



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

DOCTORADO EN CIENCIAS VETERINARIAS

**EL MODELO DEL DÍA DE PRUEBA PARA CORREGIR EFECTOS
AMBIENTALES EN LA PREDICCIÓN DE LOS VALORES GENÉTICOS
DE CARACTERÍSTICAS DE LECHE EN HOLSTEIN**

TESIS
PARA OBTENER EL GRADO DE
DOCTOR EN CIENCIAS VETERINARIAS

PRESENTA

CARLOS ANTONIO APODACA SARABIA

TUTOR

CARLOS FRANCISCO SOSA FERREIRA

COMITÉ TUTORAL

**MOISÉS MONTAÑO BERMÚDEZ
FELIPE DE J. RUIZ LÓPEZ
CARLOS M. BECERRIL PÉREZ
RAÚL ULLOA ARVIZU
PEDRO OCHOA GALVÁN
IGNACIO MONDRAGÓN VÁZQUEZ**



MEXICO, D.F.

2007



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

DEDICATORIAS

A mi Familia, Lilian, Carlos Francisco, Ricardo Ernesto y L. Raquel, con quien tantas satisfacciones he tenido en todos estos años.

A mis padres, Salvador Apodaca Sanchez y Francisca Sarabia Salmeron, a quien no sólo les debo la vida sino también mi formación como ser humano.

A todos aquellos para quien más que morir por la patria viven cada día de su vida por ella.

AGRADECIMIENTOS

A la Universidad Nacional Autónoma de México y Universidad Autónoma Chapingo, pilares del sistema de educación superior y educación superior Agropecuaria y forestal de México.

Al Dr. Carlos Francisco Sosa Ferreira que con paciencia y seguridad sabe dar confianza en uno mismo.

Al comité asesor por su apoyo y tiempo dedicado a la realización del proyecto. En especial a los Doctores Moisés Montaña Bermúdez y Pedro Ochoa Galván por su apoyo e interés en la continuación del proyecto.

Al Dr. Carlos M. Becerril Pérez por su incansable participación en todas las actividades del programa de doctorado.

Al Dr. Felipe de J. Ruiz López, por sus valiosas aportaciones al escrito final del trabajo.

Al Dr. Raúl Ulloa Arvizu cuyo apoyo fue de gran importancia en la finalización del trabajo.

A todos aquellos que de uno u otra forma me manifestaron su apoyo y confianza.

CONTENIDO

Resumen general.....	vii
General abstract.....	Viii
I. Introducción general.....	1
1.1. Literatura citada.....	6
II. Las ecuaciones de modelos mixtos y el modelo del día de prueba.....	9
Resumen.....	9
Abstract.....	10
2.1. Introducción.....	11
2.2. Derivación de las ecuaciones de los modelos mixtos.....	13
2.3. Los diferentes modelos del día de registro.....	17
2.3.1. Modelo de repetibilidad.....	17
2.3.2. Modelo multivariado.....	19
2.3.3. Modelo de regresión aleatoria.....	21
2.3.4. Modelo de funciones de covarianza.....	24
2.3.5. Diferencias entre los modelo del día de registro.....	26
2.4. Estimación de componentes de varianza.....	28
2.4.1. Máxima Verosimilitud.....	28
2.4.2. Máxima verosimilitud restringida.....	29
2.4.3. Funciones de covarianza.....	30
2.5. La función de máxima verosimilitud de los diferentes modelos y su solución.....	32
2.5.1. Análisis univariado.....	32
2.5.2. Análisis multivariado.....	33
2.5.3. Modelo de funciones de covarianza.....	35

2.6. Solución de las ecuaciones de los modelos mixtos.....	38
2.6.1. Absorción.....	38
2.6.2. Eliminación de Gauss.....	39
2.6.3. Descomposición Cholesky.....	39
2.6.4. Métodos iterativos.....	40
2.6.5. Comparación de modelos.....	41
2.7. Literatura citada.....	44
A2.1. Anexo 1. Regresión aleatoria.....	51
A2.2. Anexo 2. Predicción de la estructura de covarianza con funciones de covarianza según Kirkpatrick y colaboradores.....	60
a. Orden completo.....	61
b. Orden reducido.....	65
III. Diferentes modelos de regresión y su ajuste en producción de leche por día	68
Resumen.....	68
Abstract.....	69
3.1. Introducción.....	70
3.2. Materiales y métodos.....	70
3.3. Resultados y discusión.....	73
3.4. Conclusiones.....	77
3.5. Literatura citada.....	77
IV. Estimación de (co)varianzas genéticas y de ambiente permanente con diferentes modelos de funciones de covarianza.....	80
Resumen.....	80
Abstract.....	81

4.1. Introducción.....	82
4.2. Materiales y métodos.....	84
4.3. Resultados y discusión.....	89
4.4. Conclusiones.....	100
4.5. Literatura citada.....	101
V. Conclusiones generales.....	105

ÍNDICE DE CUADROS

Cuadro	Titulo	Pag.
2.1	Varianzas de los efectos aleatorios en los diferentes modelos del modelo del día de prueba y el número de parámetros a estimar.....	27
A1.1	Producción diaria de leche de cuatro vacas.....	51
A1.2	Ecuaciones de predicción y valores genéticos (VG) al día 15, 60 y 150 y acumulada en 150 días.....	59
3.1	Medias de mínimos cuadrados y su error estándar de correlación entre observados y predichos según la ecuación de regresión y el número de lactancia.....	75
3.2	Medias de mínimos cuadrados de porcentaje de la suma de cuadrados del error en relación a la suma de cuadrados total.....	75
4.1	Ecuaciones de uso en modelos de regresión aleatoria para producción diaria de leche en una lactancia.....	83
4.2	Elementos del vector de parámetros desconocidos (β) y de los renglones de la matriz de covariables (X) según la ecuación de regresión.....	86
4.3	Efectos fijos y ecuación de regresión, fija y aleatoria incluida en cada uno de los modelos usados en la estimación de los componentes de varianza.....	87
4.4	Número de observaciones (Obs.) media y desviación estándar (DS) de producción diaria de leche por mes de lactancia.....	90
4.5	Logaritmo natural de la función de verosimilitud restringida para los modelos que se logró convergencia y el orden de ajuste (n) por el valor de la función de máxima verosimilitud.....	90

ÍNDICE DE FIGURAS

Figura	Título	Pag.
2.1	Producción de leche por día (PLD) y valor genético (Vg) de un animal.....	21
2.2	Covarianzas entre diferentes días de producción de leche en forma discreta (A) y continua (B).....	30
A1.1	Producción diaria de leche (PLD).....	57
A1.2	Valor genético de producción diaria de leche (VG) para cada uno de los 6 animales.....	58
3.1	Curvas de lactancia por número de lactancia según las ecuaciones de Ali y Schaeffer (AS), Guo y Swalve (GS), y Wood (WO).....	74
4.1	Producción diaria de leche (PLD) y su valor genético (VG) de un animal hipotético.....	82
4.2	Promedio de producción diaria de leche observada y producción predicha con diferentes modelos de funciones de covarianza.....	92
4.2.	Promedio de producción diaria de leche observada y producción predicha con diferentes modelos de funciones de covarianza (continuación).....	93
4.3	Heredabilidad (h^2) y repetibilidad (r) estimadas con diferentes modelos de funciones de covarianza.....	95
4.4	Heredabilidad (h^2) y repetibilidad (r) estimadas con diferentes modelos de funciones de covarianza.....	97
4.5	Correlaciones genéticas estimadas con diferentes modelos de funciones de covarianza.....	99

EL MODELO DEL DÍA DE PRUEBA PARA CORREGIR EFECTOS AMBIENTALES EN LA PREDICCIÓN DE LOS VALORES GENÉTICOS DE CARACTERÍSTICAS DE LECHE EN HOLSTEIN¹

Apodaca Sarabia Carlos A.²

Resumen general

Con los objetivos de: a) comparar el ajuste de diferentes ecuaciones de regresión a la producción diaria de leche (PLD) durante la lactancia y b) comparar diferentes modelos de funciones de covarianza (COF) para describir genéticamente la PLD, se analizaron 18421 registros mensuales de vacas con 1 a 12 lactancias en el periodo de 1983 a 1993, de un establo en el norte de México. En el primer análisis, realizado en dos etapas, se comparó el ajuste a PLD de las ecuaciones de regresión; Ali y Schaeffer (AS), Guo y Swalve (GS), Wilmink (WL), y Wood (WO). En la primera etapa se calculó para cada ecuación-vaca-lactancia, la correlación entre las PLD observadas y predichas ($r_{\hat{y}y}$) y el porcentaje de la suma de cuadrados del error con respecto a la suma de cuadrados total (PSB). En la segunda etapa se usó un modelo mixto con PSB y $r_{\hat{y}y}$ las variables de respuesta, y los independientes; año-estación de parto (AEP) y número de lactancia-ecuación (NLE) efectos fijos, y vaca (V) aleatorio. El ajuste de las ecuaciones se comparó dentro de lactancia. En el segundo análisis se compararon 24 modelos de COF formados por la combinación de 3 efectos fijos y 8 ecuaciones de regresión, en cada modelo, la regresión fija y las aleatorias (valor genético y efectos de ambiente permanente) fue la misma. Los efectos fijos, además de la regresión, fueron: a) año de parto (AP), época de parto (EP) y número de lactancia (NL), b) la combinación de los tres anteriores (AENL), y c) día de prueba (DR). Las ocho ecuaciones de regresión fueron: AS, GS, WL, WO, y los polinomios de Legendre de segundo a quinto orden (PL_i; i=2, 3, 4 y 5). En el primer análisis, los efectos AEP y NLE fueron significativos ($P < 0.0001$) en $r_{\hat{y}y}$ y PSB. La varianza de V fue el 43.61 y 39.43% de la varianza total de $r_{\hat{y}y}$ y PSB, respectivamente. La ecuación de regresión AS ajustó mejor en todas las lactancias ($P < 0.0001$). En todas las ecuaciones, el ajuste fue ascendente hasta la cuarta o quinta lactancia y descendió en la sexta. En el segundo análisis, los modelos con los tres efectos fijos y PL₄, PL₅ o WO no convergieron. Los modelos con los tres efectos fijos y WL, produjeron estimaciones de varianza aditiva mayor que la varianza fenotípica observada. Los modelos DR-AS y DR-GS, presentaron el mejor ajuste. Las estimaciones de heredabilidad de PLD, mostraron incremento desde el inicio de la lactancia (0.03 a 0.27). Las estimaciones de repetibilidad de PLD (0.285 a 0.520), aumentaron desde el primer mes hasta el segundo o quinto mes y disminuyeron hasta el final de lactancia. Las estimaciones de varianza genética mostraron incrementos desde el inicio hasta el final de la lactancia (1.51 a 13.88 kg²). Los menores valores de la varianza genética fueron estimados en los modelos con DR. Las correlaciones genéticas fueron altas y nunca negativas.

Palabras clave. Modelos, día de prueba, producción, leche, comparación.

