ITEA (2000), Vol. **96A N.° 3,** 318-336

EVALUACIÓN DE MODELOS ALTERNATIVOS PARA EL ANÁLISIS DE DATOS DE PRODUCCIÓN DE LECHE A LO LARGO DE LA LACTACIÓN

P. López-Romero, M.J. Carabaño

Departamento de Mejora Genética y Biotecnología, INIA, Carretera de La Coruña km 7, 28040 Madrid, España

RESUMEN

Se han estimado parámetros genéticos para producción de leche en primeras lactaciones de ganado frisón español a través de modelos basados en el uso directo de las producciones en el día del control. Se han utilizado distintos modelos alternativos que fueron comparados en términos de bondad de ajuste y capacidad de predicción usando validación cruzada. Los modelos evaluados fueron un modelo de repetibilidad (MMR) y distintos modelos de regresión aleatoria (MRA), que incluyeron funciones lactacionales y polinomios de Legendre como funciones lineales para definir las regresiones aleatorias. Los MRA fueron superiores al MMR. Dentro de los MRA, aquellos modelos que ajustaron polinomios de Legendre en las regresiones aleatorias fueron mejores que los que ajustaron funciones lactacionales, presentado el inconveniente de modelizar mal los parámetros genéticos a partir de órdenes de ajuste superiores a cuatro para el efecto aditivo. Este problema no se presentó en órdenes elevados de ajuste sobre la parte permanente, habiéndose observado además que incrementos de orden del polinomio sobre el efecto permanente producían mejoras más relevantes que si el aumento de orden se producía en el efecto aditivo.

Palabras clave: Parámetros genéticos, Producciones en el día del control, Modelos de regresión aleatoria, Polinomios de Legendre, Comparación de modelos, Validación cruzada.

SUMMARY

EVALUATION OF ALTERNATIVE MODELS FOR ANALYZING MILK YIELD DATA ACROSS LACTATION

Genetic parameters for milk yield of first lactation Spanish Hosltein cattle were estimated using test day models. Different alternative models were employed and were compared in terms of goodness of fit and predictive ability using cross validation techniques. Models used were repeatilbility model (MMR) and different random regression models (MRA), using Wilmink function, Ali-Schaeffer function and Legendre polynomials. The MRA were better than MMR. Within the MRA, those models which fitted Legendre polynomials were better than those which fitted lactational funtions. Moreover, the MRA had the disadvantage that they did not describe well the genetic parameters at the onset and at the end of lactation when the order of fit for additive effect was greater than four. This problem was not observed when the high orders of fit for permanent effect were used. Furthemore, when the order of fit for permanent effect increased, the improving of the model was better than when the increasing of order was carried out for the additive effect.

Key words: Genetic parameters, Test day yields, Random regression models, Legendre polynomials, Model comparison, Cross validation.

Introducción

La utilización de los datos de los controles mensuales, a partir de ahora, producciones en el día del control (PDC), para la valoración de caracteres de producción de leche en ganado vacuno presenta ventajas en relación al uso del dato resumen de producción en 305 días (ver, p.ej. REKAYA y col., 1999). En contrapartida, la modelización de este tipo de información es, potencialmente, más compleja y los costes computacionales asociados a la aplicación de estos modelos se incrementa de forma muy importante, pudiendo ser limitante en algunos casos.

La forma de modelizar las PDC puede variar entre el modelo más sencillo que considera los datos de controles sucesivos como medidas repetidas del mismo carácter (dentro y/o entre lactaciones) hasta un modelo multicarácter de orden hipotéticamente infinito o muy elevado que considerase la producción en cada momento de la lactación como un carácter diferente. Existen numerosos estudios que han estudiado la forma de modelizar estos datos en vacuno lechero. PTAK y SCHAEFFER (1993) usaron un modelo de repetibilidad simple considerando que las producciones observadas en controles sucesivos de un mismo animal como medidas repetidas de un mismo carácter. REENTS y col. (1995a y b) desarrollaron un modelo que considera los datos de controles dentro de una misma lactación como medidas repetidas y como caracteres diferentes cuando se trata de datos de distintas lactaciones, siendo este el modelo usado actualmente en las valoraciones oficiales basadas en los datos de controles individuales en Alemania.

Un refinamiento con respecto al modelo de repetibilidad simple son los modelos de regresión aleatoria (MRA) (SCHAEFFER y DEKKERS, 1994). En este tipo de modelos se reconoce la posibilidad de ajustar curvas lactacionales diferentes para cada animal, tanto en relación a efectos sistemáticos como a los efectos genéticos, reconociendo, por ejemplo, la existencia de diferencias en el potencial genético para factores como la persistencia de las curvas de lactación. Además, estos modelos reconocen la existencia de (co)varianzas que cambian con el tiempo, tanto a nivel del componente genético, como de otros componentes ambientales no sistemáticos. Un modelo de regresión aleatoria es el empleado en la evaluación oficial de Canadá (JAMROZIK y col., 1997).

Previo a la introducción de modelos de regresión aleatoria para modelizar las PDC a lo largo de la lactación, KIRKPATRICK y col. (1990) propusieron el uso de polinomios ortogonales (polinomios de Legendre) para ajustar funciones de covarianza que van a definir las (co)varianzas entre observaciones tomadas a lo largo del tiempo. Este tipo de modelos se conoce como modelos de funciones de covarianza (MFC). MEYER y HILL (1997) mostraron que un MRA puede ser también usado para modelizar las (co)varianzas entre observaciones. VAN DER WERF y col. (1999) demostraron que para cada MFC existe un MRA equivalente. De esta manera, se puede considerar como MRA tanto aquellos que ajustan funciones lactacionales a los efectos aleatorios como los que usan polinomios de Legendre para ajustar funciones de covarianza.

Paralelamente al desarrollo de MRA bajo curvas lactacionales y de MFC, WIGGANS y GODDARD (1997) desarrollan un procedimiento alternativo orientado también a reducir la elevada dimensionalidad del modelo multicarácter. Estos autores proponen el uso de la transformación canónica sobre la matriz de (co)varianzas para el posterior análisis de un número reducido de las resultantes variables canónicas. VAN DER WERF y col. (1999) indican que esta aproximación puede

interpretarse como una versión discontinua de los MFC propuestos por KIRKPATRICK y col. (1990).

Más recientemente, PLETCHER y GEYER (1999) sugieren el uso de métodos paramétricos, basados en la utilización de funciones de distribución en lugar de los métodos no paramétricos como es el de los polinomios de Legendre como funciones de covarianza. Estos autores indican que de esta forma es posible modelizar de forma más acorde con la biología del proceso las covarianzas, reduciendo también el número de parámetros, frente al uso de polinomios que son funciones totalmente arbitrarias, con un comportamiento excesivamente oscilante cuando el grado del polinomio es alto y poco acorde con la naturaleza gradual y con límites asintóticos de los procesos biológicos. Sin embargo, no es obvio cual sería la función de distribución adecuada y estos autores muestran que una mala elección de esta función puede conducir a resultados mediocres en el ajuste de las verdaderas covarianzas.

