se calcula directamente de las observaciones y la varianza residual, como quedó mostrado en la metodología, se estima también directamente. Los otros dos componentes, la varianza genética aditiva y la varianza ambiental permanente, dependen directamente de la estructura de los datos para que el método reparta correctamente el resto de la varianza total en uno y otro componente.

La similitud observada entre los dos modelos padre y el modelo animal es fácilmente explicable, ya que tan sólo del 15% de las vacas se conoce su madre (Tabla 3).

Una posible causa de haber obtenido una mala estimación de la varianza genética aditiva podría ser la desigual distribución de los machos según el número de hijas. Si sólo unos pocos de ellos intervienen como padres de muchas vacas, éstos podrían sesgar el resultado. En las figuras 14, 15, y 16 se representan comparativamente el caso real junto con los dos casos extremos en los que la distribución de los machos es óptima y pésima. Aunque el caso real se encuentra entre los dos, conviene llamar la atención sobre la existencia de machos que tienen muchas más hijas que el

Figura 14.- Distribución óptima de los toros según el número de hijas

DISTRIBUCION DE LOS TOROS SEGUN EL NUMERO DE HIJAS (caso óptimo)



Figura 15.- Distribución real de los toros según el número de hijas

DISTRIBUCION REAL DE LOS TOROS SEGUN EL NUMERO DE HIJAS



Figura 16.- Distribución pésima de los toros según el número de hijas

DISTRIBUCION DE LOS TOROS SEGUN EL NUMERO DE HIJAS (caso pésimo)



resto, ya que si el mérito genético de éstos se encuentra en un extremo de la distribución normal de méritos genéticos, el de aquéllos podría sesgar el resultado.

Sin embargo, éste no es el principal motivo de una mala estimación de varianza genética, ya que el desequilibrio expuesto es corregido cuando todos los animales de la población se encuentran ligados entre sí mediante la matriz de relaciones aditivas. Se podría pensar que se trataba de un problema de desconexión pero, una vez comprobado que todos los animales estaban ligados a través de la matriz de relaciones aditivas se rechazó esta hipótesis.

Durante la simulación se asignó un valor genético a todos los animales de la población base. Posteriormente, durante la depuración de los datos, se retiraron todos aquellos que sólo intervenían como padres una vez, ya que inducían a confusión al tiempo que no aportaban información alguna. El desconocimiento de uno de los padres en el 85 % de los animales con registro origina una matriz de relaciones aditivas incompleta, ya que las madres de todas las vacas estaban identificadas aunque se había asumido lo contrario. Jensen y col. (1990) demostraron que sólo se producían estimas insesgadas cuando se introducía la matriz de relaciones aditivas completa.

Según lo expuesto, la poca información aportada por las

madres y el desequilibrio existente en el número de hijas por cada macho parecen ser las causas más importantes de la mala estimación obtenida para la componente de varianza genética.

Más grave resulta la situación en el caso de la componente de varianza ambiental permanente. Los programas de preparación de los datos no obligan a que las vacas tengan más de un registro, por lo que muchos animales que sólo hayan producido una lactación forman parte del archivo definitivo. Una buena estimación de la varianza ambiental permanente se conseguirá cuando muchos animales tengan muchos registros, mientras que aquella será imposible cuando todas las vacas hayan producido sólo una vez. Como se puede apreciar en las figuras 17, 18 y 19, la estructura de nuestros datos se acerca más al segundo caso que al primero.

Así pues, las causas fundamentales de las malas estimaciones obtenidas, parecen ser el desequilibrio existente en la población de trabajo y la incompleta información disponible sobre el parentesco de los animales. No obstante, existen otros factores que pueden llevar a la dispersión en las estimaciones, bien porque aumenten el sesgo producido, o porque tiendan a corregirlo.

Es conocida la influencia del número de lactación de la vaca sobre la producción de leche, produciéndose un aumento en cada lactación hasta una determinada a partir de la cuál se disminuye la

Figura 17.- Distribución óptima de las vacas según el número de lactaciones

Distribución óptima de las vacas según el número de lactaciones



Figura 18.- Distribución real de las vacas según el número de lactaciones

Distribución real de las vacas según el número de lactaciones

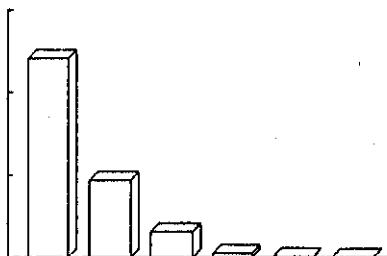
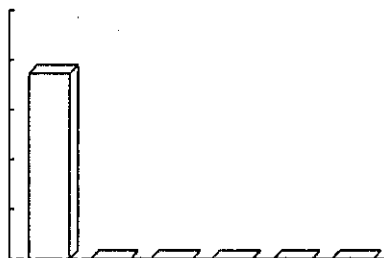


Figura 19.- Distribución de las vacas según el número de lactaciones que imposibilita la estimación de la varianza ambiental permanente

Distribución pésima de las vacas según el número de lactaciones



producción (Lush y Shrode, 1950). Este hecho no es tenido en cuenta en la simulación, en la cuál, sin embargo, se incluye el efecto fijo del año de producción tal y como influye en la población original, ya que en la simulación es introducida la estimación mínimo cuadrática de este efecto. Los distintos años aparecen así representados, dentro del mismo rebaño, pudiendo inducir a confusión con el efecto aleatorio ambiental permanente, efecto que queda definido cuando algunos animales presentan más de un registro. Además, la introducción de varianza por parte de los efectos fijos quedó patente por el hecho de haber obtenido una varianza total superior a la esperada tras la ejecución de los programas utilizados.

Otro hecho que cabe destacar es la desigual precisión del método al aplicarlo sobre distintos caracteres. Aunque la inclusión de los efectos fijos podría haber influido más en un carácter que en otro, estas diferencias podrían ser debidas a la distinta magnitud de la heredabilidad o simplemente al azar.

Para tratar de aclarar estos aspectos se hicieron análisis más detallados.