¹ Tesis de doctor en ciencias veterinarias. Universidad Autónoma de México. Facultad de Medicina Veterinaria

² apodaca1953@hotmail.com

TEST DAY MODEL FOR CORRECTION OF NO GENETICS EFFECTS IN THE BREEDING VALUE PREDICTION OF MILK PRODUCTION TRAITS IN HOLSTEIN¹

Apodaca-Sarabia Carlos A. ²

General abstract

With the objectives of: a) compare the fit of different regression equations to phenotypic daily milk production (PLD), and b) compare 24 covariance function models (COF) to genetically describe daily milk production, 18421 daily milk production records were analyzed from cows with 1 to 12 lactations between 1973 and 1993 of a herd in the arid region of Mexico. The first analysis conducted in two parts compared the fit of regression equations proposed by Ali y Schaeffer (AS), Guo y Swalve (GS), Wilmink (WL), and Wood (WO). In the first part, it was calculated for each equation-cow-lactation, the correlation between predicted and observed ($r_{\hat{y}y}$) PLD and the percentage of square bias (PSB). In the second part, a mixed model was used with $r_{\hat{y}y}$ and PSB as the response variables and the independent variables were; year-season (AEP) and lactation-equation (NLE) as fixed effects, and cow (V) as random effect. The fit of equation was compared within lactation. In the second analysis, 24 COF models (3 fix effects and 8 regression equations) were compared. The fixed and random regressions (genetic and permanent) were the same in all models. The three fix effects were; i) year (AP), season (EP), and lactation (NL), ii) its combination (AENL), and iii) test day (DR). The regression equations were, AS, GS, WL, WO, and the Legendre polynomials of second to fifth order (PL_i; i=2, 3, 4, and 5). In the first, analysis the effects, AEP and NLE were significant ($P < 0.0001$) for $r_{\hat{y}y}$ and PSB. The variance of V explained 43.61 and 39.43 % of the total variance of $r_{\hat{y}y}$ and PSB, respectively. The regression equation proposed by AS, fit better than others, in all lactation. In all the equations the fit was ascending until the fourth or fifth lactation and decreased in the sixth. In the second analysis, the models with any of the three fixed effects and PL₄, PL₅, or WO did not converge. The models with the three fixed effects and WL, the genetic variances estimated were greater than phenotypic variances. The best fit were in the models DR-AS and DR-GS. The estimation of heritability for PLD was ascending from the beginning of lactation (0.03 to 0.27). The estimation of repeatability for PLD (0.285 to 0.52), increasing until the second or fifth, and decreasing the end of lactation. The estimation of genetic variance (1.51 to 13.88 kg²) was ascending from first to the tenth month of lactation. The estimated of genetic variance, with models including DR was the lowest. The genetic correlations were high and never negative.

Keywords. Models, test day, milk, production, comparison.

¹ Doctoral thesis in Veterinarian Sciences. Universidad Autónoma de México. Facultad de Veterinaria

² apodaca1953@hotmail.com

I. Introducción general

Las características de expresión continua en una trayectoria, (tiempo el más comunes), se denominan longitudinales. En producción animal, producción de leche, calidad de leche y ganancia diaria de peso son ejemplos de características de expresión continua en un periodo de tiempo, la trayectoria de expresión. La descripción genética de estas características se realiza para la producción acumulada entre periodos o en el tiempo que dure la expresión de la característica. Algunos ejemplos de ellas son: ganancia diaria de peso del destete al año o del año a los dieciocho meses, producción de leche, grasa o proteína en una lactancia.

En ganadería especializada en producción de leche, la producción de leche por lactancia debe calcularse tomando en cuenta los cambios en producción durante la lactancia, cambios representados gráficamente en la curva de lactancia. Así, la producción por lactancia se obtiene de registros realizados periódicamente, comúnmente mensual, a los que, para obtener la producción acumulada en la lactancia, se les aplica factores para considerar las diferentes etapas de la lactancia o factores de proyección cuando la lactancia esta incompleta y los días en lactancia lo permiten^{1,2}. Si el interés es describir genéticamente partes de la lactancia, es también necesario recurrir al uso de los registros de producción diaria, tomando como ejemplo producción máxima y persistencia, de forma tradicional se requiere de diferentes análisis para su descripción genética. Estos son: a) estimar la ecuación de predicción para cada vaca y lactancia (Wood³ la de mayor uso), con ellas calcular producción máxima y persistencia, y b) estimar sus parámetros genéticos y predecir los valores genéticos. Así, la descripción genética se realiza en dos etapas, la primera consiste en obtener la producción acumulada, o las ecuaciones de predicción y en la segunda se estiman los parámetros genéticos y calculan los valores genéticos de los animales. En el procedimiento antes descrito, se supone que los efectos no genéticos sobre la producción diaria obtenida en los registros periódicos, no son de importancia.

Una alternativa para la descripción genética de las características longitudinales, es el modelo del día de prueba (TDM por sus siglas del inglés, "test day model"). En mejoramiento genético animal, se conoce como TDM a los modelos mixtos que incluyen los efectos ambientales temporales y los genéticos en el análisis del comportamiento diario de una característica longitudinal. Esta es una ventaja conceptual de gran importancia, pues es de esperarse un

incremento considerable en la exactitud con que se predicen los valores genéticos de los animales. Incremento explicado no sólo por considerar los efectos del día de registro, sino también, por considerar toda la información de las lactancias en vez de sólo la producción total. Además, con relativa facilidad, se predicen los valores genéticos de parte de la lactancia como son la producción máxima y persistencia. Así, las ventajas del TDM son: considerar los efectos ambientales que influyeron sobre el registro el día en que se realizó la medición, evitar el cálculo de la producción total, utilizar lactancias incompletas sin necesidad de su proyección, y describir el comportamiento genético de la característica día a día en todo el periodo de su expresión⁴.

El término día de prueba (test day) para el análisis con las observaciones de producción diaria fue primero usado por Ali y Schaeffer⁵, no obstante, debido a que el objetivo es predecir el valor genético considerando los efectos sistemáticos temporales sobre el comportamiento el día de su registro, es posible el término adecuado en español sea día de registro. El trabajo pionero usó el término día de prueba para las observaciones de producción diaria de leche, sin embargo la metodología es general y se puede aplicar en las características de la lactancia en la industria lechera y todas las que se pueden registrar diario en un periodo de tiempo u alguna otra trayectoria. En producción animal, el TDM ha sido utilizado, principalmente, en la industria lechera, siendo diferentes los modelos a los que se ha recurrido. Una forma de clasificar las diferentes formas de uso del TDM, es por el número de fases (análisis) requeridos para describir genéticamente la característica, en general son una o dos fases. Las dos fases consisten en: a) análisis previo con la finalidad de corregir el comportamiento diario por los efectos ambientales de importancia y b) en el segundo análisis se estiman las componentes de varianzas y predicen los valores genéticos de los animales para producción por lactancia. La fase única consiste en la estimación de los efectos ambientales y predicción de los valores genéticos de los animales en un análisis.

La estimación de los efectos ambientales y componentes de varianza en la predicción de los valores genéticos con el TDM, es de gran interés. Las diferentes formas en que se ha realizado implican suponer que la característica longitudinal es: la misma registrada repetidamente, o los cambios fenotípicos corresponden a cambios no sólo ambientales, sino también a genéticos graduales en el periodo de su expresión. Los cambios graduales en la estructura de covarianza, es posible estimarlos suponiendo se tienen características correlacionadas fenotípica y

genéticamente o bien incluyendo en el modelo covariables que cuantifiquen los cambios durante el tiempo de expresión en promedio de toda la información, y los cambios en valor genético y de ambiente permanente en cada animal. Estas covariables se denominan regresión fija y regresiones aleatorias respectivamente. Así, por las anteriores suposiciones, los modelos usados en el TDM son; de repetibilidad⁶, multivariado^{7,8}, regresión aleatoria y funciones de covarianza respectivamente^{9,10,11,12,13,14,15}. Las diferencias entre ellos son de interés biológico y económico.

En las anteriores formas de uso del TDM, la suposición de la estructura de covarianza es su mayor diferencia. Una forma simple de comprender las diferentes estructuras de la matriz de covarianzas genéticas (**G**), consiste en expresarla mediante la matriz de correlaciones (**C**) pre y post multiplicada por la matriz diagonal **S**. Los elementos de esta última son la desviación estándar genética de cada característica o registro, así, para m características (registros)¹⁶:

$$\mathbf{G} = \mathbf{S}\mathbf{C}\mathbf{S}$$

$$\mathbf{G} = \begin{bmatrix} \sigma_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \sigma_2 & & \mathbf{0} \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \mathbf{0} & \mathbf{0} & \dots & \sigma_m \end{bmatrix} \begin{bmatrix} \mathbf{1} & r_{12} & \dots & r_{1m} \\ r_{21} & \mathbf{1} & & \mathbf{0} \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ r_{m1} & r_{m2} & \dots & \mathbf{1} \end{bmatrix} \begin{bmatrix} \sigma_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \sigma_2 & & \mathbf{0} \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \mathbf{0} & \mathbf{0} & \dots & \sigma_m \end{bmatrix}$$

Si existe heterogeneidad de varianzas, en la expresión anterior, los elemento de **S** (σ_i) son diferentes para los m registro e iguales existe homogeneidad. Es importante notar que bajo homogeneidad de varianzas, $\mathbf{S} = \mathbf{I}\sigma$. La comprensión de la estructura de **G** es más compleja, sin embargo, en la notación anterior al basarla en las correlaciones genéticas, es más fácil de entender, así, las diferentes estructuras de **G** implican cambios en **C**. Las estructuras son entonces: sin estructura (**C_{US}**), correlación en bandas (**C_{BC}**), y autoregresiva (**C_{AR}**), y se escriben:

$$\mathbf{C}_{US} = \begin{bmatrix} \mathbf{1} & a & b & c \\ a & \mathbf{1} & d & e \\ b & d & \mathbf{1} & f \\ c & e & f & \mathbf{1} \end{bmatrix}; \mathbf{C}_{BC} = \begin{bmatrix} \mathbf{1} & g & h & i \\ g & \mathbf{1} & g & h \\ h & g & \mathbf{1} & g \\ i & h & g & \mathbf{1} \end{bmatrix}; \mathbf{C}_{AR} = \begin{bmatrix} \mathbf{1} & r^{|t_2-t_1|} & r^{|t_3-t_1|} & r^{|t_4-t_1|} \\ r^{|t_2-t_1|} & \mathbf{1} & r^{|t_3-t_2|} & r^{|t_4-t_2|} \\ r^{|t_3-t_1|} & r^{|t_3-t_2|} & \mathbf{1} & r^{|t_4-t_3|} \\ r^{|t_4-t_1|} & r^{|t_4-t_2|} & r^{|t_4-t_3|} & \mathbf{1} \end{bmatrix}$$

Las diferentes estructuras provocan cambios en el cálculo del valor genético del i -ésimo animal en el j -ésimo día (a_{ij}). Independientemente de la estructura, a_{ij} es función de los genes involucrados en la expresión en el día $j-k$ (a_{ij-k}) más el efecto de los genes actuando en el registro nuevo (α_j), la diferencia radica en la correlación (ρ) que multiplica al valor genético del animal (i -ésimo) en el día $j-k$ con el actual. Así, a_{ij} es igual a; i) $\rho_{jk}a_{ij-k} + \alpha_j$, ii) $\rho_k a_{ij-k} + \alpha_j$, y iii) $\rho^{j-k} a_{ij-k} + \alpha_j$, para, \mathbf{C}_{US} , \mathbf{C}_{BC} y \mathbf{C}_{AR} respectivamente. En \mathbf{C}_{US} se supone los registros anteriores son independientes y $j-k \geq 0$, mientras en \mathbf{C}_{AR} , α_j representa los genes actuando en la j -ésima observación con los efectos de los valores de los genes en las observaciones anteriores y en \mathbf{C}_{BC} , se supone la misma correlación para todas las observaciones. El último supuesto no es aplicable a diferentes estados de crecimiento¹⁶. En las características medidas en el tiempo, \mathbf{G} es sin estructura y la forma de estimarla con un reducido número de parámetros, es usar su relación con el tiempo mediante una función continua^{9,16}.