Los MRA que usan funciones lactacionales y aquellos que usan los polinomios de Legendre para ajustar las funciones de covarianza son los más utilizados en la modelización de las PDC. Además de las ventajas teóricas de estos modelos en cuanto a flexibilidad en la modelización del mérito genético y de la estructura de covarianzas, su ventaja sobre los modelos de repetibilidad simple ha quedado reflejada en varios estudios, en cuanto a una reducción sobre la varianza residual del modelo (JAMROZIK y col., 1996 y REKAYA y col., 1999). Sin embargo, no está claro cuál es la función a utilizar y tanto los modelos lactacionales como los que usan polinomios de Legendre han mostrado comportamientos anómalos en los extremos de la lactación. Por otra parte, los trabajos que han abordado la comparación de modelos alternativos (JAMROZIK y col, 1997; OLORI y col, 1999a; VAN DER WERF, 1998) han basado la misma en la utilización de tests estadísticos (principalmente tests χ^2) y en el comportamiento de las varianzas y correlaciones para distintos momentos de la lactación, tomando como referencia los valores del modelo multicarácter. No se han considerado, sin embargo, criterios como la capacidad de ajuste de los datos o la capacidad de predicción de datos futuros.

En este trabajo se presentan resultados que permiten la comparación de funciones lactacionales y polinómicas en la modelización de las PDC en nuestra población de ganado frisón, intentando abordar esta comparación desde una perspectiva más amplia que la llevada a cabo en estudios previos sobre esta y otras poblaciones de ganado vacuno lechero.

Material y métodos

Se han utilizado producciones en el día del control (PDC) de primeras lactaciones procedentes de Navarra a partir de datos suministrados por CONAFE. Se han seleccionado únicamente aquellos rebaños con registros en todos los años del intervalo comprendido entre 1988 y 1998, además, de estos rebaños, sólo fueron tenidos en cuenta los datos pertenecientes a ese periodo. El motivo de esta selección fue debido a la necesidad de reducir la base de datos a un número computacionalmente manejable, manteniendo una estructura adecuada para el análisis que se iba a llevar a cabo, en cuanto a conexiones genealógicas, número de PDC por animal y tamaño de los niveles del grupo de comparación Rebaño-Fecha de control (RFC). Para mantener todos las PDC de los animales, manteniendo simultánea-

mente tamaños de RFC con un número de observaciones suficientes, se tuvieron que muestrear rebaños completos. Además, para mantener conexiones genealógicas entre los animales de la base de datos reducida que iba a ser analizada, los rebaños tuvieron que ser muestreados de una única región, escogiéndose además sólo aquellas lactaciones pertenecientes a un periodo de tiempo determinado. Tras esta primera selección se obtiene un archivo con 65 rebaños que es sometido a un proceso de depuración, en el que se exige una edad al parto entre 18 y 40 meses, un intervalo entre el parto y el primer control entre 5 y 67 días y un intervalo entre controles sucesivos entre 26 y 67 días. También se exige que todas las PDC estén comprendidas entre los 4 y 99 kg, y que se hayan registrado entre los 5 y 335 días de la lactación. Tras este proceso se obtuvo una muestra de 150.291 PDC correspondientes a 16.166 lactaciones, con 6.258 niveles del grupo RFC que mantenían 5 o más observaciones en cada uno de sus niveles. La duración media de la lactación fue de 291,83 ± 36,37 días. Sobre esta base de datos se muestrean aleatoriamente rebaños completos para obtener una muestra que será la base del análisis. Posteriormente se eliminan "outliers" y se consolida la muestra, exigiéndose que tenga un número mínimo de 6 PDC por animal, a la vez que se mantienen tamaños mínimos de 5 observaciones por nivel de RFC. Tras este proceso queda una muestra cuyas características se resumen en el cuadro 1.

Una de las técnicas de comparación de modelos que se van a utilizar es la validación cruzada (VC). Este método consiste en partir la muestra en dos subconjuntos. El primero es utilizado para estimar los parámetros del modelo, utilizando posteriormente estas estimas para predecir las PDC del segundo subconjunto. De esta forma podemos estudiar la capacidad predictiva del modelo. La partición se realiza de forma aleatoria bajo ciertas restricciones, debido a que para realizar las predicciones sobre los elementos del segundo subconjunto (PRE), es necesario que todos los niveles de lo efectos aquí contenidos lo estén también en el primer subconjunto (EST). Se realiza por tanto un muestreo aleatorio de 2 PDC por animal para formar el conjunto PRE, exigiéndose que todos los niveles de PRE estén a su vez en EST. Adicionalmente también se exige que los niveles de los RFC del conjunto EST tengan un tamaño mínimo de 3 observaciones. El número de PDC obtenidas tras el proceso de partición de la muestra para VC fue de 35.614 PDC para el conjunto EST con un 78,89% de las PDC contenidas en la muestra original, y de 9.536 PDC en el conjunto PRE co un 21,11% de los datos iniciales. En el

Cuadro 1
Resumen de datos de la muestra de datos analizada
Table 1
Summary of the data analysed

Rebaños	22
Lactaciones (animales en dato)	4.768
Numero PDC (≥ 6 PDC /animal)	45.150
Numero niveles RFC (≥ 5 PDC/nivel)	1.853
Animales en Genealogía	8.415

cuadro 2 se muestran las producciones medias y desviaciones estándar para las PDC y días en lactación (DEL) para el conjunto EST. Todos los parámetros de todos los modelos han sido estimados con la muestra de datos del conjunto EST, utilizándose únicamente los datos de los conjuntos PRE para la metodología de VC.

Se llevó a cabo la estimación de parámetros genéticos para PDC bajo distintos modelos de regresión aleatoria (MRA) y un modelo animal con medidas repetidas (MMR). El distinto comportamiento de los modelos fue evaluado en términos de la bondad de ajuste y de la capacidad de predicción. Las comparaciones también estuvieron basadas, en el caso de los MRA, en la habilidad de cada MRA para modelizar las estructuras de varianzas de los efectos aleatorios a lo largo de la lactación. Aunque es bien conocido el hecho de que son mejores aquellos modelos que permiten heterogeneidad de la varianza residual a lo largo de la lactación (OLORI y col., 1999a; OLORI y col., 1999b), hemos trabajado con varianza residual constante debido a que el programa utilizado en la estimación de los componentes de varianza no permite de momento la heterogeneidad en la varianza residual (MISZTAL, 2000). OLORI y col (1999a) afirman que mantener homogeneidad en la varianza residual a lo largo de la lactación, produciría subestimaciones de la varianza residual al principio y final de la lactación sin afectar de modo significativo a la estima del resto de componentes, por lo que pensamos que las conclusiones que se puedan extraer del presente trabajo son perfectamente válidas.

Modelos

Todos los modelos utilizados incluyen en su ecuación los mismos efectos fijos. Las

Cuadro 2
Media y desviaciónes estándar (paréntesis) para los días en lactación (DEL) y kg. de leche del conjunto EST
Table 2
Averages and standard deviations (in parenthesis) for days in milk (DEL) and milk yield of EST subset

Periodo	Intervalo (días)	DEL medio	Producción media (kg)	Número PDC
1	5-35	20,33 (8,83)	25,18 (6,21)	3.428
2	36-65	50,23 (8,68)	27,43 (6,49)	3.492
3	66-95	80,46 (8,73)	27,15 (6,66)	3.494
4	96-125	110,42 (8,61)	26,70 (6,69)	3.432
5	126-155	140,27 (8,55)	25,70 (6,73)	3.548
6	156-185	170,17 (8,66)	24,77 (6,72)	3.563
7	186-215	200,36 (8,61)	24,08 (6,76)	3.461
8	216-245	230,43 (8,64)	23,02 (6,59)	3.415
9	246-275	260,18 (8,55)	22,08 (6,52)	3.196
10	276-305	289,49 (8,57)	21,33 (6,50)	2.559
11	306-335	318,82 (8,64)	20,49 (6,40)	1.986

diferencias entre modelos se deben únicamente a si se utilizan o no regresiones aleatorias y en cuanto a las funciones que se han utilizado para definir esas regresiones.