2.3.- Segundo análisis. Comparación de dos modelos con y sin efectos fijos

Los resultados obtenidos en el punto anterior demostraron que la inclusión de la matriz de relaciones aditivas existentes entre los machos prácticamente no producía variaciones significativas en los valores finales. Por otro lado, cada una de las ejecuciones de los programas de estimación utilizados, independientemente del modelo ajustado, resultaba muy cara y lenta. Por estas razones, el modelo padre incluyendo relaciones de parentesco fue eliminado en los análisis posteriores.

En este segundo paso se intentaron clarificar algunas cuestiones que habían quedado oscuras en el apartado anterior. Se intentó evaluar la influencia del azar simulando dos réplicas más con los mismos parámetros de entrada que para la primera.

Ibáñez y col. (1991) comprobaron la existencia de interacción genotipo-ambiente en esta misma población. En estas condiciones la estimación de los efectos ambientales viene confundida con la de otros efectos del modelo. Al ser esta estimación introducida en los datos simulados, éstos se podrían haber visto afectados. Para poder valorar la influencia de los efectos fijos sobre el resultado final, cada una de las réplicas se simuló incluyendo los efectos fijos y sin ser éstos incluidos.

Los resultados expuestos en las tablas 11 y 12 mostraron que la inclusión de los efectos fijos, tal y como se consideraban en la simulación, producían confusión al intentar obtener estimaciones finales, ya que introducían varianzas que se iba a repartir entre los dos componentes más conflictivos. Se aprecia directamente que la supresión de los efectos fijos hizo que la varianzas total se acercase mucho más a la simulada. Las componentes de varianzas genética y ambiental permanente también fueron mejores sin la participación de los efectos fijos. Esta conclusión se hace más evidente si las diferencias absolutas entre el resultado obtenido y el esperado se expresan en tanto por uno (tabla 21).

Aunque la influencia de los efectos fijos queda patente, los resultados obtenidos, una vez eliminados los efectos fijos, permanecen alejados de los valores esperados. El modelo animal sigue teniendo un comportamiento muy similar al modelo padre hasta el punto de no quedar nada claro cuál de ellos es mejor. Después de haber realizado doce comparaciones diferentes se puede llegar a una conclusión: la información que aportan los datos es tan pobre, que modelos más completos no producen mejores estimaciones y, por tanto, el modelo más indicado para estimar componentes de varianzas en esta población será aquél que resulte más económico. La estimación de componentes de varianzas obtenida anteriormente con los datos reales es, según lo expuesto, tan buena como cualquier otra que pudiera lograrse con modelos más complejos. En cualquier caso se puede decir al mismo tiempo que por muy completo que sea el

TABLA 21.- Variación, absoluta y en porcentaje, de las estimaciones obtenidas respecto a las esperadas, en las tres réplicas, con los dos modelos utilizados al incluir los efectos fijos y al no incluirlos.

LECHE												
	Varianza genética aditiva			Varianza ambiental permanente			Varianza residual					
	FIJOS		NO FIJOS	FIJOS		NO FIJOS	FIJOS		NO FIJOS			
PADRE	88.995	0,26	-4.608	-0,01	39.192	0,13	-54.314	-0,18	23.304	0,02	-146	-0,00
	34.129	0,10	8.397	0,02	95.731	0,32	-49.610	-0,17	17.976	0,02	43.959	0,05
	-35.484	-0,11	-8.123	-0,02	173.986	0,59	-23.405	-0,08	49.836	0,05	33.597	0,04
ANIMAL	93.721	0,28	3.669	0,01	34.631	0,12	-63.479	-0,22	23.562	0,02	163	0,00
	62.096	0,18	39.693	0,12	73.260	0,25	-80.471	-0,27	18.862	0,02	43.995	0,05
	24.631	0,07	5.634	0,02	118.860	0,40	-34.673	-0,12	50.276	0,06	34.854	0,04
GRASA												
	Varianza genética aditiva			Varianza ambiental permanente			Varianza residual					
	FIJOS		NO FIJOS	FIJOS		NO FIJOS	FIJOS		NO FIJOS			
PADRE	-110,9	-0,30	-116,0	-0,31	295,6	0,84	88,9	0,25	42,1	0,03	59,8	0,05
	96,8	0,26	-115,5	-0,31	133,2	0,38	48,5	0,14	35,6	0,03	44,5	0,03
	243,2	0,65	73,4	0,19	4,8	0,01	-34,9	-0,10	9,6	0,01	-24,2	-0,02
ANIMAL	-158,4	-0,42	-110,4	-0,30	341,6	0,97	83,2	0,24	42,2	0,03	59,2	0,05
	71,7	0,19	-108,9	-0,29	152,1	0,43	36,5	0,10	36,7	0,03	45,1	0,03
	123,6	0,33	63,0	0,17	109,0	0,31	-29,3	-0,08	12,8	0,01	-23,1	-0,02

modelo la estimación que se obtenga será muy poco precisa.

TABLA 22.- Variación, absoluta y en porcentaje, de las estimaciones obtenidas respecto a las esperadas, en las tres réplicas, con los dos modelos utilizados al no incluir los efectos fijos.

LECHE						
	Var. gen.		Var. amb.per.		Var. resid.	
PADRE	-4.608	-0,01	-54.314	-0,18	-146	-0,00
	8.397	0,02	-49.610	-0,17	43.959	0,05
	-8.123	-0,02	-23.405	-0,08	33.597	0,04
ANIMAL	3.669	0,01	-63.479	-0,22	163	0,00
	39.693	0,12	-80.471	-0,27	43.995	0,05
	5.634	0,02	-34.673	-0,12	34.854	0,04
GRASA						
	Var. gen.		Var. amb.per.		Var. resid.	
PADRE	-116,0	-0,31	88,9	0,25	59,8	0,05
	-115,5	-0,31	48,5	0,14	44,5	0,03
	73,4	0,19	-34,9	-0,10	-24,2	-0,02
ANIMAL	-110,4	-0,30	83,2	0,24	59,2	0,05
	-108,9	-0,29	36,5	0,10	45,1	0,03
	63,0	0,17	-29,3	-0,08	-23,1	-0,02

Otra pregunta que buscaba respuesta al realizarse este segundo experimento, era por qué la estimación obtenida para el carácter producción de leche resultaba más precisa que para el carácter producción de grasa en leche. Una razón podría haber sido una influencia superior de los efectos fijos sobre un carácter que sobre el otro, pero una vez eliminados éstos las diferencias se mantienen (Tabla 22).