La estructura de \mathbf{G} es otro caso específico para el modelo de repetibilidad, en este la correlación genética entre los diferentes días es uno, así sólo se estima la varianzas genética y la ambiental permanente, entonces:

$$\mathbf{C}_{REP} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

De esta forma, si el objetivo es describir la estructura de covarianza de producción diaria de leche en la lactancia, por la evidencia de cambios en el valor genético de producción diaria de leche durante la lactancia^{5,17,18,19}, el modelo de repetibilidad no es el más adecuado. El modelo multivariado, para la descripción completa de la estructura de covarianza requiere de tantas variables como días dure la lactancia, siendo entonces impráctico. La regresión aleatoria y las funciones de covarianza describen genéticamente las características de la lactancia, con mucho menos parámetros que el análisis multivariado.

La clave de los modelos de regresión aleatoria y funciones de covarianza es expresar el valor genético del comportamiento por día de una característica de la lactancia con la ecuación de regresión adecuada, así, es entonces posible predecir el valor genético de la característica en cualquier momento de la lactancia. En ambos modelos cuando se aplica el concepto en la

metodología de los modelos mixtos se denomina regresión aleatoria¹⁰. Para el modelo de funciones de covarianza es necesario considerar que las (co)varianzas entre producción diaria, son continuas y se pueden expresar mediante una función, que se conoce como función de covarianza, y se define como la función continua que representa las (co)varianzas de una característica medida en diferentes puntos de una trayectoria, generalmente tiempo^{9,12,20}. La generalización del modelo de funciones de covarianza en los modelos mixtos, permitió estimar la estructura de covarianza y las ecuaciones de regresión que predicen los valores genéticos de los animales al tiempo que se estiman los efectos sistemáticos temporales^{12,13,14}.

En los modelos de regresión aleatoria y funciones de covarianza, la varianza residual es función tanto de los efectos fijos como del modelo matemático incluido en las regresiones fijas y aleatorias. En los efectos fijos es importante definir; a) el grupo de comparación de los animales y b) las covariables en función del tiempo que cuantifican el efecto general de la etapa de lactancia (regresión fija). En los efectos aleatorios se puede usar el mismo o diferente grupo de covariables para describir el valor genético y los efectos de ambiente permanente (regresiones aleatorias). Son muchos los modelos matemáticos de uso en regresión aleatoria y funciones de covarianza, y se diferencian por el número de coeficientes de regresión, el tipo de covariable, y si la regresión es o no lineal¹⁷. El número de coeficientes de regresión, en la regresión aleatoria es de interés porque determina el de parámetros a estimar en la matriz de coeficientes de la función de covarianza, de gran importancia en análisis multivariado. Así, el uso de la metodología del TDM, exige la definición previa del modelo adecuado a las condiciones particulares.

No son pocos los trabajos relacionados con la definición del modelo adecuado para diferentes países y regiones. Es posible exista consenso en la inclusión del efecto fijo día de registro día de registro, aún cuando no se define con claridad. La comparación de modelos con hato-año-estación o hato-día de registro como efecto fijo (grupo de contemporáneos), mostró reducción de hasta un 30 % con hato-día de registro como grupo de contemporáneos en relación a la combinación hato-año-estación²¹. Resultados en los que se analizó producción de sólo la primera lactancia. Cuando se analiza la producción de las tres primeras lactancias, el análisis generalmente es multivariado, donde la producción diaria en cada lactancia se considera característica longitudinal, y la producción en diferentes lactancias características correlacionadas²². Sin embargo, este último enfoque requiere de la estimación de gran cantidad

de parámetros, complicando la presentación y aplicación de los valores genéticos predichos para los animales. Una alternativa es considerar a la producción diaria de leche en la lactancia, como característica longitudinal y los cambios causados por número de lactancia definirlos ambientales sistemáticos. De esta forma, el número de lactancia debe ser un efecto fijo en el modelo mixto.

Por la importancia de las ecuaciones de regresión, son muchos los estudios realizados para determinar la mejor ecuación de regresión tanto para la parte fija como para los valores genéticos y los efectos de ambiente permanente. Los modelos pueden ser diferentes en cada zona o región, y una parte importante es la factibilidad de uso.

Aún cuando las principales aplicaciones del TDM han sido en la industria lechera, Schaeffer²³ sugirió algunas innovaciones. Un interesante intento de mejorar el ajuste del modelo consistió en incluir curvas aleatorias de hato en Holanda²⁴.

El presente estudio tiene como objetivos: a) comparar diferentes modelos de funciones de covarianza para describir genéticamente la producción diaria de leche y b) estimar los parámetros genéticos de producción diaria de leche durante la lactancia. Para cumplir con los objetivos el trabajo se dividió en; i) revisión bibliográfica de la teoría, que incluye diferentes TDM, y diferencias básicas con los modelos animales “tradicionales”, ii) estimar el ajuste de diferentes ecuaciones lactacionales a la producción diaria de leche, ello con la finalidad de encontrar ecuaciones de regresión de posible uso en TDM y iii) comparar diferentes modelos de funciones de covarianza y estimar parámetros genéticos.

1.1. Literatura citada

1. Bath L. D., F. N. Dickinson, H. A. Tucker, and D. Appleman. Dairy Cattle: Principles Practical, Problems, Profits. Second edition. Lea and Fibger. Philadelphia USA. 495-500. 1978.
2. Ferris A. T. *Selecting for lactation curve sharp and milk yield in dairy cattle*. A dissertation submitted to Michigan State University. Doctor of Philosophy. Department of Dairy Science. Michigan State University. 1981.
3. Wood P. D. P. *Algebraic model of the lactation curves in cattle*. Nature. 216:164-165. 1967.

4. Freeman A. E. *Dairy cattle breeding*. In: 6th Congress on Genetics Applied to Livestock Production, Armidale, Australia. 23:293-294. 1998.
5. Ali T. E., and L. R. Schaeffer. *Accounting for among test day yields in dairy cows*. Can. J. Anim. Sci. 67:637-644. 1987.
6. Reents R., L. Dopp, and F. Reinhardt. *Experience with a genetic evaluation using test day data*. In proceedings of the open session of the 1997 Interbull meeting, Vienna, Austria, August 28-29. 48- 52. 1997.
7. Jamrozik J., L. R. Schaeffer, Z. Liu, and G. Jensen. *Multiple trait random regression test day model for production traits*. In: Proceedings of the open session of the INTERBULL, annual meeting, Vienna, Austria, August 28-29. Bulletin No 16. 43-4. 1997.
8. Tsuruta S., J. F. Keon, L. D. Van Vleck, and I. Misztal. *Bias in genetic evaluation by records of cows treated with bovine somatotropin*. J. Dairy Sci. 83:2650-2656. 2000.
9. Kirkpatrick M., D. Lofsvold, and M. Bulmer. *Analysis of the inheritance, selection and evolution of growth trajectories*. Genetics 124:979-993. 1990.
10. Schaeffer L. R. and J. C. M. Dekkers. *Random regression in animal models for test-day production in dairy cattle*. In 5th World Congress on Genetics Applied to Livestock Production, Guelph Canada. 18:443-446. 1994.
11. Jamrozik J., G. J. Kistemaker, J. C. M. Dekkers, and L. R. Schaeffer. *Comparison of possible covariates used in random regression model for analyses of test day yields*. J. Dairy Sci. 80:2550-2556. 1997.
12. Meyer K., and W. G. Hill. *Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood*. Livest. Prod. Sci. 47:185-200. 1997.
13. Meyer K. *Estimating covariance functions for longitudinal data using random regression model*. Gent. Sel. Evol. 30:221-240. 1998.
14. Meyer K. *Modeling 'repeated' records: covariance functions and random regression models to analyze animal breeding data*. 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia. 25:517-520. 1998.

15. Wiggans G. R., and M. E. Goddard. *A computational feasible test day model for genetic evaluation of yield traits in the United States*. J. Dairy Sci. 80:1795-1800. 1997.
16. Apolaza, L. A., and D. Garrick. Analysis of longitudinal data from progeny test: some multivariate approaches. For. Sci. 47:129-140. 2001.
17. Guo Z., and H. H. Swalve. *Modelling of the lactation curves as sub-model in the evaluation of test day records*. In: Proceedings of the open session of the INTERBULL, annual meeting, Prague, Czech Republic, Sep. 7-8. Bulletin No 11. Topic - session IV. 1-4. 1995.
18. Gengler N., A. Tijani, G. R. Wiggans, C. P. Van Tassell, and J. C. Philpot. *Estimation of (co)variances of test day for first lactation Holstein in the United States* (online only). J. Dairy Sc. 82:225 (abstract). 1999.
19. Brotherstone S., White, I. M. S., and K. Mayer. *Genetic modeling of daily milk using orthogonal polynomials and parametric curves*. Anim. Sci. 70:407-415. 2000.
20. Gengler N., A. Tijani, and G. R. Wiggans. *Use of sequential estimation of regressions and effects on regressions to solve multitrait test-day models*. J. Dairy Sc. (on line only). 2000.
21. Rekaya, R., M. J. Carabaño, and M. A. Toro. Use of test day yields for the genetic evaluation of production traits in Holstein-Friesian cattle. Livest. Prod. Sci. 57:203-217. 1999.
22. Veerkamp R. F., and M. E. Goddard. Covariance functions across herd production levels for test day records on milk, fat, and protein yields. J. Dairy Sc. 81:1690-1701.
23. Schaeffer, L. R. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35-45. 2004.
24. Ross de A.P.W, A.G.F. Harbers, and G. de Jong. Random regression herd curves in a test day model for milk, fat, and protein production of dairy cattle in the Netherlands. J. Dairy Sci. 87:2693-2701.