Uno de los efectos fijos que vamos a tener en cuenta es el punto de la lactación en el que se ha producido el dato, mediante la inclusión de covariables que definen la forma general de la curva de lactación dentro del grupo de comparación edad-época de parto (EE). Para definir estas covariables, vamos a utilizar basándonos en análisis preliminares, la función lineal de ALI-SCHAEFFER de 5 covariables (ALI-SCHAEFFER, 1987), que define la producción de leche en el día *t* mediante la ecuación:

$$y(t) = b_0 + b_1\, t/305 + b_2 ln(t/305)^2 + \\ + b_3 ln(305/t) + b_4\, (ln(305/t))^2 \qquad [1]$$

donde:

y(t) producción en el día *t*.

b_i coeficientes de regresión estimados con el modelo. Dividir por la constante 305 no es estrictamente necesario pero reduce la magnitud de la covariable.

Teniendo esto en cuenta, los efectos fijos se escribirán para todos los modelos como:

$$EF_{ij} = RFC_i + EE_j + \sum_{m=1}^{4} \beta_{jm} X_{klm} \qquad [2]$$

donde:

EF_{ij} representa los efectos fijos (comunes a todos los modelos).

RFC_i es el grupo de comparación rebaño-fecha de control *i*.

EE_j es el grupo de comparación edad-época de parto *j*.

β_{jm} es el coeficiente de regresión fijo *m* anidado a EE_j.

X_{klm} es la covariable m de la función lineal de Ali-Schaeffer [1] definida para la PDC_l del animal *k*.

Modelo de repetibilidad

La ecuación del modelo es la siguiente:

$$y_{ijkl} = EF_{ij} + a_k + p_k + e_{ijkl} \qquad [3]$$

donde:

y_{ijkl} es la PDC_l del animal *k*, perteneciente al grupo RFC_i y cuya curva de lactación fija esta anidada al grupo EE_j.

EF_{ij} son los efectos fijos definidos en (2).

a_k es el aditivo del animal *k*.

p_k es el efecto ambiental permanente asociado a todas las *PDC* de la vaca *k*.

e_{ijkl} es el efecto ambiental transitorio de la PDC_l.

En notación matricial, **y** será el vector que contiene las PDC observadas en todos los animales, ***b***, ***a***, ***p*** y ***e*** son vectores que contienen los efectos ambientales (RFC_i, EE_j y los coeficiente de regresión fijos $\beta_{jm\ (m = 1,4)}$ asociados a la función de ALI-SCHAEFFER), aditivo, permanente y residual. ***X***, ***Z*** y ***W*** son matrices de incidencia, incluyéndose en *X* las covariables $x_{klm\ (m = 1,4)}$ de la función de ALI-SCHAEFFER [1].

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wp} + \mathbf{e} \qquad [4]$$

Se asume la siguiente estructura de (co)varianzas:

$$\text{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{P} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{E}\end{bmatrix}$$

donde **G** es la matriz de (co)varianzas genéticas entre los animales, definida como $\mathbf{G} = \sigma^2_a\mathbf{A}$, siendo σ^2_a la varianza genética aditiva y **A** la matriz de parentesco. **P** es una matriz diagonal que contiene la varianza del efecto permanente, $\mathbf{P} = \sigma^2_p\mathbf{I}$ y **E** es una matriz diagonal que contiene la varianza residual σ^2_e.

Modelos de regresión aleatoria

Los MRA utilizados se pueden agrupar en modelos que ajustan coeficientes de regresión aleatorios sobre covariables definidas a partir de funciones lineales que describen la curva de lactación, y en modelos que ajustan coeficientes sobre covariables definidas a través de una combinación lineal de polinomios ortogonales (polinomios de Legendre). Estos últimos surgen del procedimiento de obtención de funciones de covarianza (FC) como funciones reales de dos variables, a través de bases definidas por un conjunto de polinomios ortogonales dentro de un espacio funcional. (KIRKPATRICK y HECKMAN 1989; KIRKPATRICK y BULMER, 1990; APOSTOL, 1973)

La ecuación general del modelo es:

$$y_{ijkl} = EF_{ij} + a_k + \sum_{m=1}^{r} \alpha_{km} Z_{1,klm} + p_k + \sum_{m=1}^{s} \omega_{km} Z_{2,klm} + e_{ijkl} \quad [5]$$

donde a_k y p_k se definen igual que en [3] y α_{km} coeficiente *m* de regresión aleatorio anidado al animal *k*. Obsérvese que *m* puede variar desde *1* hasta *r*, siendo r el número de términos dependientes del tiempo de la función ajustada a la parte aditiva. Considerando el termino independiente, el número de coeficientes de regresión estimados, u *orden de ajuste*, será $r + 1$.

$Z_{1,klm}$ covariable m definida por la función ajustada sobre el efecto aditivo

ω_{km} coeficiente de regresión aleatorio *m* anidado a p_k. El orden ajustado sobre el efecto permanente será $s + 1$.

$Z_{2,klm}$ covariable *m* definida por la función que se ajusta al efecto permanente.

Siguiendo la misma notación que en [4] para el MMR, un MRA se puede escribir en forma matricial como:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_x \mathbf{a} + \mathbf{W}_x \mathbf{p} + \mathbf{e} \quad [6]$$

con ***y***, ***b***, ***e*** y ***X*** definidos como en [4], siendo a el vector que contiene los $r+1$ coeficientes de regresión aleatorios ($a_k + \alpha_{km\ (m=1,r)}$). $\mathbf{Z}_x$ es la matriz de incidencia que relaciona las PDC con el animal e incluye las covariables $Z_{1,klm\ (m\ =\ 1,r)}$ correspondientes a la función lineal ajustada al efecto aditivo. El vector p incluye los $s + 1$ coeficientes de regresión aleatorios para el efecto permanente ($p_k + \varpi_{km\ (m\ =\ 1,s)}$) y la matriz $\boldsymbol{W}_x$ es la matriz de incidencia del efecto permanente que incluye también las covariables $Z_{2,klm\ (m\ =\ 1,s)}$ que se definen para modelizar este efecto a lo largo de la lactación. La estructura de (co)varianzas asumidas en los MRA es:

$$\text{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{P} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{E}\end{bmatrix}$$

donde **G** se define como $\mathbf{G} = \Sigma_a \otimes \mathbf{A}$, siendo Σ_a una matriz que contiene las (co)varianzas genéticas entre los $r+1$ coeficientes de regresión aleatorios que se ajustan sobre la parte aditiva, y **A** la matriz de parentesco. P se define como $\mathbf{P} = \Sigma_p \otimes \mathbf{I}$, donde Σ_p es una matriz que contiene la (co)varianza ambiental permanente entre los $s+1$ coeficientes de regresión aleatorios ajustados sobre el efecto permanente. **E** es una matriz diagonal que contiene la varianza residual σ^2_e, asumida constante a lo largo de toda la lactación por limitaciones del programa empleado en la estimación de los componentes de varianza.

MRA basados en funciones lactacionales

Las funciones que determinan las covariables sobre las que se establece la regresión aleatoria son funciones lineales que definen una curva de lactación. Se han comparado dos modelos.

Modelo de Ali-Schaeffer (MRA-A)

La función lineal utilizada para definir las covariables en función de los DEL es la función lineal de 5 covariables de ALI-SCHAEFFER definida en [1]. Se ajusta esta función en las partes aditiva y permanente.