Si los datos tuviesen su origen en una población real,

probablemente sería muy sencillo encontrar una o varias explicaciones a este resultado. Sin embargo, se trata de datos simulados, y por tanto, ambos caracteres presentan la misma estructura y están libres de la influencia de los efectos fijos, de forma que podrían considerarse más como seis réplicas del mismo carácter que como tres réplicas de dos caracteres diferentes. La única diferencia que se aprecia es que poseen distinta heredabilidad, mientras que la repetibilidad es muy similar en ambos. Traducido a términos de varianzas, el primer carácter posee mayor varianza genética y menor varianza ambiental permanente que el otro en proporción a la varianza fenotípica, produciéndose, en el segundo, un gran parecido entre ambos componentes de la varianza total.

Para poder aclarar el diferente comportamiento del método en uno y otro carácter, se precisa realizar un nuevo análisis.

2.4.- Tercer análisis. Comparación con distintas heredabilidades en un sólo carácter

Las diferencias de precisión obtenidas para los dos caracteres en estudio hacían sospechar que la eficacia del método REML dependía, de alguna manera, de la magnitud de ciertos componentes de la varianza total, tal y como fue reflejado anteriormente.

Con el fin de poder evaluar estas diferencias, se hicieron dos réplicas más del segundo carácter manteniendo constante la varianza ambiental permanente y la varianza total, de manera que sólo se modificase la heredabilidad del carácter. Los valores de la heredabilidad para los cuales se hizo la simulación fueron 0,10 y 0,28, cubriendo un rango en cuyo centro se encuentra el valor 0,19 incluido en la simulación anterior. Para que la comparación fuera lo más fiable posible, sólo se modificó la magnitud de las varianzas en el sentido antes mencionado, utilizando los mismos valores de la distribución normal de media cero y varianza uno que los introducidos en la tercera réplica del último análisis.

En las tres estimaciones obtenidas para la varianza genética se produjo un sesgo en la misma dirección y sentido, demostrando que las diferencias encontradas anteriormente entre ambos caracteres se deben exclusivamente a la elevada dispersión de las estimaciones. El mejor comportamiento del método en uno de ellos parece que fue debido, por tanto, exclusivamente al azar.

2.5.- Resultados de computación

En la tabla 18 quedó reflejado el superior rendimiento del programa utilizado para el modelo animal frente al utilizado para el modelo padre. Las diferencias en tiempo no deben ser tomadas como tales, ya que las recientes modificaciones del primer

programa lograron una velocidad 600 veces superior.

Las diferencias en iteraciones son más significativas aunque, como puede apreciarse en las figuras 7, 8, 9, 10, 11 y 12 pertenecientes al modelo padre, a partir de la mitad de las iteraciones apenas se produjeron cambios. Se consideraba que se había alcanzado la convergencia cuando el cambio en los valores obtenidos en dos iteraciones sucesivas era inferior a un 1%. En la tabla 23 se representan las estimaciones finales de componentes de varianza frente a los valores introducidos *a priori*, comparándolos a su vez con valores intermedios. En la tabla 24 se presentan de igual modo los parámetros genéticos de interés.

Tabla 23.- Estimaciones de componentes de varianza finales en comparación con valores iniciales y con valores intermedios.

LECHE						
Iter.	σ^2 mac	σ^2 vacas	σ^2 gen.	σ^2 perm.	σ^2 res.	σ^2 total.
0	48.634	421.245	194.535	275.342	679.118	1.148.997
20	84.862	534.397	339.448	279.811	947.402	1.566.661
35	83.921	544.626	335.683	292.863	940.191	1.568.737
GRASA						
Iter.	σ^2 mac	σ^2 vacas	σ^2 gen.	σ^2 perm.	σ^2 res.	σ^2 total
0	37,3	395,1	149,2	283,2	778	1.210
35	94,8	618,8	379,2	334,4	1.273	1.986
60	93,8	629,6	375,2	348,2	1.265	1.988

El parecido entre los valores intermedios y los valores finales parece indicar que el criterio exigido podría haber sido

menos rígido, con la consiguiente disminución en iteraciones y tiempo.

Tabla 24.- Estimaciones de la heredabilidad y la repetibilidad en comparación con valores iniciales e intermedios.

LECHE			GRASA		
Iteración	h^2	r	Iteración	h^2	r
0	0,17	0,41	0	0,12	0,35
20	0,22	0,40	35	0,19	0,36
35	0,21	0,40	60	0,19	0,36

No está claro cuál debe ser el criterio de convergencia a utilizar. Quizá lo más lógico sería pensar cuál es el objetivo que se persigue. Así, si lo que se pretende es utilizar las estimaciones finales para ordenar los animales por su mérito genético, el criterio podría ser precisamente ese, es decir, que podría considerarse que se ha alcanzado la convergencia, cuando de una iteración a otra el orden de los animales no variase. Si lo que se desea es obtener una buena predicción de los méritos genéticos de los animales para poder estimar el progreso genético en las siguientes generaciones, sería deseable que los valores finales pudieran gozar de mayor precisión. Sin embargo, las varianzas obtenidas parecen alejarse tanto de su valor real que el hecho de afinar demasiado en el criterio de convergencia parece estar fuera de lugar.

El criterio de convergencia con el método Simplex,

algoritmo utilizado cuando se ajustó el modelo animal, no depende, sin embargo, de la variación de las estimaciones de una iteración a otra sino de la varianza del Simplex. En las figuras 20, 21 y 22 se aprecia que, al contrario que en el algoritmo EM, las estimaciones no llevan una tendencia aparente a partir de la primera iteración, sino que los valores intermedios oscilan irregularmente en torno al valor final. El criterio de convergencia seguido en este caso parece menos modificable que el utilizado con el algoritmo EM para el modelo padre.