II. Las ecuaciones de modelos mixtos y el modelo del día de prueba¹

Apodaca Sarabia C. A.²

Resumen

El modelo del día de prueba (TDM) es una metodología con principal aprovechamiento en las evaluaciones genéticas de producción de leche. En el presente trabajo se revisaron las diferentes formas en que se ha usado el TDM, dando énfasis a las relacionadas con la estimación de parámetros genéticos del comportamiento diario y de los efectos temporales sistemáticos en un análisis, así como sus suposiciones e implicaciones biológicas. Los modelos que se han utilizado en el TDM son: repetibilidad, multivariado, regresión aleatoria y funciones de covarianza. Los modelos tienen diferente función de verosimilitud, sin embargo, la flexibilidad que caracteriza a los modelos mixtos es mantenida. En el modelo de repetibilidad se supone que los valores genéticos y las desviaciones de ambiente permanente no cambian durante la lactancia. Los modelos restantes suponen cambios graduales en el valor genético y de ambiente permanente. La diferencia de mayor importancia entre los modelos utilizados en el TDM es el número de parámetros requeridos para describir la estructura de (co)varianza. El modelo multivariado por suponer que las producciones diarias son variables correlacionadas, es impráctico y por la alta correlación entre registros contiguos, es difícil lograr convergencia. Los modelos de regresión aleatoria y funciones de covarianza describen la estructura de (co)varianza con base a funciones continuas. En el modelo de funciones de covarianza la función que describe la estructura día a día, se denomina función de covarianza. Así, se reduce considerablemente el número de parámetros necesarios y aumenta la posibilidad de lograr convergencia de la función de verosimilitud.

Palabras clave: Modelo del día de prueba, modelos mixtos, estructura de covarianza.

¹ Tesis de doctor en ciencias veterinarias. Universidad Autónoma de México. Facultad de Medicina Veterinaria

² apodacacs@correo.chapingo.mx

II. The mixed model equations and the test day model³

Apodaca-Sarabia C. A.⁴

Abstract

The test day model (TDM) is a methodology with main approach in the genetic evaluation of milk production. The objective of this work was to review the different TDM used in animal production, with main attention in the estimation of the genetic parameters of daily production, and systematic no genetic effects as well as its suppositions and biological implications. The models of repeatability, multivariate, random regression and covariance function are used in TDM methodology. These models have different maximum likelihood functions, however, the generality and flexibility of mixed model equations is maintained. The repeatability model assumes constant the genetic and permanent values in the lactation. In the other three models, gradual changes in the genetic and permanent values are assumed. The most important difference between the models applied in TDM methodology, is the numbers of parameters necessary to describe the (co)variance structure. In the multivariate model, due to the supposition that the daily productions are considerate correlated variables, is impractical because the numbers of parameters to estimate is too large and the high correlation between contiguous productions reduce the possibility to get convergence. The random regression and covariance models describe the covariance structure with a continuous function. In the covariance function model, the function used for describe every day, is denominated covariance function. In this way, increases the possibility to get convergence in the maximum likelihood function.

Keywords: Test day models, mixed model, covariance structure.

³ Doctoral thesis in Veterinarian Sciences. Universidad Autónoma de México. Facultad de Veterinaria

⁴ apodacacs@correo.chapingo.mx

2.1. Introducción

En el mejoramiento genético del ganado lechero se han realizado diferentes esfuerzos para aprovechar la variabilidad en producción diaria de leche durante una lactancia, algunos ejemplos son: i) estimar los parámetros genéticos de los coeficientes de regresión de la ecuación de Wood, de producción máxima y de persistencia, para luego, mediante simulación, evaluar la efectividad de la selección para modificar la curva de producción diaria de leche¹, ii) evaluar la eficiencia de diferentes medidas de persistencia para mejorar, indirectamente, el comportamiento reproductivo del hato² y iii) estimar las correlaciones genéticas entre diferentes medidas de persistencia y la eficiencia alimenticia de los hatos productores de leche³.

Las características que se expresan en una trayectoria continua se denominan longitudinales, la producción diaria de leche al expresarse diariamente durante la lactancia es un ejemplo de importancia en producción animal. Se cree que estas características muestran cambios fenotípicos y genéticos conforme avanza la trayectoria, no obstante, en la industria lechera una práctica común es estimar los parámetros genéticos sólo para la producción total por lactancia, producción obtenida al considerar al menos registros mensuales de producción por día.

La flexibilidad y generalidad de los modelos mixtos⁴ y los avances en las técnicas de programación, han hecho posible contar con estimaciones y predicciones con mejores propiedades estadísticas, así como poder solucionar sistemas de ecuaciones simultáneas de gran magnitud, así el uso de modelos más complejos es ahora posible. Uno de estos modelos es el modelo del día de prueba (TDM), y sus principales ventajas son: considerar los efectos ambientales que influyeron sobre la observación el día en que se registró; evitar el cálculo de la producción total; utilizar lactancias incompletas sin necesidad de factores de proyección; y describir el comportamiento genético de la característica durante su expresión⁵.

La metodología del TDM incluye diferentes formas de aplicación algunas de ellas son: **a)** corregir las pesadas mensuales por los efectos ambientales que influyeron sobre estas y, posteriormente, estimar los parámetros genéticos y predecir los valores genéticos para producción por lactancia⁶, **b)** considerar diferentes formas de la matriz de covarianzas de las observaciones (**V**) y ecuaciones de regresión con fines de determinar el mejor modelo en la evaluación genética⁷, **c)** predecir el mérito genético de los animales considerando la producción por día durante la lactancia ya sea como característica repetida⁸ o bien como características

correlacionadas genética y fenotípicamente^{9,10}, **d**) inclusión de regresiones fijas y aleatorias con la idea de modelar tanto la forma general de la curva de lactancia como las variaciones individuales, denominada regresión aleatoria^{11,12}, y **e**) uso de funciones de covarianza para describir las (co)varianzas entre las pesadas mensuales con base en funciones continuas^{13,14}.

De los usos anteriores, en los incisos del **c** al **e**, se busca estimar las (co)varianzas genéticas y predecir el valor genético de producción de leche al tiempo que se corrigen por los efectos no genéticos, y, no obstante de que hacen uso de la metodología de los modelos mixtos tienen importantes diferencias conceptuales entre ellos. El modelo de repetibilidad (MR) supone que el valor genético y los efectos ambientales permanentes no cambian durante la lactancia. En el análisis multivariado (MV), los registros repetidos de producción de leche en una lactancia se consideran características correlacionadas genética y fenotípicamente; además describen sólo parte de la estructura de covarianza y requieren de muchos parámetros. Los modelos de regresión aleatoria (RA) y funciones de covarianza (FCO), asumen que los cambios graduales observados en el fenotipo, son causados, en parte, por cambios en el valor genético y del efecto ambiental permanente. La RA tiene como principal objetivo predecir los valores genéticos de los animales para cada día de lactancia^{7,12,15,16}. Las FCO fueron el siguiente paso, y permiten estimar las (co)varianzas y predecir los valores genéticos durante todo el tiempo de expresión de la característica en un análisis^{13,17,18,19,20}.

Entre las razones para usar los modelos de RA y FCO se encuentran sus ventajas sobre los modelos multivariados, estas son: a) descripción genética continua, del comportamiento diario, sobre el rango de medición^{13,21}, b) mayor exactitud en la predicción de la respuesta a la selección²⁰, y c) uso más eficiente de los datos²⁰.

Son diferentes los países, en la evolución genética nacional utilizan alguna de las formas del TDM. En Canadá, se usa un modelo de RA, y la evaluación genética incluye producción de leche, grasa y proteína, y conteo de células somáticas en cada una de las primeras tres lactancias. Una vez estimadas las ecuaciones de predicción de los valores genéticos para el comportamiento diario, se calcula el valor genéticos de producción por lactancia y persistencia, esta última calculada con la producción los días 60 y 280 de la lactancia^{9,22,23}. La importancia y las diferentes definiciones de persistencia se han discutido por varios autores, entre su importancia se encuentra su relación con eficiencia alimenticia³, resistencia a enfermedades,

estrés fisiológico²⁴, y eficiencia reproductiva^{2,25}. Estas relaciones son las que definen el valor económico de la persistencia, por ejemplo tomando en cuenta sólo su relación con eficiencia alimenticia, se estimó en aproximadamente 5% del valor económico de producción²⁶. Los TDM facilitan el estudio de los componentes de la curva de lactancia y de la relación entre producción de leche, grasa y proteína con balance de energía y reinicio de la actividad ovárica posparto²⁷.

Los modelos de RA y FCO al igual que el MR y el MV, hacen uso de la metodología de los modelos mixtos, sin embargo para lograr las ventajas antes mencionadas, es necesario realizar algunos cambios en las suposiciones de las propiedades de distribución (estructura de (co)varianza), y en la construcción de las ecuaciones de los modelos mixtos. La presente revisión de literatura tiene como objetivo identificar las formas del modelo del día de registro y sus diferencias conceptuales.

2.2. Derivación de las ecuaciones de los modelos mixtos

Dos de las suposiciones de la metodología de cuadrados mínimos e índice de selección son; a) los sementales son una muestra aleatoria de una población, y b) su hijas se distribuyen aleatoriamente en los hatos, suposiciones que en la industria lechera generalmente no se cumplen⁴ e implican ausencia de selección y similitud de los hatos en relación al valor genético de los animales, siendo entonces indispensable encontrar una metodología que tomara en cuenta el efecto de selección y realizara las comparaciones de los animales en forma adecuada. La metodología de los modelos mixtos estima los efectos fijos y predice los valores genéticos al mismo tiempo. En general combina las mejores características de los métodos, índice de selección y cuadrados mínimos^{4,28}.

Considerando el modelo lineal mixto:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \quad (1)$$

Donde para N observaciones, t efectos fijos, y q animales con registros:

y : Es el vector de observaciones, Nx1,

b: Es el vector desconocido de efectos fijos, tx1,

u: Es el vector desconocido de efectos aleatorios, qx1,

X y Z: Son matrices conocidas que relacionan las observaciones con los efectos fijos y aleatorios, de tamaño Nxt y Nxq, respectivamente, y

e: Vector de residuales, Nx1.

Para especificar completamente cualquier modelo es necesaria la definición de sus propiedades de distribución²⁹. Las definiciones usuales para (1) involucran las siguientes esperanzas y variancias³⁰.

Las esperanzas de **y**, **u**, y **e**:

$$E \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ 0 \\ 0 \end{pmatrix} \quad (2)$$

En tanto que las variancias y covariancias de **u** y **e**:

$$\text{var} \begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{pmatrix} \quad (3)$$

De (3) se tiene, $\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$, donde la covarianza entre las observaciones, primero de los sumandos de la ecuación, es causada por efectos aleatorios en común. En mejoramiento genético animal y considerando una característica, $\mathbf{G} = \sigma_a^2 \mathbf{A}$, y $\mathbf{R} = \mathbf{I}\sigma_e^2$, por lo que³⁰;

$$\text{var} \begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \lambda^{-1} \mathbf{A} & 0 \\ 0 & \mathbf{I} \end{pmatrix} \sigma_e^2, \text{ con } \lambda^{-1} = \frac{\sigma_a^2}{\sigma_e^2}$$

Donde:

- A** : matriz de relaciones genéticas aditivas entre los animales,
- σ_a^2 ; variancia genética aditiva de la característica,
- I** : matriz identidad, y
- σ_e^2 : variancia residual.

Las suposiciones impuestas sobre el modelo, ecuaciones (2) y (3), implican:

- i. Los valores genéticos de los animales provienen de la misma distribución y en ausencia de consanguinidad, tienen la misma variancia genética,
- ii. los residuales tiene una variancia común y no están correlacionados, y,

iii. la covarianza entre los efectos aleatorios, \mathbf{u} y \mathbf{e} , es cero.