Modelo del Wilmink (MRA-W)

El modelo propuesto por WILMINK (1987) es una función lineal de tres covariables que describe la producción de leche $y(t)$ en función del tiempo t como:

$$y(t) = b_0 + b_1 t + b_2 exp(-0.05 t).$$

Se ajustó esta función tanto al efecto aditivo como al permanente.

MRA basados en polinomios de Legendre (MRA-L)

Se utiliza la serie de orden n de polinomios de Legendre normalizados (KIRKPATRICK y HECKMAN, 1989, KIRKPATRICK y BULMER, 1990; KIRKPATRICK y col 1994; MEYER y HILL, 1997; APOSTOL ,1973) donde el término de orden n de la serie viene dado por:

$$P_n(x) = \frac{1}{2^n}\sqrt{\frac{2n+1}{2}}\sum_{m=0}^{(n/2)}(-1)^m\binom{n}{m}\binom{2n-2m}{j}x^{n-2m} \qquad [7]$$

siendo x el DEL estandarizado dentro del intervalo (-1,1) que es donde está definida la serie ortogonal de polinomios de Legendre, calculada a partir de la expresión

$$x = \frac{2(t - t_{min})}{t_{max} - t_{min}} - 1,$$

siendo t días en lactación, t_{min} día mínimo de la lactación y t_{max} día máximo de la lactación. En la expresión [7] *(n/2)* indica tomar la parte entera de n/2, y los términos del sumatorio en los que se verifique que *n-2m=0*, son independientes de x.

La notación que vamos a seguir para describir un modelo al que se ajustan polinomios de Legendre es:

$$y = EF + L_a(n) + L_p(m) + e \qquad [8]$$

donde *EF* son los efecto fijos definidos en [2], y $L_a(n)$ y $L_p(m)$ representan los polinomios de Legendre de orden n y m ajustados sobre la parte aditiva y permanente, respectivamente. Para simplificar todavía más la expresión (8), en lo sucesivo nos referiremos a un MRA-L de la siguiente forma:

$$\begin{array}{l} La(2) + Lp(3) = A2 \\ La(2) + Lp(3) = A3 \\ \cdots\cdots \quad \cdots\cdots \quad \cdots\cdots \\ La(7) + Lp(6) = D7 \end{array}$$

donde la letra se refiere al orden del efecto permanente (A = 3; B = 4; C = 5; D = 6) y el número al orden de ajuste sobre el efecto aditivo. Por *orden* entendemos el número de coeficientes del polinomio.

Comparación de modelos

Para comparar los distintos modelos se han utilizado estadísticos que definen la bondad de ajuste y la capacidad para predecir datos faltantes. Además, en los MRA se estudio la forma en la que predecían las varianzas aditiva y permanente diarias a lo largo de la lactación, así como el cociente entre la varianza aditiva diaria (VAD) entre la suma de la varianza total.

Las estimas de los componentes de varianza, soluciones de los efectos y cálculo de las verosimilitudes para todos los modelos se realizaron con el programa REMLF90 (MISZTAL, 2000).

Bondad del ajuste

Para cuantificar la bondad del ajuste del modelo se usaron dos estadísticos: el porcentaje de sesgo al cuadrado para la capacidad de ajuste (PSB-F) y la correlación empírica para la capacidad de ajuste (RHO-F) entre el valor observado y el ajustado por el modelo. El PSB-F se define como:

$$\text{PSB-F} = \frac{\sum_{i=1}^{n} (y_{ijkl} - \hat{y}_{ijkl})^2}{\sum_{i=1}^{n} (y_{ijkl})^2} \times 100 \qquad [9]$$

siendo n el número total de observaciones, y_{ijkl} la PDC_l observada en el animal k dentro de los grupos RFCi y EE_j e $\hat{y}_{ijkl}$ la PDC calculada por la ecuación del modelo a partir de las estimas de los efectos asociados a esa observación. Un mayor PSB-F evidenciaría un peor ajuste del modelo a los datos.

Dentro de los MRA-L existen modelos anidados que hacen posible determinar por medio de pruebas chi-cuadrado sobre el ratio de verosimilitudes, el orden del polinomio hasta donde se produce un incremento significativo en la mejora del valor de la verosimilitud del modelo.

Capacidad de predicción

Un método para comparar modelos es la validación cruzada (VC). Una vez estimados los parámetros del modelo, podemos predecir datos faltantes, es decir, datos que no han intervenido en la estima de los parámetros del modelo, y en concreto vamos a predecir los datos contenidos en el conjunto PRE. Comparando el dato que predice la ecuación del modelo con el dato real contenido en PRE, podemos tener una medida comparativa de la capacidad de ajuste del modelo usando los mismos estadísticos que en el apartado anterior, donde ahora la correlación y el porcentaje de sesgo al cuadrado se definen en función del valor predicho y el dato faltante (RHO-P y PSB-P).

Estimas de las varianzas diarias

Con MRA en los que las covariables están definidas por funciones lineales, es posible estimar funciones de covarianza que permiten tener en cuenta la estructura de (co)varianzas entre las PDC a lo largo de la lactación. Si $\mathbf{z}_i$ es el vector de covariables para un tiempo t_i y $\mathbf{z}_j$ lo es para otro t_j, y $\mathbf{K}_a$ es la matriz de (co)varianzas entre los coeficientes de regresión aleatorios aditivos del modelo, entonces la (co)varianza aditiva a lo largo de la lactación viene dada por la siguiente función de dos variables:

$$g(t_i,t_j) = \mathbf{z}_i' \mathbf{K}_a \mathbf{z}_j \qquad [10]$$

En el caso en que $i = j$, la superficie que define [10], y en la que se encuentran las (co)varianzas de las PDC en función del tiempo, se reduce a una función de una variable que representa la evolución de la varianza a lo largo del ciclo de lactación. La expresión [10] definida para el efecto aditivo y para el permanente, junto con la varianza residual, nos va a permitir el cálculo de parámetros genéticos a lo largo de la lactación. Los cálculos fueron llevados a cabo entre los días 5 y 335 de la lactación.

Resultados y discusión

Comparación de modelos

Los estadísticos para *bondad de ajuste* obtenidos dentro de las series de MRA-L,

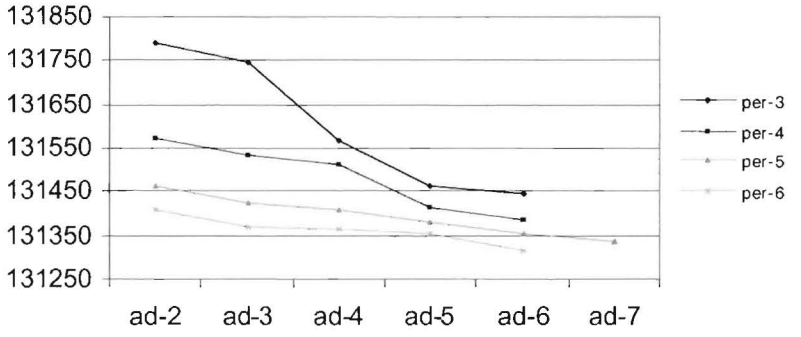


Figura 1. –2 log L*

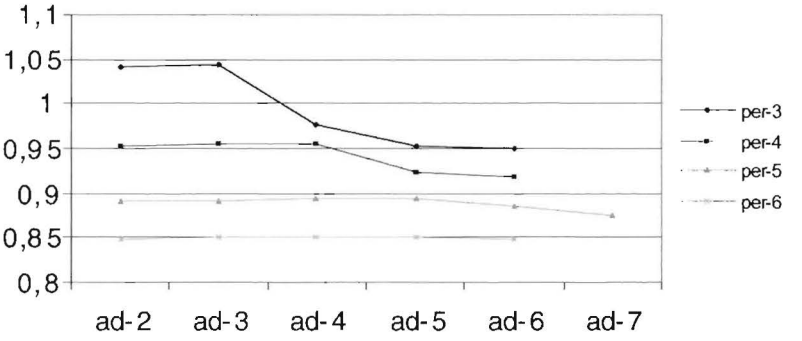


Figura 2. PSB para la bondad de ajuste- PSB-F
Figure 2. Goddness of fit using PSB- PSB-F