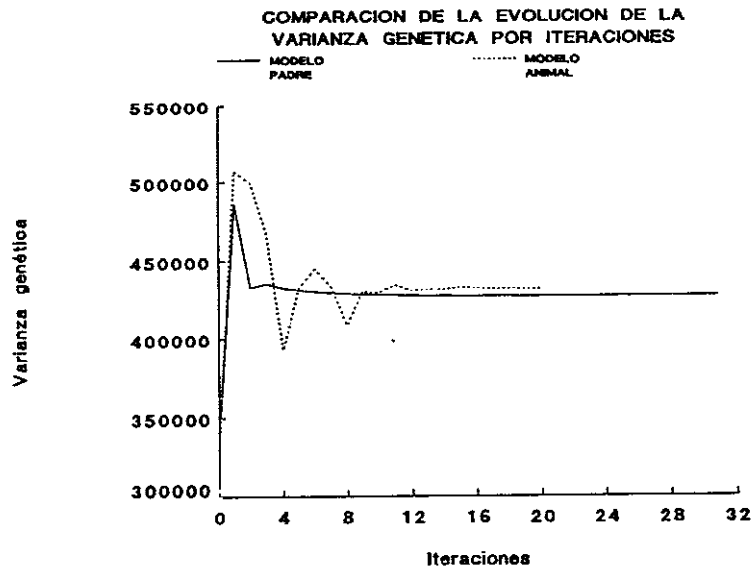


Figura 20.- Comparación (entre modelos) de la evolución de la varianza genética a través de las iteraciones

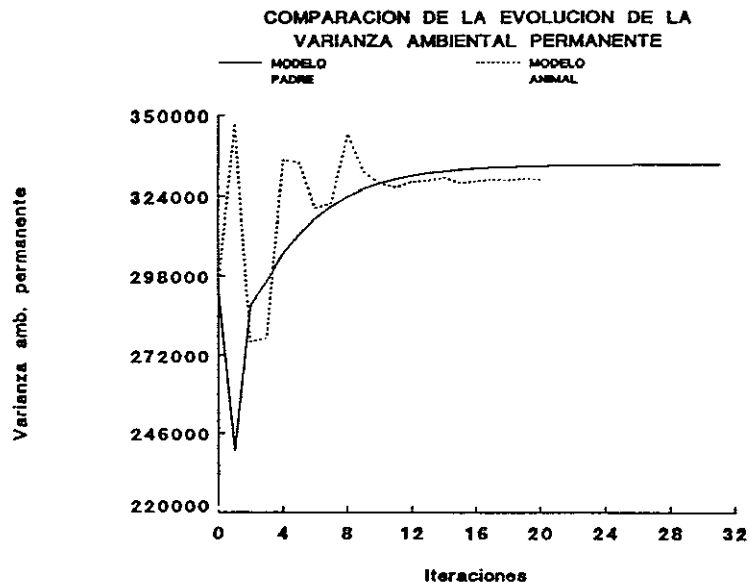


Figura 21.- Comparación (entre modelos) de la evolución de la varianza ambiental permanente a través de las iteraciones

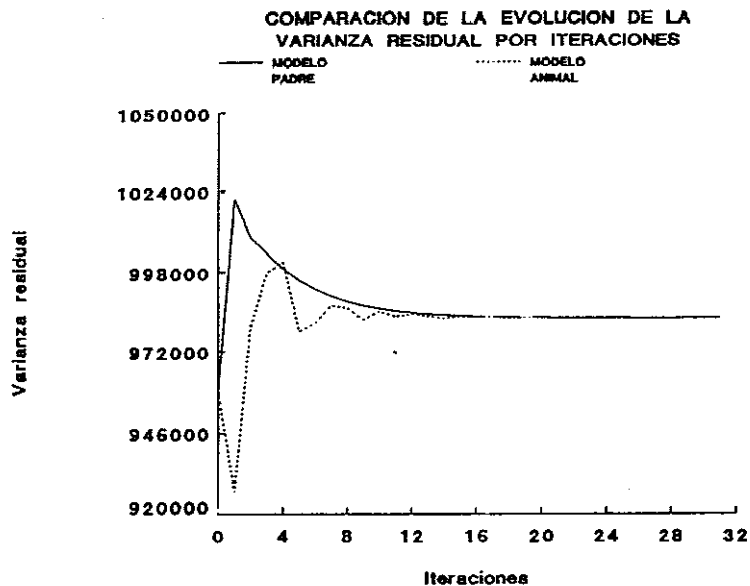


Figura 22.- Comparación (entre modelos) de la evolución de la varianza residual a través de las iteraciones

VII. CONCLUSIONES

Primera.- El coeficiente de endogamia medio resultó ser despreciable. Debido a la escasa información que aporta este parámetro parece fuera de lugar su inclusión en la matriz de relaciones aditivas.

Segunda.- La información aportada por la genealogía de los machos resultó ser tan escasa como para no ser tenida en cuenta en los análisis.

Tercera.- La tendencia genética observada resultó ser negativa, lo que pone de manifiesto la necesidad de monitorizar el control lechero con el fin de detectar posibles fallos en el programa de selección aplicado.

Cuarta.- La utilización de modelos más complejos, como puede ser el modelo animal, parece estar fuera de lugar en esta población en la que la ausencia de información y el desequilibrio son los factores limitantes.

Quinta.- La metodología seguida se mostró como un procedimiento robusto para proponer métodos de estimación de parámetros genéticos en cualquier población, antes de perderse en procedimientos complicados y onerosos que podrían resultar inútiles.

VIII. RESUMEN

Mediante el desarrollo de este trabajo se propuso una metodología a seguir para obtener estimaciones lo menos costosas posibles a la vez que suficientemente precisas, de los principales parámetros genéticos de interés en las poblaciones de animales domésticos.

Para este estudio se utilizó una muestra de ganado vacuno de raza Frisona.

En primer lugar se obtuvieron los coeficientes de endogamia de los animales. Los objetivos de esta primera parte eran, por un lado, conocer el valor de este parámetro en la población, y, por otro, saber si era necesario incluir su valor en la matriz de relaciones aditivas para realizar las estimaciones correspondientes en la segunda parte. El coeficiente de consanguinidad medio resultó ser de $4,03 \times 10^{-4}$.

Posteriormente se obtuvieron estimaciones de los componentes de varianza de la población en los caracteres producción de leche y producción de grasa en leche. El objeto de este análisis, además de obtener unas estimaciones para estos datos, era obtener valores de estas varianzas para poder ser incluidas en la población simulada. La heredabilidad obtenida para

ambos caracteres fue de 0,21 y 0,19 respectivamente.

Como un subproducto, la estimación de los distintos niveles del efecto fijo grupo genético permitieron dibujar una tendencia genética que resultó ser negativa.