En el modelo (1), entre los criterios deseables en la predicción de los valores genéticos de los animales, la minimización del cuadrado del error de predicción, es el más deseable^{31,32}. Si en (3) \mathbf{G} y \mathbf{R} son conocidos, es posible obtener una predicción de \mathbf{u} ($\hat{\mathbf{u}}$) con las siguientes propiedades deseables:

- i. Insesgamiento, $E(\hat{\mathbf{u}}) = \mathbf{u}$.
- ii. Entre los predictores lineales, $\hat{\mathbf{u}}$ minimiza la varianza del error de predicción, $\text{var}(\hat{\mathbf{u}} - \mathbf{u})$ es mínima.
- iii. Maximiza la correlación entre el predictor y el valor verdadero, $r_{\hat{\mathbf{u}}\mathbf{u}}$ es máxima.
- iv. Suponiendo distribución normal multivariada:
 - a. Produce el mejor estimador lineal insesgado de máxima verosimilitud de la media condicional del parámetro.
 - b. Entre los estimadores lineales insesgados, maximiza la probabilidad de lograr la jerquización correcta.

Para encontrar el mejor estimador lineal insesgado (BLUE, por sus siglas en ingles) de β , y usarlo en la predicción de \mathbf{u} satisfaciendo las propiedades antes mencionadas, se aplicó la teoría de índice de selección en combinación con el método de cuadrados mínimos. Así, el método busca predecir una combinación lineal de los valores genéticos, $\mathbf{m}'\mathbf{u}$, con base a una combinación lineal de las observaciones, $\mathbf{d}'\mathbf{y}$, tal que:

$$E(\mathbf{d}'\mathbf{y} - \mathbf{m}'\mathbf{u})^2 \text{ sea mínimo, y}$$

$$E(\mathbf{d}'\mathbf{y}) = \mathbf{d}'\beta\mathbf{X}.$$

Aplicando las anteriores restricciones en el modelo (1), el BLUE de β se puede obtener de las ecuaciones de cuadrados mínimos generalizados: $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta^0 = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$, resultando:

$$\beta^0 = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (4)$$

Donde $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$ es la inversa generalizada de $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ pues generalmente no es de rango completo, y β^0 denota una solución. La solución es denotada por β^0 en lugar de β , para indicar que tiene muchas soluciones.

El mejor predictor lineal insesgado (BLUP por sus siglas en ingles) de \mathbf{u} , se puede obtener de:

$$\hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta^0) \quad (5)$$

La expresión para predecir los efectos aleatorios, (5), esencialmente es la regresión de \mathbf{u} sobre \mathbf{y} después de ajustar las observaciones por los efectos fijos, $(\mathbf{y} - \mathbf{X}\beta^0)$. En este método la mayor dificultad radica en obtener la inversa de \mathbf{V} , pues en general es grande y no diagonal.

Un método alternativo para resolver simultáneamente para β^0 y $\hat{\mathbf{u}}$ sin la necesidad de invertir \mathbf{V}^{31} , consiste para el caso del modelo (1), en la predicción de una combinación lineal de los efectos aleatorios mediante una función lineal de las observaciones, $\mathbf{m}'\mathbf{u}$ y $\mathbf{d}'\mathbf{y}$ como se mencionó antes, tal que $E(\mathbf{m}'\mathbf{u} - \mathbf{d}'\mathbf{y})^2$ sea mínima e imponiendo la restricción, $E(\mathbf{d}'\mathbf{y}) = \mathbf{d}'\mathbf{X}\beta = 0$ y $E(\mathbf{u}) = 0^{33}$. El método consiste en derivar la función de densidad conjunta de \mathbf{u} e \mathbf{y} (función de verosimilitud) con respecto a β y $\mathbf{u}^{9,29,31}$. La función de densidad conjunta de \mathbf{u} e \mathbf{y} es:

$$f(\mathbf{y}|\mathbf{u})f(\mathbf{u}) = \frac{\exp\{-1/2((\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{u}) + \mathbf{u}'\mathbf{G}^{-1}\mathbf{u})\}}{(2\pi)^{0.5(t+q)}|\mathbf{R}|^{0.5}|\mathbf{G}|^{0.5}} \quad (6)$$

Derivando (6) con respecto a β y \mathbf{u} e igualando las ecuaciones resultantes a cero, se produce, después de desarrollo algebraico, las ecuaciones de los modelos mixtos:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \beta^0 \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad (7)$$

En (7), para valores conocidos de \mathbf{R} y \mathbf{G} , β^0 y $\hat{\mathbf{u}}$ son el mejor estimador y predictor lineal insesgado de las ecuaciones de cuadrados mínimos generalizados, solo que se incluyó a la inversa de la matriz de residuales (\mathbf{R}) en lugar de la inversa de la matriz de (co)varianzas entre las observaciones (\mathbf{V}). Lo que es una ventaja, pues la estructura de \mathbf{R} , comparada con la de \mathbf{V} , generalmente es más simple y por lo tanto se invierte más fácilmente^{28,33,34}.

La generalidad y flexibilidad de las ecuaciones de los modelos mixtos, así como la demostración de que β^0 y $\hat{\mathbf{u}}$ son el BLUE y BLUP respectivamente, ha sido publicada por varios

autores^{28,29,31,32,35,36,37}. Así, (1) y (7) se aplican a diferentes situaciones, inclusive los modelos para la metodología del día de prueba.

Por ejemplo, si (1) corresponde a una variable y los elementos de $\hat{\mathbf{u}}$ son los valores genéticos de los animales, \mathbf{G}_0 es la varianza genética (σ_a^2) y \mathbf{R} la varianza residual (σ_e^2). La ecuación (6) es entonces, después de multiplicar ambos lados por σ_e^2 :

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + (\sigma_e^2 / \sigma_a^2)\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b}^0 \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Tomando a $\sigma_e^2 / \sigma_a^2 = \lambda$, las ecuaciones de los modelos mixtos son:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \lambda \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b}^0 \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

2.3. Los diferentes modelos del día de registro

2.3.1. Modelo de repetibilidad

Al considerar, por ejemplo, a la producción diaria de leche como una característica registrada repetidamente durante una lactancia, N observaciones, t efectos fijos, q animales de los que sólo p ($p < q$) tienen registros, el modelo (1) se debe escribir:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{p} + \mathbf{e} \quad (8)$$

Donde:

- \mathbf{y} : Vector de observaciones, Nx1,
- \mathbf{X} : matriz de incidencia que relaciona las observaciones con los respectivos niveles de los efectos fijos, Nxt,
- $\boldsymbol{\beta}$: vector desconocido de efectos fijos, tx1,
- \mathbf{Z} : matriz de incidencia que relaciona las observaciones con el valor genético de los animales, Nxq,
- \mathbf{u} : vector desconocido de efectos aleatorio de valores genéticos de los animales, qx1,
- \mathbf{W} : matriz de incidencia que relaciona las observaciones con el efecto de ambiente permanente de los animales, Nxp,

p : vector aleatorio desconocido de efectos de ambiente permanente de los animales, $p \times 1$, y

e : vector de residuales, $N \times 1$.

Con las siguientes propiedades de distribución:

$$\mathbf{E} \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

y

$$\text{var} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\sigma_a^2 & 0 & 0 \\ 0 & \mathbf{I}\sigma_p^2 & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{pmatrix}$$

En la matriz de varianzas, σ_a^2 , σ_p^2 , y σ_e^2 son la varianza del efecto genético, ambiental permanente y residual de producción diaria de leche, respectivamente.

El BLUE de $\boldsymbol{\beta}$ y los BLUP de \mathbf{u} y \mathbf{p} , se pueden obtener simultáneamente, de las ecuaciones de los modelos mixtos de Henderson³⁷:

$$\begin{pmatrix} \mathbf{X}\mathbf{X} & \mathbf{X}\mathbf{Z} & \mathbf{X}\mathbf{W} \\ \mathbf{Z}\mathbf{X} & \mathbf{Z}\mathbf{Z} + \lambda\mathbf{A}^{-1} & \mathbf{Z}\mathbf{W} \\ \mathbf{W}\mathbf{X} & \mathbf{W}\mathbf{Z} & \mathbf{W}\mathbf{W} + \kappa\mathbf{I}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^0 \\ \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\mathbf{y} \\ \mathbf{Z}\mathbf{y} \\ \mathbf{W}\mathbf{y} \end{pmatrix} \quad (9)$$

Donde $\kappa = (\sigma_e^2/\sigma_p^2)$ y λ como se definió anteriormente.

En el modelo de repetibilidad, los cambios en producción diaria de leche al avanzar la lactancia se atribuyen a efectos no genéticos, suponiendo que el valor genético y ambiental permanente no cambian durante la lactancia. Sin embargo, la evidencia indica que los genes que actúan para producción diaria de leche no son los mismos durante toda la lactancia, así, existen cambios en la varianza genética aditiva y en la de efectos ambientales permanentes durante la lactancia^{7,38,39}. De esta forma, si el objetivo es describir la estructura de covarianza de la

producción diaria de leche en una lactancia, el modelo de repetibilidad no es adecuado para describir genéticamente a la producción diaria de leche durante una lactancia.

2.3.2. Modelo multivariado

Una forma de considerar los cambios en la producción diaria de leche en una lactancia, es suponer, las producciones por día en una lactancia, son variables correlacionadas fenotípica y genéticamente, siendo entonces el modelo con dos o más características una de las opciones para su análisis.

El modelo multivariado, ordenando los animales dentro de las características^{33,40,41} se muestra en (1), en donde el vector de observaciones, \mathbf{y} , para t características, está compuesto por los vectores de las t características, $\mathbf{y}' = (\mathbf{y}_1' \mathbf{y}_2' \dots \mathbf{y}_t')$, similarmente los vectores de los efectos fijos (\mathbf{b}) y aleatorios (\mathbf{u}), se componen de los vectores correspondientes de cada variable y las matrices de incidencia, \mathbf{X} y \mathbf{Z} , por matrices de incidencia y nulas. El modelo, escrito en forma ampliada es:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \mathbf{X}_2 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \mathbf{X}_t \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{b}_t \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \mathbf{Z}_2 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \mathbf{Z}_t \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{u}_t \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{e}_t \end{bmatrix} \quad (10)$$

Si el modelo es el mismo para todas las características, el modelo y las ecuaciones de los modelos mixtos se pueden escribir usando el producto directo (\otimes). El modelo se escribe: $\mathbf{y} = \mathbf{I}_t \otimes \mathbf{X} \boldsymbol{\beta} + \mathbf{I}_t \otimes \mathbf{Z} \mathbf{u} + \mathbf{e}$, en el que $\text{Var}(\mathbf{e}) = \mathbf{R} = \mathbf{I}_n \otimes \mathbf{R}_0$, si los residuales están correlacionados sólo entre características medidas en el mismo animal³³, con $\text{Var}(\mathbf{u}) = \mathbf{G} = \mathbf{G}_0 \otimes \mathbf{A}$, donde \mathbf{u} es el vector de valores genéticos aditivos, y:

$$\mathbf{R}_0 = \begin{bmatrix} \sigma_{e1}^2 & \sigma_{ele2} & \cdot & \cdot & \cdot & \sigma_{elet} \\ \sigma_{ele2} & \sigma_{e2}^2 & \cdot & \cdot & \cdot & \sigma_{e2et} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{elet} & \sigma_{e2et} & \cdot & \cdot & \cdot & \sigma_{et}^2 \end{bmatrix}; \mathbf{G}_0 = \begin{bmatrix} \sigma_{g1}^2 & \sigma_{g1g2} & \cdot & \cdot & \cdot & \sigma_{g1gt} \\ \sigma_{g1g2} & \sigma_{g2}^2 & \cdot & \cdot & \cdot & \sigma_{g2gt} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{g1gt} & \sigma_{g2gt} & \cdot & \cdot & \cdot & \sigma_{gt}^2 \end{bmatrix}$$

Las matrices de (co)varianzas residuales y genéticas, respectivamente.