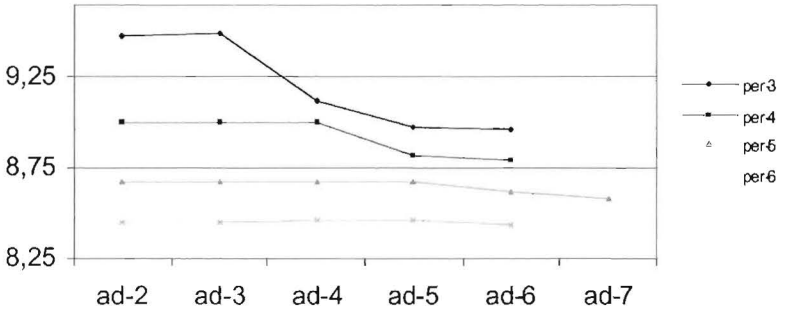


Figura 3. Varianza residual- var(e)*
*Figure 3. Residual variance- var(e)**

*En "ad-n" y "per-n", "n" indica el orden del polinomio que se ha ajustado a la parte aditiva y permanente respectivamente.
**In "ad-n" and "per-n", "n" indicates the order of the polynomial fitted to additive and permanent effect, respectively.*

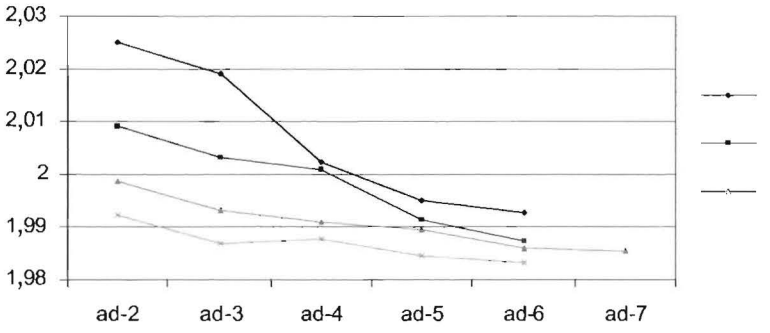


Figura 4. PSB para la capacidad predictiva PSB-P*
*Figure 4. PSB Predictive ability using- PSB-P**

*En "ad-n" y "per-n", "n" indica el orden del polinomio que se ha ajustado a la parte aditiva y permanente respectivamente.
**In "ad-n" and "per-n", "n" indicates the order of the polynomial fitted to additive and permanent effect, respectively.*

(logaritmo de verosimilitud, porcentaje de sesgo al cuadrado y varianza residual) se presentan en las figuras 1, 2 y 3 donde se muestra su tendencia ante el aumento de orden de ajuste sobre la parte aditiva dentro de distintos órdenes de ajuste constantes de la parte permanente. En el mismo sentido, la figura 4 muestra el porcentaje de sesgo al cuadrado para los MRA-L calculado en validación cruzada como medida de la *capacidad predictiva* de los modelos.

El valor de las verosimilitudes (figura 1) pone de manifiesto la mejora en la bondad de ajuste de los modelos paralela al incremento en el orden de los polinomios ajustados sobre los efectos aditivo y permanente. El valor de –2 log L disminuyó según se incremento el orden del polinomio sobre la parte aditiva manteniendo constante el orden de ajuste en la permanente, siendo la disminución en los valores de este estadístico más manifiesta en órdenes de ajuste sobre el efecto permanente inferiores (3 y 4) que en órdenes de ajuste superiores (5 y 6), donde las mejoras fueron más moderadas. Una tendencia similar en cuanto a la mejora del modelo, se encontró al analizar incrementos de orden sobre el efecto permanente dentro de un orden de ajuste constante sobre el efecto aditivo. Sin embargo, la disminución en el valor del –2 log L fue más pronunciada y se mantuvo para órdenes de efecto aditivo superiores (5 y 6). Sobre los ratios de verosimilitudes de los MRA-L anidados, se llevaron a cabo test chi cuadrado para determinar hasta que orden de ajuste se producía un descenso estadísticamente significativo en el valor de –2 log L (figura 1). En todos los casos resultó significativa la mejora que se produjo al incrementar el orden desde un grado 2 hasta grado 6, tanto en la parte aditiva como en la permanente. El paso de un modelo C6 a un C7 no resulto significativo. Los modelos de orden de ajuste 6 sobre el efecto permanente resultaron mejores que los de orden 5. El mejor modelo encontrado fue el D6. No se probó el modelo D7.

Otros estadísticos para medir la capacidad de ajuste, porcentaje de sesgo al cuadrado para la capacidad predictiva, (PSB-F) y varianza residual (var(e)), también informaron de que los mejores modelos eran los de órdenes superiores. En las figuras 2 y 3 se observa que la los estadísticos PSB-F y var(e) evolucionaron hacia valores cada vez más altos, al pasar de modelos con órdenes de ajuste inferiores sobre el efecto aditivo a modelos con órdenes superiores, mientras se mantuvo constante el orden de ajuste en el efecto permanente, hasta llegar a un determinado orden de ajuste en el que se produjo un descenso brusco en el valor de los estadísticos, que es suficiente para poner de manifiesto que los modelos de orden superior tienen mejores cualidades de ajuste. Este fenómeno no se puso de manifiesto si el orden del efecto permanente se aumenta sobre modelos con el efecto aditivo constante, en este caso, tanto PSB-F como var(e) disminuyeron a la vez que se incrementaba el orden del permanente. Dentro de los MRA-L, los valores de la correlación para la bondad de ajuste (RHO-F) mostraron el mismo comportamiento que el PSB-F y la var(e), oscilando entre un 92,76 y un 94,21%. En los estadísticos de medida sobre la capacidad de predicción, porcentaje de sesgo al cuadrado para la capacidad de predicción (PSB-P) (figura 4), y correlación (RHO-P), se encontró que mejo-

Cuadro 3
Estadísticos para los MRA-L más representativos de cara a la comparación, MRA lactacionales y el MMR (siglas de modelos explicadas en modelos de regresión aleatoria)
Table 3
Comparisons between MRA and MRR (model symbol explained in Random regression models section)

	Bondad del ajuste del modelo			Capacidad de predicción		
Modelo	-2 log L	PSB-F	RHO-F	var(e)	PSB-P	RHO-P
A2	131792	1,0424	0,9276	9,4842	2,0252	0,8553
A3	131744	1,044	0,9274	9,486	2,0191	0,8558
A4	131567	0,9774	0,9324	9,1277	2,0021	0,8571
A5	131461	0,9524	0,9343	8,981	1,9949	0,8576
B2	131571	0,9523	0,9343	9,001	2,0089	0,8566
B4	131514	0,9549	0,9341	9,0087	2,001	0,8572
B6	131389	0,9168	0,9369	8,7917	1,9874	0,8582
C2	131463	0,8908	0,9389	8,6715	1,9987	0,8574
C5	131381	0,8933	0,9387	8,6756	1,9897	0,858
C6	131356	0,8844	0,9394	8,6256	1,9857	0,8584
C7	131340	0,8751	0,9401	8,5761	1,9853	0,8584
D3	131370	0,8499	0,942	8,4537	1,9867	0,8583
D2	131408	0,8488	0,942	8,4511	1,9922	0,8579
D6	131316	0,8478	0,9421	8,436	1,9833	0,8585
MMR	134168	1,586	0,8859	12,493	2,2783	0,8354
MRA-W	131706	1,0658	0,9257	9,5713	2,03	0,8549
MRA-A		0,9755	0,9326	9,0957	2,0034	0,857

raban simultáneamente al incremento de orden de efectos aditivos y permanente. Los valores de RHO-P dentro de los MRA-L, oscilaron entre un 85,53 y un 85,85%. La variabilidad observada en los resultados de las medidas de capacidad de predicción fue menor que la observada en las de bondad de ajuste (cuadro 3).