Para comparar los diferentes modelos utilizables en la estimación de componentes de varianza, se simularon los efectos considerados normalmente como aleatorios en la población, utilizando las estimaciones anteriormente obtenidas. Se mantuvo, sin embargo, la propia estructura de los datos.

Se contrastó la eficacia de los modelos más utilizados comúnmente aplicándolos sobre tres réplicas. Los resultados mostraron un sesgo similar con todos los modelos. El parecido existente en las estimaciones encontradas entre un modelo padre y un modelo animal son razonables dada la escasa información acerca de las madres de las vacas. Las diferencias entre los valores obtenidos y los valores esperados parecen deberse a la falta de equilibrio en los datos y a la peculiar estructura de la población.

Finalmente, habiendo observado diferencias en la precisión de la estimación de la componente de varianza genética aditiva entre un carácter y otro, se realizó un tercer experimento. Se simuló el carácter producción de grasa en leche con tres heredabilidades diferentes de forma que el resto de los factores

que intervinieran en la simulación se mantuvieran constantes, incluidos los valores obtenidos aleatoriamente de una distribución normal de media cero y varianza uno. El sesgo se produjo en todos ellos en la misma magnitud, dirección y sentido por lo que se concluye que tales diferencias se debían exclusivamente al azar.

La falta de precisión observada en todos los casos, y el parecido entre las estimaciones obtenidas con todos los modelos, parecen indicar que, mientras no se mejore la estructura de los datos de esta población, el método de elección para realizar estimaciones de parámetros genéticos nuestra población será aquél que resulte computacionalmente más económico.

IX. BIBLIOGRAFIA

ABE, T.; NISHIDA, A. 1971. A method of calculating coefficients of inbreeding and relationship based on pedigree lines. Jap. Poultry Sci., 6: 245-249.

ALEDA, R; CARABAÑO, M. J.; GOMEZ, L.; JURADO, J. J.; RODRIGUEZ, M. C.; VILLANUEVA, B. 1986. Evolución genética de la población española. Jornadas sobre mejora genética del ganado vacuno de leche. pp. 45-52

ALFONSO PONCE, P. 1971. Redes genealógicas y computación de coeficientes de parentesco y consanguinidad. Arch. Zootec., 20: 257-259.

ALFONSO PONCE, P. 1985. Un método de cálculo de coeficientes de identidad genética sobre genealogías extensas. An. Vet. (Murcia), 1: 29-43.

ANDERSON, R. D.; HENDERSON, H. V.; PUKELSHEIM, F.; SEARLE, S. R. 1984. Best estimation of variance components from balanced data, with arbitrary kurtosis. Math. Operationforsch. u. Statistic. ser. statist., 15: 163-176.

BARKER, J.E.F.; ROBERTSON, A. 1966. Genetic and phenotypic parameters for the first three lactations in Friesian cows. Anim. Prod., 8: 221-240.

BEAUMONT, C. 1989. Restricted maximum likelihood estimation of genetic parameters for the first three lactations in the Montbeliarde dairy cattle breed. Génét. Sélect. Evol., 21: 493-506.

BEN GARA, A.; ALENDA, R.; ALONSO, A. 1991. Estima de los componentes de varianza con dos modelos diferentes en dos poblaciones de ganado frisón en España. Proc. IV Jornadas sobre producción animal. 565-567.

BERESKIN, B.; FREEMAN, A. E. 1965. Genetic and environmental factors in dairy sire evaluation. I. Effects of herds, months, and years-seasons on variance among lactation records; repeatability and heritability. J. Dairy Sci., 48: 347-351.

BOLDMAN, K. G.; FREEMAN, A. E. 1990. Adjustment for Heterogeneity of Variances by Herd Production Level in Dairy Cow and Sire Evaluation. J. Dairy Sci., 73: 503-512.

BOLDMAN, K. G.; VAN VLECK, L. D. 1991. Derivate-free Restricted Maximum Likelihood in Animal Models with a Sparse Matrix Solver. (En prensa).

- BOYCE, A. J. 1983. Computation of inbreeding and Kinship Coefficients on extended pedigrees. J. Hered., 74: 400-404.
- BUTRAM, S. T. 1987. Genetics of racing performance in the American Quarter Horse. Ph.D. Dissertation, Iowa State University, Ames, IA.
- CAÑON, J. 1989. Influencia de factores ambientales sistemáticos sobre la producción de leche en la población Frisona española. II. Cálculo de factores de ajuste aditivos y multiplicativos para la edad y mes de parto. ITEA, 82: 3-12.
- CAÑON, J.; BERGER, P. J.; GUTIERREZ, J. P.; MUÑOZ, A. 1989. Estimate of (Co)variance components for milk and fat yield in Spanish Holstein population using Reml. Arch. Zootec., 38: 249-254.
- CAÑON, J.; CHESNAIS, J. 1989. Indirect approach to simultaneous sire and cow evaluation using an animal model with repeated records and groups for unknown parents. Z. Tierzüchtg. u. Züchtbiol., 106: 335-344
- CAÑON, J.; GUTIERREZ, J. P. 1989. Influencia de la definición de los grupos genéticos en la clasificación de reproductores por el mérito genético. Proc. II Jornadas Producción animal, 379-381.

CAÑON, J.; MUÑOZ, A. 1991. Genetic trends for milk production in the Spanish Holstein population. J. Anim. Breed. Genet., 18: 41-47.

CARABAÑO, M. J.; VAN VLECK, L. D.; WIGGANS, G. R.; ALENDA R. 1989. Estimation of Genetic Parameters for Milk and Fat Yields of Dairy Cattle in Spain and the United States. J. Dairy Sci., 72: 3013-3022.

CARABAÑO, M. J.; WADE, K. M.; VAN VLECK, L. D. 1990. Genotype by environment interactions for milk and fat productions across regions of the United States. J. Dairy Sci., 73: 173-180.

COLACO, J. A.; FERNANDO, R. L.; GIANOLA, D. 1987. Variability in milk production among sires, herds and cows in Portuguese dairy cattle. In: 38th Annual meeting of the European Association for Animal Production, Lisboa, Portugal (Abstract).

CORBELL, R. R.; SEARLE, S. R. 1976. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. Technometrics, 18: 31-38.