Las ecuaciones de los modelos mixtos se pueden representar como sigue:

$$\begin{bmatrix} \mathbf{R}_0^{-1} \otimes \mathbf{X}'\mathbf{X} & \mathbf{R}_0^{-1} \otimes \mathbf{X}'\mathbf{Z} \\ \mathbf{R}_0^{-1} \otimes \mathbf{Z}'\mathbf{X} & \mathbf{R}_0^{-1} \otimes \mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_0^{-1} \otimes \mathbf{X}'\mathbf{y} \\ \mathbf{R}_0^{-1} \otimes \mathbf{Z}'\mathbf{y} \end{bmatrix} \quad (11)$$

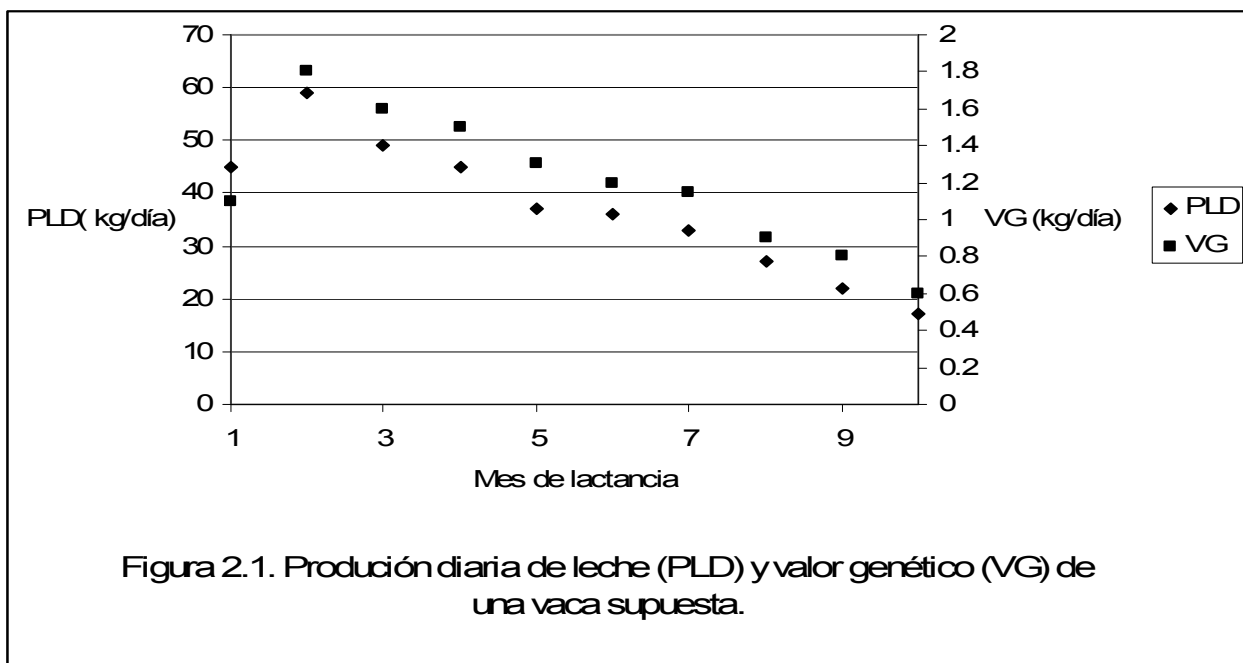
Si los modelos son diferentes para cada una de las características, el modelo y las ecuaciones de los modelos mixtos, no se puede escribir usando el producto directo^{33,41}. Las ecuaciones de los modelos mixtos para dos características son entonces:

$$\begin{bmatrix} \mathbf{X}_1' \mathbf{r}^{11} \mathbf{X}_1 & \mathbf{X}_1' \mathbf{r}^{12} \mathbf{X}_2 & \mathbf{X}_1' \mathbf{r}^{11} \mathbf{Z}_1 & \mathbf{X}_1' \mathbf{r}^{12} \mathbf{Z}_2 \\ \mathbf{X}_2' \mathbf{r}^{21} \mathbf{X}_1 & \mathbf{X}_2' \mathbf{r}^{22} \mathbf{X}_2 & \mathbf{X}_2' \mathbf{r}^{21} \mathbf{Z}_1 & \mathbf{X}_2' \mathbf{r}^{22} \mathbf{Z}_2 \\ \mathbf{Z}_1' \mathbf{r}^{11} \mathbf{X}_1 & \mathbf{Z}_1' \mathbf{r}^{22} \mathbf{X}_2 & \mathbf{Z}_1' \mathbf{r}^{11} \mathbf{Z}_1 + (\sigma_{g1}^2 \mathbf{A})^{-1} & \mathbf{Z}_1' \mathbf{r}^{12} \mathbf{Z}_2 + (\sigma_{g1g2} \mathbf{A})^{-1} \\ \mathbf{Z}_2' \mathbf{r}^{21} \mathbf{X}_1 & \mathbf{Z}_2' \mathbf{r}^{22} \mathbf{X}_2 & \mathbf{Z}_2' \mathbf{r}^{21} \mathbf{Z}_1 + (\sigma_{g1g2} \mathbf{A})^{-1} & \mathbf{Z}_2' \mathbf{r}^{22} \mathbf{Z}_2 + (\sigma_{g2}^2 \mathbf{A})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1' (\mathbf{r}^{11} \mathbf{y}_1 + \mathbf{r}^{12} \mathbf{y}_2) \\ \mathbf{X}_2' (\mathbf{r}^{11} \mathbf{y}_1 + \mathbf{r}^{12} \mathbf{y}_2) \\ \mathbf{Z}_1' (\mathbf{r}^{11} \mathbf{y}_1 + \mathbf{r}^{12} \mathbf{y}_2) \\ \mathbf{Z}_2' (\mathbf{r}^{11} \mathbf{y}_1 + \mathbf{r}^{12} \mathbf{y}_2) \end{bmatrix}$$

El modelo multivariado, al incluir mayor información, debido a las correlaciones entre las características, y corregir por la selección ocurrida en etapas tempranas de la vida del animal, produce mejores predicciones de los valores genéticos de los animales^{33,41}.

De esta forma, para producción diaria de leche en una lactancia, tomando un registro mensual de producción diaria de leche en una lactancia de diez meses, cada animal contará con diez valores genéticos para producción diaria de leche durante la lactancia. La predicción de estos valores genéticos requiere, en las ecuaciones de los modelos mixtos respectivas, de la estimación de 55 componentes de (co)varianza, aumentando considerablemente los requisitos de cómputo y la posibilidad de no lograr convergencia. Esto último es aun más difícil de lograr en características con alta correlación genética entre ellas, como es el caso de producción diaria de leche los últimos tres meses de la lactancia^{39,42,43}. No obstante, una vez obtenida la

solución, las producciones diarias de leche y el valor genético predicho de un animal podrían ser las mostradas en la Figura 2.1.



2.3.3. Modelo de regresión aleatoria

Con base en lo mostrado en la figura 2.1, es posible encontrar una ecuación que permita predecir el valor genético del animal para cualquier día de la lactancia, sin tener que incluir en el análisis las producciones por día en cada día de la lactancia. Así, el valor genético de una característica que se expresa continuamente en un periodo de tiempo dado, se puede representar mediante la ecuación de regresión que predice los valores genéticos en el tiempo de medición y no como el valor genético de la misma para un tiempo dado. La regresión aleatoria permite el uso eficiente de esta información^{44,45}, e implica la posible ventaja sobre el modelo MV al tener la oportunidad de reducir los parámetros necesarios.

En ganado especializado en producción de leche, considerando que se tiene una muestra aleatoria de animales en un hato o conjunto de ellos, las variaciones en su comportamiento durante la lactancia pueden ser también consideradas como aleatorias. Así, si la producción de leche por día se representa por la ecuación de regresión que la predice, la descomposición en sus componentes genéticos y ambientales, es entonces posible y se escribe mediante la ecuación que predice los valores genéticos, la que predice los efectos ambientales permanentes y el residual. Tomando como ejemplo a la ecuación de Guo y Swalve⁴⁶, la

producción de leche el día t de la i -ésima vaca es $y_{it} = b_{0i} + b_{1i} t^{(0.5)} + b_{2i} \ln t + e_{it}$. Los coeficientes de regresión de la ecuación fenotípica se descomponen en, $b_{ki} = \alpha_{ki} + p_{ki}$, donde α_{ki} y p_{ki} son el k -ésimo coeficiente de regresión para el efecto genético y ambiental permanente del i -ésimo animal, respectivamente, para $k = 0, 1, \text{ y } 2$. Sustituyendo los coeficientes fenotípicos y agrupando términos, la ecuación de regresión se escribe⁴⁷:

$$y_{it} = (\alpha_{0i} + \alpha_{1i} t^{(0.5)} + \alpha_{2i} \ln t) + (p_{0i} + p_{1i} t^{(0.5)} + p_{2i} \ln t) + e_i \quad (12)$$

Los primeros tres términos del lado derecho de la ecuación corresponden a la ecuación de predicción para el valor genético, los siguientes tres a la ecuación de predicción de los efectos de ambiente permanentes y el último es el término de error.

El modelo mixto debe ser escrito, entonces, en términos diferentes. La diferencia, consiste escribir, en lugar del valor genético del animal y el efecto de ambiente permanente, la ecuación de regresión que los predice, es decir:

$$y_{ijkt} = f_i + \beta_0 + \beta_1 t^{(0.5)} + \beta_2 \ln t + \alpha_{0j} + \alpha_{1j} t^{(0.5)} + \alpha_{2j} \ln t + p_{0j} + p_{1j} t^{(0.5)} + p_{2j} \ln t + e_{ijk} \quad (13).$$

En (13), y_{ijk} es la producción de leche por día, f_i es el i -ésimo efecto fijo, β_r , y son los coeficientes de regresión ($r = 0, 1, 2$) de la producción por día sobre el día t de la lactancia promedio para todos los datos (regresión fija), α_{rj} , son los coeficientes de regresión aleatoria para predecir el valor genético del j -ésimo animal, p_{rj} son los coeficientes de regresión aleatoria de los efectos de ambiente permanente del j -ésimo animal, y e_{ijk} el error aleatorio⁴⁷.