Como era previsible, el MMR fue el modelo que presentó peores resultados, tanto desde el punto de vista de la bondad de ajuste como desde el de la capacidad predictiva. Los MRA que utilizaron funciones lactacionales para ajustarse a la parte aditiva y permanente, mostraron peores resultados que los MRA-L. A igualdad de número de parámetros ajustados, el modelo de Wilmink equivaldría a un MRA-L A3, presentando el modelo A3 mejores resultados. Por su parte, el MRA basado en la función lineal de Ali-Schaeffer, equivaldría en cuanto a número de parámetros a un MRA-L C5, siendo el modelo C5 mejor en cuanto a los resultados presentados. Un MRA de Ali-Schaeffer, se situaría en términos comparativos en una posición intermedia entre un A4 y un A5.

Estimas de parámetros genéticos

Se evaluaron las varianzas genética aditiva y permanente diarias (VAD y VPD) como funciones de los días en lactación (DEL). Los resultados para los modelos más representativos se muestran en las figuras 5 y 6.

La varianza genética aditiva estimada con el MMR fue de 5,80, mientras que la permanente y residual fue de 9,89 y 12,49, respectivamente. REKAYA y col. (1999) encuentran un valor similar para la varianza aditiva de 5,65, y valores algo más bajos para las varianzas permanente y residual de 6,60 y 6,68, respectivamente. Los MRA permiten oscilaciones de esta varianza a lo largo de la lactación. En la figura 5 se observa que dependiendo del MRA, el inicio de la curva es ligeramente distinto. Los modelos de Ali-

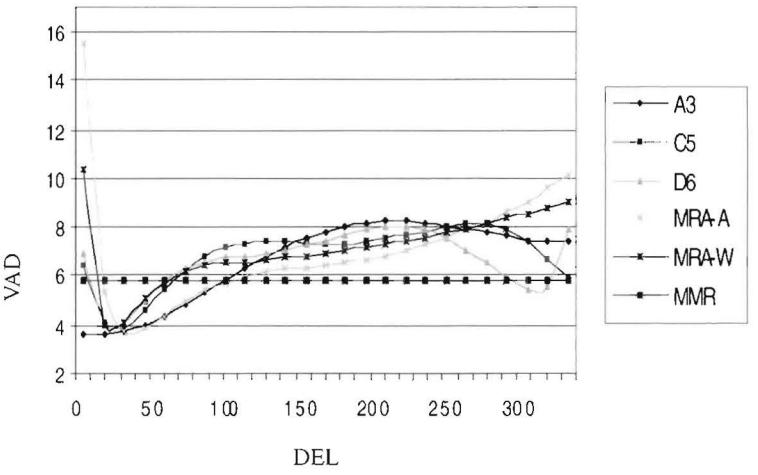


Figura 5. Varianza aditiva diaria- VAD *
*Figure 5. Daily additive variance- VAD **

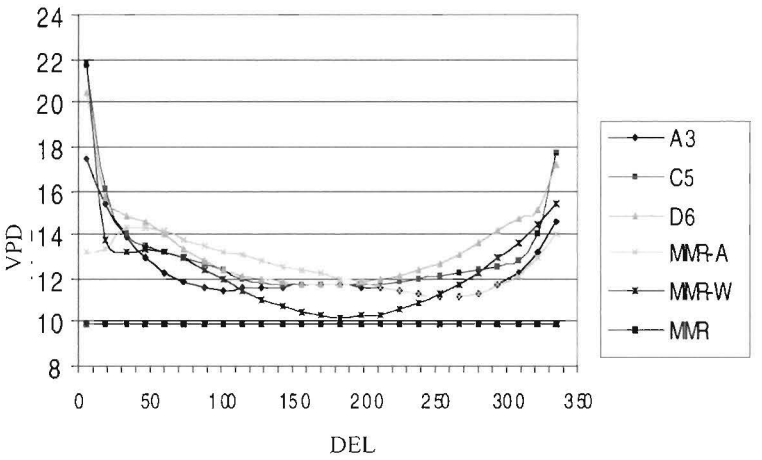


Figura 6. Varianza ambiental permanente diaria- VPD *
*Figure 6. Daily permanent environmental variance- VPD **

*Siglas de modelos explicadas en Modelos de regresión aleatoria.
** Model symbol explained in Random regression mode section.*

Schaeffer, Wilmink y los MRA-L de orden elevado, D6 y C5, toman valores altos al principio de la lactación decreciendo hasta el día 50. En el día 50 todos los modelos alcanzan el mínimo de VAD, y desde aquí todos muestran una línea ascendente hasta los 275-300 días. A partir de aquí cada modelo difiere en su comportamiento. En la figura 6 se ve como la VPD fue modelizada por todos los modelos de forma muy similar, presentando el MMR el valor más bajo.

La tendencia general de la "heredabilidad" (h^2) (figura 7) fue similar al de la VAD. Al principio de la lactación los modelos presentaron unos valores de h^2, relativamente altos, que fueron decreciendo hasta el día 50 con valores en torno a 0.15. Desde aquí aumentaron a valores en torno a 0.27, manteniéndose en la mayor parte de la lactación. A partir del día 300 los modelos manifestaron comportamientos distintos. Estas tendencias de la h^2 al principio y final de la lactación, se ponen de manifiesto en MRA que ajustan un elevado número de parámetros en al parte aditiva, por lo que parece ser debido al propio funcionamiento de los MRA. STRABEL y MISZTAL (1999) sugieren que este fenómeno se debe a la pérdida de precisión en la estima de los componentes de varianza que se produce los MRA cuando se incrementa el número de parámetros que han de ser estimados.

Este comportamiento también se puso de manifiesto en las "correlaciones genéticas diarias" (CGD). A partir de un orden cuarto de ajuste sobre la parte aditiva se obtuvieron correlaciones muy bajas e incluso negativas entre PDCs situados en la periferia de la lactación. Este resultado se muestra para el modelo D6 en la figura 8, habiéndose encontrado los mismos resultados para órdenes inferiores de efecto permanente desde órdenes de ajuste 4 para el efecto aditivo. Con órdenes 2 y 3 de ajuste sobre la parte aditiva se encontró que la correlación entre dos