CUE, R. I.; MONARDES, H. G.; HAYES, J. F. 1987. Correlations between production traits in first lactations Holstein cows. J. Dairy Sci., 70: 2132-2137.

CUNNINGHAM, E. P.; HENDERSON, C. R. 1968. An iterative procedure for estimating fixed effects and variance components in mixed model situations. Biometrics, 24: 13-21.

CRAUHAN, V.P.S.; HAYES, J. F. 1991. Genetic parameters for first lactation milk production and composition traits for Holsteins using multivariate restricted maximum likelihood. J. Dairy Sci., 74: 603-610.

DANELL, B. 1981. Interaction between genotype and environment in sire evaluation for milk production. Act. Agric. scan., 32: 33-46.

DE VEER, J. C.; VAN VLECK, L. D. 1987. Genetic parameters for first lactation milk yields at three levels of herd production. J. Dairy Sci., 70: 1434-1441.

DEMPFLE, L.; HAGGER, C.; SCHNEEBERGER, M. 1983. On the estimation of genetic parameters via variance components. Génét. Sél. Evol., 15: 425-444.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Royal Statist. Soc., B 39: 1-38.

DONG, M. C.; VAN VLECK L. D. 1989. Estimates of genetic and environmental (co)variances for first lactation milk yield, survival, and calving interval. J. Dairy Sci., 72: 678-684.

DUFF, I. S.; ERISHMAN, A. M.; REID, J. K. 1986. Direct methods for sparse matrices. Clarendon Press, Oxford, England.

EVERETT, R. W.; HENDERSON, C. R. 1972. The Northeast AI Sire Comparison. Why?. Animal Science Mimeograph Series, N° 19. Cornell University. N.Y.

EVERETT, R. W.; QUASS, R. L. 1979. Sire evaluation in the Northeast. Animal Science Mimeograph Series, N° 44. Cornell University. N.Y.

FALCONER, D. S. 1986. Introduction to Quantitative Genetics. 2nd. ed., 340 p., Longman, London.

GEORGE, A.; LIU, J.; NG, B. 1980. User guide for SPARSPAK: Waterloo sparse linear equations package. CS-78-30. Dept. Computer Science, Univ. Waterloo, Ontario, Canada.

GIANOLA, D.; FERNANDO, R. L. 1986. Batesian methods in animal breeding theory. J. Anim. Sci., 63: 217-244.

GRABER, H.-U.; SMITH, S.P.; TIER, B. 1987. A derivate-free approach for estimating variance components in animal models by Restricted Maximum Likelihood. J. Anim. Sci., 64: 1362-1370.

GUTIERRES, J. P.; CAÑON, J.; RICO, M. 1988. Cálculo del coeficiente de consanguinidad en poblaciones de animales domésticos. Actas I Congreso Monográfico Internacional "Mejora Genética del Ganado Porcino". Lorca (Murcia).

HARTLEY, H. O.; RAO, J. N. K. 1967. Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54: 93-108.

HARVEY, W. R. 1960. Least squares analysis of data with unequal subclass numbers. United States Department of Agriculture, ARS, 20-28.

HARVEY, W. R. 1977. User's guide for LMSL76. Mixed Model least squares and maximum likelihood computer program. United States Department of Agriculture, ARS.

HARVEY, W. R. 1987. User's guide for LMSL76. Mixed Model least squares and maximum likelihood computer program. United States Department of Agriculture, ARS.

HARVILLE, D. A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. J. An. Statist. Assoc., 72: 320-338.

HARVILLE, D. A.; CALLANAN, T. P. 1990. Computational aspects of likelihood-based inference for variance components. In: Prog. Int. Symposium on Advances in Statistical Methods for Genetic Improvement of Livestock (Gianola, D.; Hammond K., eds.). Springer Verlag, Heidelberg. 136-176.

HAZEL, L. N.; LUSH, J. L. 1950. Computing inbreeding and relationship coefficients from punched cards. J. Hered., 41: 301-306.

HENDERSON, C. R. 1953. Estimation of variance components. Biometrics, 9: 226-252.

HENDERSON, C. R. 1973. Sire evaluation and genetic trends. Prog. Anim. and Genet. Symp. in Honour of Dr. Jay L. Lush. ASAS-ADSA, Champaign, Illinois: 10-41.

HENDERSON, C. R. 1975. Comparison of alternative sire evaluation methods. J. Anim. Sci. 45: 760-770.

HENDERSON, C.R. 1980a. A simple method for unbiased estimation of variance components in the mixed model. Mimeo. Cornell University.

HENDERSON, C. R. 1980b. A simple method for unbiased estimation of variance components in the mixed model. 72nd Annual Meeting of the American Society of Animal Science, Champaign, Illinois. (Abstract)

HENDERSON, C. R. 1985. Equivalent linear models to reduce computations. J. Dairy Sci., 68: 2267-2277.

HENDERSON, C. R. 1986. Recent developments in variance and covariance estimation. J. Anim. Sci., 63: 208-216.

HENDERSON, C. R.; SEARLE, S. R.; SCHAEFFER. 1974. The invariance and calculation of method 2 for estimating variance components. Biometrics, 30: 583.

HENDERSON, C. R.; QUASS, R. L. 1976. Multiple trait evaluation using relative's records. J. Anim. Sci., 43: 1188-1197.

HILL, W. G. 1970. Design of experiments to estimate heritability by regression of offspring on selected parents. Biometrics, 26: 566-571.

HILL, W. G.; EDWARDS, M. L.; AHMED, K. A.; THOMPSON, R. 1983. Heritability of milk yield and composition at different levels and variability of production. Anim. Prod., 36: 59-68.

IBAÑEZ, M. A.; ALENDA, R.; CARABAÑO, M. J.; JURADO, J. J. 1991. Interacción genotipo medio ambiente, nivel de producción de leche y tamaño de los animales en la población frisona española. Proc. IV Jornadas sobre producción animal. 559-561.

JAGER, D. DE; KENNEDY, B. W. 1987. Genetic parameters of milk yield and composition and composition and their relationships with alternative breeding goals. J. Dairy Sci., 70: 1258-1266.

JENSEN, J.; MOORE, T. L.; MAO, I. L. 1990. Sampling of traits and data in estimation of genetic parameters. In: Proc. 4th World Congress on Genetics applied to Livestock Production. Plenary lectures, molecular genetics and mapping, selection, prediction and estimation. Edinburgh. UK. 281-284.