El modelo (13) escrito en forma matricial se presenta en (14), y es similar al modelo de repetibilidad (8), las diferencias entre ellos son: a) suponer que el valor genético y el ambiental permanente cambian durante la lactancia, y b) las matrices **Z** y **W** correspondientes a los efectos aleatorios referidos en (8), no son de incidencia sino que contienen las covariables usadas en la regresión aleatoria. El modelo entonces:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wp} + \mathbf{e} \quad (14)$$

Donde:

- y**: vector de observaciones, producciones diarias de leche,
- X**: matriz conocida de incidencia y/o covariables,
- b**: vector desconocido de efectos fijos incluyendo los coeficientes de regresión fija,

- Z**: matriz de las covariables, raíz cuadrada de días en lactancia y su logaritmo natural, que relaciona las observaciones con el coeficiente aleatorio de regresión para el valor genético de cada animal,
- u**: vector desconocido de coeficientes de regresión aleatoria del valor genético de cada animal,
- W**: matriz de las covariables, raíz cuadrada de días en lactancia y su logaritmo natural para cada animal con registro, que relaciona las observaciones con el coeficiente de regresión aleatorio para el efecto ambiental permanente de cada animal,
- p**: vector desconocido de coeficientes de regresión aleatoria del efecto ambiental permanente de cada animal con registro, y,
- e**: vector de residuales.

Con: $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$ y $\text{Var}(\mathbf{y}) = \mathbf{Z}(\mathbf{G}_0 \otimes \mathbf{A})\mathbf{Z}' + \mathbf{W}(\mathbf{I} \otimes \mathbf{P}_0)\mathbf{W}' + \mathbf{R}$.

En regresión aleatoria, los elementos de \mathbf{G}_0 son las (co)varianzas entre los coeficientes de regresión aleatoria de los valores genéticos y los elementos de la matriz \mathbf{P}_0 son las (co)varianzas entre los coeficientes de regresión de los efectos ambientales permanentes. Así, su orden depende de la ecuación de regresión usada en cada uno de los efectos aleatorios, 3 para el usado en el ejemplo de la partición del fenotipo^{12,47}.

Las ecuaciones de los modelos mixtos para el modelo de regresión aleatoria en (14) son:

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{I} \otimes \mathbf{P}_0^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{a} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \quad (15).$$

Las ecuaciones de los modelos mixtos para un caso particular, donde el modelo incluyó el efecto de ható-día de prueba como fijo (\mathbf{b}), y para predecir el valor genético, efecto aleatorio, incluyeron días en lactancia (γ_1) y el logaritmo natural de 305/días en lactancia (γ_2), además del valor genético del animal (\mathbf{a})¹². El modelo fue:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_1\gamma_1 + \mathbf{Z}_2\gamma_2 + \mathbf{e}.$$

Y las ecuaciones de los modelos mixtos:

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_a & \mathbf{X}'\mathbf{Z}_1 & \mathbf{X}'\mathbf{Z}_2 \\ \mathbf{Z}_a'\mathbf{X} & \mathbf{Z}_a'\mathbf{Z}_a + \mathbf{A}^{-1}\lambda_{aa} & \mathbf{Z}_a'\mathbf{Z}_1 + \mathbf{A}^{-1}\lambda_{a1} & \mathbf{Z}_a'\mathbf{Z}_2 + \mathbf{A}^{-1}\lambda_{a2} \\ \mathbf{Z}_1'\mathbf{X} & \mathbf{Z}_1'\mathbf{Z}_a + \mathbf{A}^{-1}\lambda_{a1} & \mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{A}^{-1}\lambda_{11} & \mathbf{Z}_1'\mathbf{Z}_2 + \mathbf{A}^{-1}\lambda_{12} \\ \mathbf{Z}_2'\mathbf{X} & \mathbf{Z}_2'\mathbf{Z}_a + \mathbf{A}^{-1}\lambda_{a2} & \mathbf{Z}_2'\mathbf{Z}_1 + \mathbf{A}^{-1}\lambda_{12} & \mathbf{Z}_2'\mathbf{Z}_2 + \mathbf{A}^{-1}\lambda_{22} \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{a} \\ \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_a'\mathbf{y} \\ \mathbf{Z}_1'\mathbf{y} \\ \mathbf{Z}_2'\mathbf{y} \end{pmatrix} \quad (16)$$

Donde, $\lambda_{ij} = g^{ij}\sigma_e^2$ con g^{ij} elemento de \mathbf{G}_0^{-1} y σ_e^2 la varianza residual de todas las producción de leche por día. Agrupando los efectos aleatorios en un vector, de forma que incluya el valor genético de producción por día y los dos coeficientes de regresión aleatoria $\mathbf{u}' = (\mathbf{a}' \gamma_1' \gamma_2')$, resulta el modelo para una variable de respuesta con valor genético como efecto aleatorio, sin embargo \mathbf{G}_0 no es un escalar, sino una matriz con las covarianzas entre los elementos de \mathbf{u} . Las ecuaciones de los modelos mixtos son:

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \quad (17)$$

Regresando, con fines de lograr mayor generalización, al modelo de regresión aleatoria con el valor genético y los efectos de ambiente permanente, las matrices \mathbf{W} y \mathbf{Z} , son, con excepción de las columnas de cero de los animales sin observaciones en \mathbf{Z} , diagonales en bloques. Cada bloque, corresponde a un animal con registros. Así, el número de columnas en \mathbf{Z} es igual al número de animales con y sin registro por el número de coeficientes de regresión en la ecuación para predecir el valor genético de los animales y en \mathbf{W} es igual al producto del número de animales con registro por el de coeficientes de regresión respectivo. En ambas el número de renglones es igual al de registros en las observaciones¹². La construcción de las matrices de diseño para regresión aleatoria se muestra en el Anexo 1 mediante un ejemplo hipotético.

2.3.4. Modelo de funciones de covarianza

El modelo mixto de funciones de covarianza se puede obtener de los modelo MV⁴⁸ y RA^{18,20,49} mediante cambios en las varianzas de ambos modelo. En el MV, ordenando las observaciones por animal y tiempo de registro, el valor genético y el efecto de ambiente permanente, se pueden expresar con el polinomio de Legendre adecuado, así, resulta un modelo univariado en cual se incluye el tiempo para tomar en cuenta los cambios graduales en la expresión de la característica y se supone; $\text{var}(\Phi_i \mathbf{u}_i) = \Phi_i \text{var}(\mathbf{u}_i) \Phi_i = \Phi_i \mathbf{K}_G \Phi_i = \mathbf{G}_0$, y, $\text{Var}(\Phi_i \mathbf{p}_i) = \Phi_i \text{var}(\mathbf{p}_i) \Phi_i =$

$\Phi_i K_P \Phi_i = P_0$. Es decir, la función de covarianza propuesta por Kirkpatrick *et al.*¹³ aplicada a los efectos genético y de ambiente permanente, respectivamente.

Para obtener las funciones de covarianza con el modelo de RA^{18,20,48} primero se incluyen los polinomios de Legendre en las regresiones aleatorias de los efectos genéticos y de ambiente permanente, así, la covarianza entre dos registros, y_{ij} e $y_{i'j'}$ es: $Cov(y_{ij}, y_{i'j'}) = \sum_{m=0}^{k-1} \sum_{m'=0}^{k-1} \phi_m(t_i) \phi_{m'}(t_{i'}) Cov_G(\beta_m, \beta_{m'}) + \sum_{m=0}^{k-1} \sum_{m'=0}^{k-1} \phi_m(t_i) \phi_{m'}(t_{i'}) Cov_P(\beta_m, \beta_{m'}) + Cov(e_i, e_{i'}) = B_G(t_i, t_{i'}) + B_P(t_i, t_{i'}) + Cov(e_i, e_{i'})$, donde, $B_i(t_j, t_{j'})$ es la función de covarianza¹³ para los efectos genéticos y de ambiente permanente, $i = G$ y P . Es importante hacer notar, que tanto con las modificaciones descritas para el MV como en RA, es posible usar, no solo polinomios de Legendre, sino cualquier ecuación de regresión que se considere adecuada, generalizando así el uso de las funciones de covarianza.

El modelo mixto resultante cuando se incluyen el efecto genético aditivo y el ambiental permanente es:

$$\mathbf{y} = \mathbf{Xb} + \Phi \mathbf{u} + \Phi \mathbf{p} + \mathbf{e} \quad (18)$$

Donde:

- y** : Vector de observaciones, $N \times 1$,
- X** : Matriz conocida que relaciona los efectos fijos con las observaciones,
- Φ** : Matriz conocida que relaciona las observaciones con los coeficientes de regresión aleatoria de los valores genéticos y de ambiente permanente,
- b** : Vector desconocido de los efectos fijos,
- u** : Vector desconocido de coeficientes de regresión de los valores genéticos,
- p** : Vector desconocido de coeficientes de regresión de los efectos de ambiente permanente,
- e** : Vector de residuales.

Con las siguientes propiedades de distribución:

$$E \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{ y } \text{var} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A} \otimes \mathbf{G}_0 & 0 & 0 \\ 0 & \mathbf{I} \otimes \mathbf{P}_0 & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{pmatrix} = \begin{pmatrix} \mathbf{A} \otimes \mathbf{K}_G & 0 & 0 \\ 0 & \mathbf{I} \otimes \mathbf{K}_P & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{pmatrix}$$

En las que, como fue antes mencionado: $\text{var}(\Phi_i \mathbf{u}_i) = \Phi_i \text{var}(\mathbf{u}_i) \Phi_i = \Phi_i \mathbf{K}_G \Phi_i = \mathbf{G}_0$, y, $\text{Var}(\Phi_i \mathbf{p}_i) = \Phi_i \text{var}(\mathbf{p}_i) \Phi_i = \Phi_i \mathbf{K}_P \Phi_i = \mathbf{P}_0$. De esta forma, las matrices de (co)varianzas en la función de verosimilitud restringida y en las ecuaciones del modelo mixto, se pueden sustituir por las matrices de coeficientes de la función de covarianza respectiva, \mathbf{K}_G y \mathbf{K}_P , y las ecuaciones de los modelos mixtos, debe entonces escribirse:

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\Phi & \mathbf{X}'\mathbf{R}^{-1}\Phi \\ \Phi'\mathbf{R}^{-1}\mathbf{X} & \Phi'\mathbf{R}^{-1}\Phi + \mathbf{K}_G^{-1} \otimes \mathbf{A}^{-1} & \Phi'\mathbf{R}^{-1}\Phi \\ \Phi'\mathbf{R}^{-1}\mathbf{X} & \Phi'\mathbf{R}^{-1}\Phi & \Phi'\mathbf{R}^{-1}\Phi + \mathbf{I} \otimes \mathbf{K}_P^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{a} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \Phi'\mathbf{R}^{-1}\mathbf{y} \\ \Phi'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \quad (19)$$

Así mismo, en el modelo y las ecuaciones de los modelo mixtos, los elementos de Φ son los coeficientes del polinomio de Legendre respectivo al tiempo de registro. Al igual que en RA, Φ es diagonal en bloques, y la matriz correspondiente al vector de valores genéticos tiene columnas de ceros correspondientes a los animales sin registro. Por último, es de gran importancia que el número parámetros a estimar depende del grado del polinomio incluido en el modelo, y en general se busca sea menor al número de veces que se pueda registrar la característica.