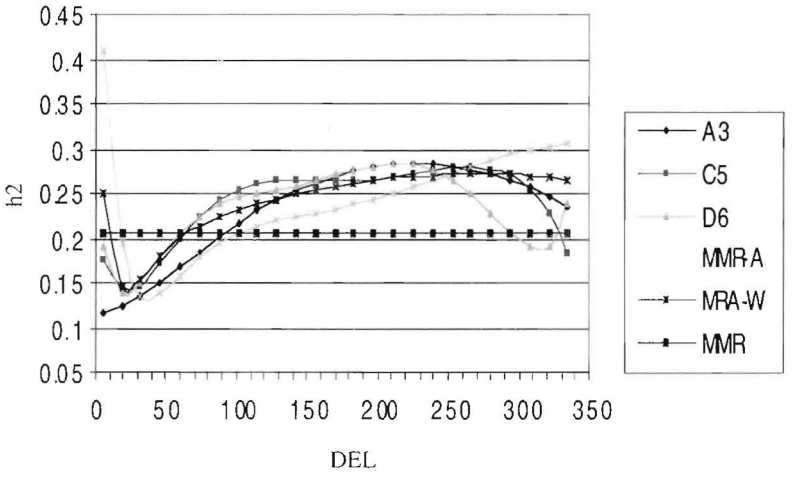


Figura 7. "Heredabilidad" diaria* (*Siglas de modelos explicadas en Modelos de regresión aleatoria)
Figure 7. Daily "heritability" (*Model symbol explained in Random regression models section)*

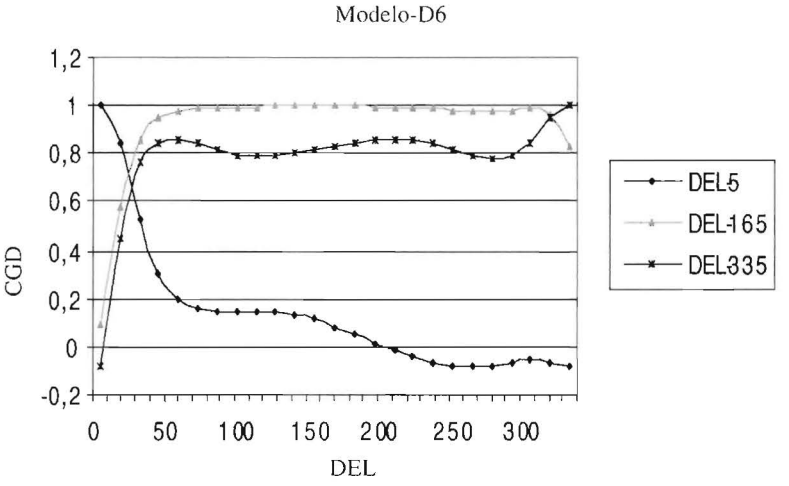


Figura 8. CGD para el modelo MRA-L D6
Figure 8. GDC for the MRA-L D6 model

PDCs medidos tiempos t_1 y t_2 disminuye a medida que t_2 aleja de t_1 (figura 9), encontrándose valores de CGD con un modelo D3 para PDCs entre los DEL 5 y 335 de 0,58. Resultados similares han sido presentados por STRABEL y MISZTAL (1999).

De igual forma los MRA-A y MRA-W mostraron resultados en las estimas de parámetros genéticos equivalentes a los MRA-L de elevado orden de ajuste (figuras 7, 10 y 11). En este caso, además del elevado número de parámetros del modelo (MRA-A), el

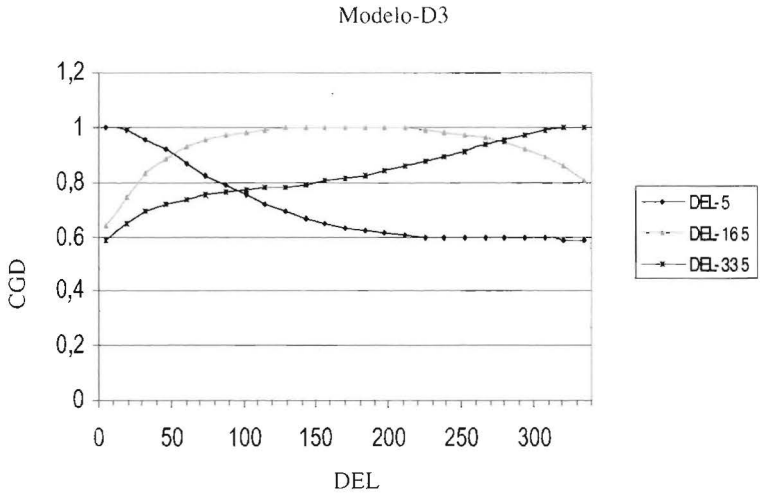


Figura 9. CGD para el modelo MRA-L D3
Figure 9. CGD for the MRA-L D3 model

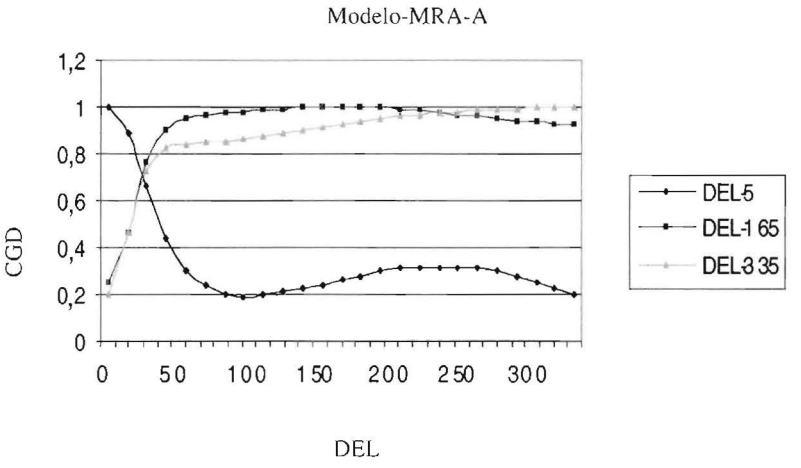


Figura 10. CGD para el modelo MRA-A
Figure 10. *CGD for the MRA-A model*

hecho de haber obtenido h^2 muy altas al principio y final de la lactación y haber obtenido CGD muy bajas para PDCs periféricos, también pudo ser debido a que los modelos lactacionales no son adecuados para modelizar al estructura de (co)varianzas, aunque sí lo sean para definir la curva de lactación.

Pequeñas imprecisiones en las estimas de los componentes de varianza del MRA, parecen traducirse en un comportamiento anómalo de los parámetros genéticos al principio y final de la lactación. En nuestro caso, las altas estimas encontradas para la VAD (figura 5) al principio de la lactación, tam-

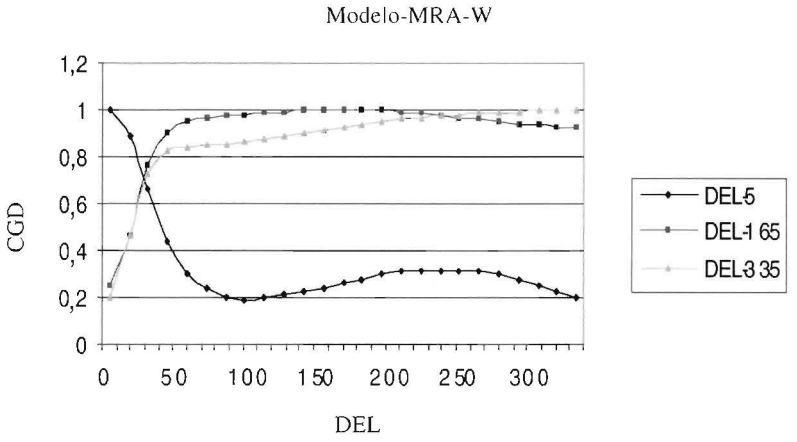


Figura 11. CGD para el modelo MRA-W
Figure 11. CGD for the MRA-W model

bién pudieron ser debidas a trabajar con modelos con homogeneidad en la varianza residual, sin embargo las estimas altas de la VAD sólo se manifestaron en MRA que ajustaron órdenes superiores a 3 en la parte aditiva. Este fenómeno no se observó si el incremento de orden se producía en la función que se ajustaba al efecto permanente. En este sentido, OLORI y col. (1999_a) afirman que mantener homogeneidad en la varianza residual a lo largo de la lactación en MRA, produciría sesgos en la propia varianza residual al principio y final de la lactación, pero no afectaría significativamente a las estimas del resto de las varianzas.