JUGA, J.; THOMPSON, R. 1989. Estimation of variance components in populations selected over multiple generations. Act. Agric. Scan., 39: 79-89.

KNUTH, D. E. 1973. The Art of Computer Programming Vol 1. Fundamental Algorithms, 2nd edition. Addison-Wesley Pub. Co., Reading, Massachusetts.

LAIRD, N. M.; WARE, J. H. 1982. Random effects models for longitudinal data. Biometrics, 38: 963-974.

LANOTTE, L. R. 1970. A class of estimators of variance components. Technical Report N° 10. Department of Statistics, University of Kentucky, Lexington.

LANOTTE, L. R. 1971. Locally best quadratic estimators of variance components. Technical Report N° 22. Department of Statistics, University of Kentucky, Lexington.

LAWLOR, T. J.; POLLACK, E. J.; QUAAS, R. L. 1984. Estimation of variance components with relationships included for a multiple trait model. J. Dairy Sci., 67 (Suppl.1), 181 (Abstract P.204)

LEE, K. J.; YANG, Y. M.; YANG, H. S. 1989. Estimation of dairy sire's breeding values by mixed model procedures with additive relationships and genetic grouping. Korean Journal of Animal Sciences., 31: 691-698.

LIN, C. Y.; MCALLISTER, A. J. 1984. Monte Carlo comparison of four methods for estimation of genetic parameters in the univariate case. J. Dairy Sci., 67: 2389-2398.

LUSH, J. L.; SHRODE, R. R. 1950. Changes in milk productions with age and milking frequency. J. Dairy Sci., 33: 338-357.

MANFREDI, E. J.; EVERETT, R. W.; SEARLE, S. R. 1984. Phenotypic and genetics statistics of components of milk and two measures of somatic cell concentrations. J. Dairy Sci., 67: 2028-2033.

MENENDEZ, A.; REYES, A.; GUERRA, D.; CORDOVI, J. 1989. Genetic variability of milk production from Holstein cows according to the level of milk yield of the herd. Cuban Journal of Agricultural Science., 23: 9-15.

MEINERT, T. R.; KROVER, S.; VAN ARENDOK, J. A. M. 1989. Parameter estimation of milk yield and composition for 305 days and peak production. J. Dairy Sci., 72: 1534-1539.

MEYER, K. 1983. Maximum likelihood procedures for estimating genetic parameters for later lactations of dairy cattle. J. Dairy Sci., 66: 1988-1997.

MEYER, K. 1984. Estimates of genetic parameters for milk and fat yield for the first three lactations in British Friesian cows. Anim. Prod., 38: 313-322.

MEYER, K. 1985a. Genetic parameters for dairy production of Australian Black and White cows. Livest. Prod. Sci., 12: 205-219.

MEYER, K. 1985b. Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. Biometrics, 41: 153-165.

MEYER, K. 1986. Between algorithms: A 'short cut' restricted maximum likelihood procedure to estimate variance components. J. Dairy Sci., 69: 1904-1916.

MEYER, K. 1987. Restricted maximum likelihood to estimate variance components for mixed models for mixed models with two random factors. Génét. Sélect. Evol., 19: 49-68.

MEYER, K. 1988a. DFREML-a set of programs to estimate variance components under an individual animal model. J. Dairy Sci., 71 (suppl. 2): 33-34.

MEYER, K. 1988b. DFREML. Programs to estimate variance components for individual animal models by restricted maximum likelihood. User notes. Univ. of Edinburgh.

MEYER, K. 1989. Restricted Maximum Likelihood to Estimate Variance Components for Animal Models with Several Random Effects Using a Derivate-free Algorithm. Génét. Sélect. Evol., 21: 317-340.

MISZTAL, I.; DA, Y. 1990. Property of trace in restricted maximum likelihood estimation of variance components. J. Dairy Sci., 73: 3583-3585.

MOKHTAR IBRAHIM, S.A. 1979. Interactions between environmental intensity and genetic potential Holstein-Friesian crosses in the Friesian population of Lower Saxony. Ph. D. Thesis. Univ. Gottingen.

NADOT, R. 1971. Les coefficients d'identité. Une méthode de calcul. In Génétique et Populations. Hommage à Jean Sutter. Presses Universitaires de France. Paris, p. 111-117.

NADOT, R.; VAYSSEIX, G. 1973. Apparentement et identité. Biometrics, 29: 347-359.

NELDER, J. A.; MEAD, R. 1965. A simplex method for function minimization. Computer J., 7: 147-151.

O'FERRALL, G. J. M. 1990. Phenotypic and genetic parameters of production traits in Irish Friesian cows. Irish Journal of Agricultural Research, 29: 95-100.

OUWELTJES, W. R.; SCHAEFFER, L. R.; KENNEDY, B. W. 1988. Sensitivity of methods of variance component estimation to culling type of selection. J. Dairy Sci., 71: 773-779.

PATTERSON, H. D.; THOMPSON, R. 1971. Recovery of interblock information when block sizes are unequal. Biometrika, 58: 545-554.

POGREBNIYAK, V. A. 1990. The dynamics of inbreeding in Russian Black Pied cattle. Zootekhnika, 4: 24-26.

POWELL, L. R.; NORMAN, H. D.; DICKINSON, F. W. 1977. Trends in breeding value and production. J. Dairy Sci., 60: 1316-1326.

PRESS, W. H.; FLANNERY, B. P.; TEUKOLSKY, S. A.; VETTERLING, W. T. 1989. Numerical recipes. Cambridge Univ. Press, Cambridge, MA.

QUAAS, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics, 32: 949-953.

QUAAS, R. L.; EVERETT, R. W.; MCCLINTOCK, A. C. 1979. Maternal grandsire model for dairy sire evaluation. J. Dairy Sci., 62: 1648-1654.

RAO, C. R. 1971. Estimation of variance and covariance components. MINQUE theory. J. Multivariate Analysis, 1: 257-275.

RENFELD, C. E.; BACUS, J. W.; PAGELS, J. A.; DIPTER, M. H. 1967. Computer calculation of Wright's inbreeding coefficient. J. Hered., 58: 81-84.

ROBINSON, J. A. B. 1988. Multiple lactation variance components estimation using restricted maximum likelihood. Dissertation Abstracts International, B (Sciences and Engineering) 49: 2007. Cornell University, Ithaca, NY, USA.

ROTHSCHILD, M. F.; HENDERSON, C. R. 1979. Maximum likelihood estimates of parameters of first and second lactation milk records. J. Dairy Sci., 62: 990-995.

RÖNNINGEN, K. 1974. Monte Carlo Simulation of Statistical-Biological Models which are of Interest in Animal Breeding. Act. Agric. Scan., 24: 135-142.

SCHAAP, T.; COHEN, M. M. 1977. A simple non graphic method for pedigree description and analysis. J. Med. Genet., 14: 25-29.

SEARLE, S. R. 1968. Another look at Henderson's methods of estimating variance components. Biometrics, 24: 749-787.

SEARLE, S. R. 1979. Notes on variance component estimation: A detailed account of Maximum Likelihood and kindred methodology. Paper BU-673M. Biometrics Unit. Cornell University, Ithaca, NY.

SCHAEFFER, L. R. 1986. Pseudo expectation approach to variance component estimation. J. Dairy Sci., 69: 2884-2889.

SHORT, T. H.; BLAKE, R. W.; QUASS, R. L.; VAN VLECK, L. D. 1990. Heterogeneous within-herd variance. 1. Genetic parameters for first and second lactation milk yields of grade Holstein cows. J. Dairy Sci., 73: 3312-3320.

SHUTZ, M. M.; HANSEN, L. B.; STEURNAGEL, G. R.; RENEAU, J. K.; KUCK, A. L. 1990. Genetic parameters for somatic cells, protein and fat in milk of Holsteins. J. Dairy Sci., 73: 494-502.

SIMIANER, H. 1986. Restricted maximum likelihood estimation of variances and covariances from selected data. In: Proc. 3rd World Congress on Genetics Applied to Livestock Production, Lincoln, Nebraska, July 16-22, 1986 (Dickerson G. E. and Johnson R. K., ed.), University of Nebraska, U.S.A., pp. 421-426.

SMITH, S. P.; GRASER, H.-U. 1986. Estimating Variance Components in a class of mixed models by Restricted Maximum Likelihood. J. Dairy Sci., 69: 1156-1165.

SMITH, S. P.; LIN, C. Y. 1989. Efficient implementation of the new restricted maximum likelihood algorithms. J. Dairy Sci., 72: 3336-3341.

SPENDLEY, W.; HEIT, G. R.; HINSWORTH, F. R. 1962. Sequential application of simplex designs in optimization and evolutionary operation. Technometrics, 4: 441-461.

STEVENS, A. 1975. An elementary computer algorithm for the calculation of coefficient of inbreeding. Inf. Processing Lett., 3: 153-163.

SWALVE, E.; VAN VLECK, L. D. 1987. Estimation of genetic (co)variances for milk yield in first three lactations using an animal model and restricted maximum likelihood. J. Dairy Sci., 70: 842-849.

THOMPSON, R. 1969. Iterative estimation of variance components for non-orthogonal data. Biometrics, 25: 767-780.

THOMPSON, R. 1976. Design of experiments to estimate heritability when observations are available on parents and offspring. Biometrics, 32: 283-304.

THOMPSON, R. 1977. The estimation of heritability with unbalanced data. I. Observations available on parents and offspring. Biometrics, 33: 485-495.

THOMPSON, R.; MEYER, K. 1986. Estimation of variance components: What is missing in the EM algorithm? J. Statist. Comput. Simul., 24: 215-230.

TIER, B.; SMITH, S. P. 1988. Use of sparse matrix absorption in animal breeding. Génét. Sél. Evol., 21: 457-466.

TONG, A. K. W.; KENNEDY, B. W.; MOXLEY, J. E. 1979. Heritabilities and genetic correlations for the first 3 lactations from records subject to culling. J. Dairy Sci., 62: 1784-1790.

VAN DER WERF, J. H. J.; DE BOER, W. 1989a. Estimation of genetic parameters in a crossbred population of Black and White dairy cattle. J. Dairy Sci., 72: 2615-2623.

VAN DER WERF, J. H. J.; DE BOER, W. 1989b. Influence of nonadditive effects on estimations of genetic parameters in dairy cattle. J. Dairy Sci., 72: 2606-2614.

VAN RADEN, P. M.; FREEMAN, A. E. 1985. Obtaining bounds on REML estimates of variance components. J. Dairy Sci., 68 (Suppl. 1): 213. (Abstr.)

VAN RADEN, P. M.; FREEMAN, A. E. 1987. Rates of convergence of REML algorithms. Iowa Agric. and Home Econ. Exp. Stn., Journal Paper N° J-12584.

VAN RADEN, P. M.; JUNG, Y. C. 1988. A general purpose approximation to restricted maximum likelihood: The tilde-hat approach. J. Dairy Sci., 71: 187-194.

VAN VLECK, L. D.; DONG, M. C. 1988. Genetic (co)variances for milk, fat, and protein yield in Holsteins using an animal model. J. Dairy Sci., 71: 3040-3046.

VAN VLECK, L. D.; HENDERSON, C. R. 1961. Estimates of genetic parameters of some functions of part lactations milk records. J. Dairy Sci., 44: 1073-1084.

VAN VLECK, L. D.; WESTELL, R. A.; SCHNEIDER, J. C. 1986. Genetic change in milk yield estimated from simultaneous genetic evaluation of bulls and cows. J. Dairy Sci., 69: 2963-2965.

VAST-E USER'S GUIDE. 1988. Edition 1.4. Pacific-Sierra Corp., Los Angeles, CA.

WADE, K. M.; VAN VLECK, L. D. 1989. Genetic parameters for production traits of holsteins in California, New York and Wisconsin. J. Dairy Sci., 72: 1280-1285.

WESTELL, R. A. 1984. Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. Ph. D. Thesis. Cornell University.

WESTELL, R. A.; QUASS, R. L.; VAN VLECK, L. D. 1988. Genetics groups in Animal Model. J. Dairy Sci., 71: 1310-1318.

WRIGHT, S. 1922. Coefficients of inbreeding and relationship. Am. Nat., 56: 330-338.