Los resultados obtenidos demuestran que aumentar el orden de los polinomios se traducen siempre en mejoras en la capacidad de ajuste, sin embargo esto lleva aparejado una mala modelización de los parámetros genéticos en la periferia de la lactación, lo que puede ser debido a una pérdida de precisión en la estima de los componentes de varianza cuando estos son incrementados en número, o al propio comportamiento del MRA que no se ajusta bien en los extremos de la lactación. Además, PLETCHER y GEYER (1999) afirman que los polinomios de alto grado son "oscilantes" por lo que no se ajustarían bien a la FC. Una solución a esto sería trabajar con polinomios de menor grado de ajuste. PLETCHER y GEYER señalan a su vez, que un polinomio de grado excesivamente reducido podría no ser suficiente para definir satisfactoriamente el proceso biológico que se este modelizando. Por tanto, parece lógico que hay que definir un polinomio con un orden de ajuste óptimo, para lo que habría que tener en cuenta, además de factores como la capacidad de ajuste y la capacidad para predecir datos faltantes, otros relacionados con la precisión de los componentes de varianza y la posterior estima de parámetros genéticos, teniendo especial atención con lo que ocurre en los extremos de la lactación, con la capacidad del polinomio para definir el proceso biológico, y finalmente con el coste computacional.

Conclusiones

Los resultados de la comparación de modelos obtenidos, tanto en lo referente a bondad de ajuste como capacidad de predicción, discriminaron muy en favor de los MRA frente al MMR.

Dentro de los MRA, los MRA-L mostraron mejores resultados que los MRA basados en funciones lactacionales (MRA-A y MRA-W), a igualdad de número de parámetros ajustados.

En los MRA-L, se encontró que los modelos mejoraron sus propiedades en cuanto a bondad de ajuste y capacidad de predicción de forma paralela al aumento en el número de parámetros, precisando que, incrementos en el orden del polinomio que se ajustó sobre la parte aditiva, produjeron mejoras más acusadas en modelos donde el orden ajustado sobre la parte permanente no fue muy elevado (3-4). Sin embargo, los mismos incrementos en el orden de ajuste sobre la parte aditiva en modelos con un orden elevado en la permanente (5-6) produjeron mejoras poco relevantes. Por su parte, incrementar el orden del efecto permanente produjo mejoras en el modelo independientemente del orden que se ajustó sobre la parte aditiva. Por tanto, desde el punto de vista del ahorro computacional, parece razonable tratar de mejorar las características del modelo incrementando el orden de ajuste sobre la parte permanente manteniendo órdenes de ajustes no muy elevados en la parte aditiva.

Por otro lado, el aumento en el número de parámetros del modelo condujo a resultados muy elevados de h^2 al principio y final de lactación, a la vez que resultados de CGD muy bajos entre PDCs situadas en la periferia de la lactación. No obstante, este efecto sólo se puso de manifiesto cuando el incremento de orden de ajuste se produjo sobre la parte aditiva del modelo.

Sumado a esto último el hecho de que, las mejoras que se produjeron en el modelo, como resultado de incrementar el orden de ajuste en la parte aditiva en modelos de elevado orden de ajuste en la parte permanente no fueron muy superiores, nos podemos permitir sugerir que el modelo MRA-L D3 sería el más adecuado para su uso, por ser un modelo que presenta buenas características en cuanto a bondad de ajuste y capacidad de predicción, sin presentar el inconveniente de la imprecisión en la estima de los componentes de varianza.

Dado que los resultados obtenidos se basan en modelos que consideraron homogeneidad en la varianza residual, y es conocido el hecho de que modelos con heterogeneidad de la varianza residual presentan mejores resultados, se deberían hacer más investigaciones al respecto. Así mismo, se deberían comparar los resultados de este análisis, con otros que se obtengan a partir de otras muestras de datos en otras regiones. Por último, se debe plantear un estudio sobre el posible cambio en el mérito genético de los animales que se podría producir con modelos basados en el análisis de las PDC.

Bibliografía

ALI T.T., SCHAEFFER L.R., 1987. Accounting for covariances among Test day milk yield in dairy cows.

APOSTOL T.M., 1973. Calculus. Vol- 1. pag-700.

JAMROZIK J., KISTEMAKER J., SCHAEFFER L.R., DEKKERS J. C.M., 1997. Comparison of possible covariates for use in a random regression model for analyses of test day yields. J. Dairy Sci. 80: 2550.

KIRKPATRICK M., HECKMAN N., 1989. A quantitative genetic model for growth, shape, reaction norms,

and other infinite-dimensional characters. J. of Math. Biol. 27: 429.

KIRKPATRICK M., BULMER M., 1990. Anaysis of the inheritance, selection and evolution of growth trajectories. Genetics. 124:979.

KIRKPATRICK M., HILL W.G. y THOMPSON R., 1994. Estimationg covariance structure of traits during growth and ageing, illustration with lactations in dairy cattle. Genet. Res. 64:57.

MEYER K., HILL W.G., 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or "repeated" records by restricted maximum likelihood. Livest. Prod. Sci. 47:185.

MISZTAL I., 2000. REMLF90 Manual. ftp://num.ads.uga.edu

OLORI V.E., HILL W.G., BROTHERSTONE S., 1999a. The structure of the residual error variance of test day milk yield in random regression models. Workshop on Computational Breeding. Finland. 18-20 March.

OLORI V.E., HILL W.G., MCGUIRK, BROTHERSTONE S., 1999b. Estimating variance components for test day milk records by restricted maximum likelihood with a random regression model. Livest. Prod. Sci. 61: 53.

PLETCHER S.D., GEYER C.J., 1999. The Genetic analysis of age-dependente traits: Modeling the character process. Genetics: 151:825.

PTAK E., SCHAEFFER L., 1993. Use of test day yields for genetic evaluation of dairy sires and cows. Livest. Prod. Sci. 34: 23.

REENTS R., JAMROZIK J., DEKKERS J.C.M., SCHAEFFER L.R., 1995a. Estimation of genetic parameters for test day records of somatic cell score. J. Dairy Sci. 78:2847.

REENTS R., DEKKERS J.C.M., SCHAEFFER L.R., 1995b. Genetic evaluation for somatic cell score with a test day model for multiple lactations. J. Dairy Sci. 78:2858.

REKAYA R., CARABAÑO M.J., TORO M.A., 1999. Use of test day yields for the genetic evaluation of production traits in Holstein-Friesian cattle. Livest. Prod. Sci. 57:203.

STRABEL, T., MISZTAL I., 1999. Genetic parameters for first and second lactation milk yields of Polish black and white cattle with random regression test-day model. J Diary Sci. 82:2805.

TIJANI A., WIGGANS G.R. , VAN TASSEL C.P., PHILPOT J. C., GENGLER N., 1999. Use of (co)variance functions to describe (co)variances for test day yield. J. Dairy Sci. 82.

VAN DER WERF, J.H.J., GODDARD M.E., MEYER, K., 1998. The use of covariance funtion and random regressions for genetic evaluation of milk production based on test day records. J Diary Sci. 81: 3300.

WILMINK J.B.M., 1987. Adjusment of test day milk, fat and protein yield for age, season ans stage of lactation. Livest. Prod. Sci. 16:335.