2.3.5. Diferencias entre los modelo del día de registro

Las diferencias entre las distintas formas de modelar las características longitudinales radican en las suposiciones hechas sobre las propiedades de distribución, en particular sobre la estructura de las matrices de (co)varianzas de los efectos aleatorios. Estas, producen diferentes ecuaciones de los modelos mixtos y tienen implicaciones sobre la interpretación biológica y el número de (co)varianzas que deben estimarse. Como fue antes mencionado, los modelos mixtos factibles de uso en los TDM son de repetibilidad, multivariado, regresión aleatoria y funciones de covarianza, y las (co)varianzas de los efectos aleatorios de ellos y el número de parámetros a estimar se muestran en el cuadro 2.1. En este, sólo se presenta el número de parámetros a estimar para un efecto aleatorio, así en los modelos multivariado, regresión

aleatoria y funciones de covarianza, es general la inclusión de al menos dos efectos, el efecto genético y el de ambiente permanente, y el número de (co)varianzas se duplica.

Cuadro 2.1. Varianzas de los efectos aleatorios número de parámetros a estimar.

Modelo	Varianza de:			Parámetros a estimar ¹	Observaciones
	u	p	e		
Repetibilidad	$\mathbf{A} \sigma_a^2$	$\mathbf{I} \sigma_p^2$	$\mathbf{I} \sigma_e^2$	3	Varianza genética y ambiental permanente sin cambio.
Multivariado	$\mathbf{G}_0 \mathbf{A}$	$\mathbf{P}_0 \mathbf{I}$	$\mathbf{I}_n \otimes \mathbf{R}_0$	$(t(t+1)/2)^a$	Covarianzas sólo en cortes de “tiempo” de la expresión de la característica.
Regresión aleatoria	$\mathbf{G}_0 \mathbf{A}$	$\mathbf{P}_0 \mathbf{I}$	$\mathbf{I} \sigma_e^2$	$(Nb(Nb-1)/2)^b$	Los elementos de \mathbf{G}_0 y \mathbf{P}_0 son las covarianzas de los coeficientes de regresión aleatoria.
Funciones de covarianza	$\Phi_i \mathbf{K}_G \Phi_i$	$\Phi_i \mathbf{K}_P \Phi_i$	$\mathbf{I} \sigma_e^2$	$(Nb(Nb-1)/2)^b$	Describe la estructura genética en todo el tiempo de expresión

¹por efecto aleatorio; ^at= al número de características en el modelo; ^bNb= número de coeficientes de regresión el la regresión aleatoria respectiva.

Las ventajas de modelo de funciones de covarianza sobre los modelos de repetibilidad y multivariado son: a) modelar con mayor exactitud la estructura de (co)varianza, b) permitir la predicción de las (co)varianzas en cualquier punto de la escala continua, registrado o no en las observaciones, c) optimización de uso de los registros tomados en cualquier momento de la trayectoria de expresión sin la necesidad de corregir a un momento de la trayectoria de expresión, d) requieren de menos parámetros para estimar las (co)varianzas de los efectos aleatorios y e) provee una metodología de análisis del patrón de (co)varianza asociado con cambios de la característica durante la trayectoria de su expresión^{13,17,48,50,51}.

2.4. Estimación de componentes de varianza

En las ecuaciones de los modelos mixtos es necesario conocer \mathbf{A}^{-1} y estimar las matrices \mathbf{G}_0 y \mathbf{R}_0 . La matriz de relaciones aditivas se obtiene de la información en los registros de genealogía. En general \mathbf{A} es grande y no diagonal, obtener por lo tanto su inversa en forma directa implica gran demanda de cómputo, sin embargo, Henderson⁵² y Quaas⁵³ desarrollaron reglas simples para su inversión. El primero cuando la población no es consanguínea y el segundo para poblaciones donde existe apareamiento entre parientes. La estimación de \mathbf{G}_0 y \mathbf{R}_0 es preferible hacerla con el método de máxima verosimilitud restringida (REML), pues tiene mejores propiedades estadísticas.

Por didáctica, se revisará primero el método de máxima de verosimilitud (ML), y de esta forma establecer la justificación de la restricción. En ambos métodos, es básico suponer una distribución de densidad de las observaciones. Así mismo, por definición, en ambos se involucra la localización del máximo de la función de verosimilitud de un conjunto de datos con un modelo de análisis y los parámetros a estimar^{29,54,55}.

2.4.1. Máxima Verosimilitud

Considerado el modelo más simple, $y_i = \mu + e_i$ con $E(y_i) = \mu$, $E(e) = 0$ y $\text{Var}(e) = \sigma^2$ (para $i = 1, 2, \dots, n$), bajo normalidad, la función de probabilidades de y_i es:

$$f(y_i | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mu)^2}{\sigma^2}} \quad (20)$$

La función conjunta de densidades de probabilidad, función de verosimilitud⁵⁶, es:

$$\mathbf{f}(y_1, y_2, \dots, y_n | \mu, \sigma^2) = \left(\frac{1}{\sigma^n (2\pi)^{n/2}} \right) e^{-\frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}} \quad (21)$$

Para la estimar μ y σ^2 como función de las observaciones, los estimadores de ML de μ y σ^2 son los valores $\hat{\mu}$ y $\hat{\sigma}^2$, que maximizan la función de verosimilitud y que están en el espacio permisible de los parámetros. Así, es necesario encontrar el máximo de la función, máximo que es más fácil de encontrar si se realiza sobre el logaritmo natural de la función.

Derivando el logaritmo de la función de verosimilitud con respecto a μ y σ^2 , e igualando a cero las ecuaciones resultantes, se produce, $\hat{\mu} = \bar{y}$ y un estimador sesgado de la varianza ($\hat{\sigma}^2$), dado que está dividido entre n y no entre $n-1$, el grado de libertad correspondiente a la estimación de la media, el efecto fijo⁵⁶. El sesgo debido a los efectos fijos, se obtiene sin importar el tipo de modelo de que se trate, como fue observado en la derivación de los componentes de varianza de (1)²⁹. En la literatura existe una estrategia de cómputo para estimar los componentes de varianza con ML⁵⁷.

2.4.2. Máxima verosimilitud restringida

Con el objetivo de eliminar el sesgo en la estimación de la varianza producido en ML, REML estima los componentes de varianza basándose en la función de verosimilitud de los residuales después de ajustar por los efectos fijos, o bien de funciones lineales de las observaciones libres de los efectos fijos, \mathbf{Ky} , denominadas contrastes de error. La variable resultante tiene, $E(\mathbf{Ky}) = \mathbf{KE}(\mathbf{y}) = \mathbf{KX}\beta = 0$, por lo cual $\mathbf{Ky} \sim (0, \mathbf{KVK}')$, y el logaritmo de la función de verosimilitud es:

$$l = -\frac{1}{2}(N - r(\mathbf{X})) \log 2\pi - \frac{1}{2} \log |\mathbf{KVK}'| - \frac{1}{2} \mathbf{y}' \mathbf{K}' (\mathbf{KVK}')^{-1} \mathbf{Ky} \quad (22).$$

En el primer elemento de la ecuación, al número de observaciones se le resta $r(\mathbf{X})$, el rango de \mathbf{X} , que corresponde a los grados de libertad en la estimación de los efectos fijos, de forma tal que la derivada sólo se realiza sobre los componentes de varianza. Igualando a cero y solucionando con respecto a los componentes de varianza, resulta en estimadores libres de efectos fijos e insesgados al tomar en cuenta los grados de libertad mencionados. Al obtener la derivada, las ecuaciones a solucionar para estimar los componentes de varianza son²⁹:

$$\{\text{tr}(\mathbf{PV}_i \mathbf{PV}_j)\} \sigma^2 = \{\mathbf{y}' \mathbf{PV}_i \mathbf{Py}\} \quad (23)$$

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

Aplicando REML al modelo $y_i = \mu + e_i$, con las esperanzas definidas en ML, el logaritmo natural de la función de verosimilitud, es⁵⁵:

$$l = -1/2(n - 1) \ln 2\pi - 1/2 |\mathbf{K}' \mathbf{K} \sigma^2|^{-1/2} \mathbf{y}' \mathbf{K} (\mathbf{K}' \mathbf{K} \sigma^2)^{-1} \mathbf{K}' \mathbf{y}.$$

Derivando con respecto a σ^2 , igualando a cero y simplificando, se obtiene:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n y_i - \frac{(\sum_{i=1}^n y_i)^2}{n}}{n-1}$$

2.4.3. Funciones de covarianza

Supóngase se estimaron las covarianzas genéticas de producción diaria de leche en los días, 75, 135 y 195, así se estimaron las (co)varianzas de algunos puntos o cortes de la estructura de covarianza genética de una variable de expresión continua en la lactancia, y podemos estar interesados en su estimación completa. La figura 2.2 muestra la estructura de covarianza en los días antes mencionados y en el periodo entre el día 5 y 195 de la lactancia.

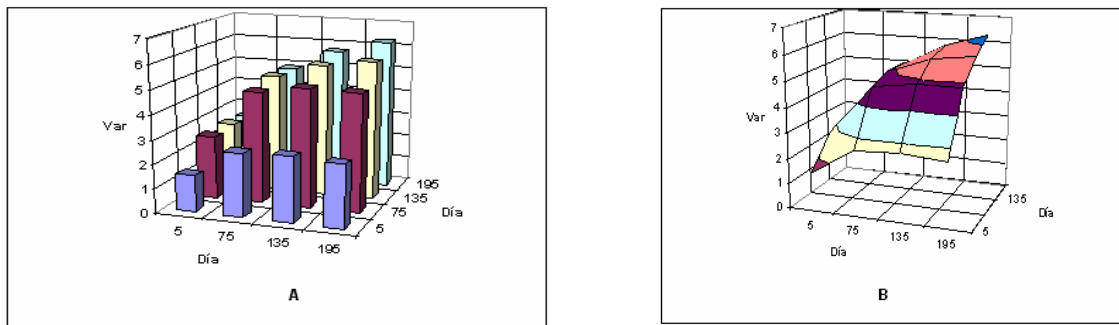


Figura 2.2. Covarianzas entre diferentes días de producción de leche en forma discreta (A) y continua (B).

Para obtener las covarianzas en forma continua, figura 2.2 B, es posible usar regresión, y en este caso se conoce como función de covarianza^{13,21}. El procedimiento para obtener la función de covarianza a partir de una estimación previa de parte de la matriz de (co)varianza genética ha sido descrito por varios autores^{13,17,21,27,47}. Su generalización a los modelos mixtos fue rápida y permitió la aplicación de REML en la estimación de las funciones de covarianza. En esta sección se describirá el procedimiento inicial, la función de máxima verosimilitud restringida resultante del modelo mixto y su solución se discutirá en la siguiente sección.

La función de covarianza fue descrita mediante: $G(a_i, a_j) = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} k_{ij} \phi(a_i) \phi(a_j)$; donde ϕ es el i -ésimo polinomio de Legendre ($i = 0, 1, \dots, k-1$) de orden k , a_i es tiempo estandarizado con valores en el intervalo entre -1 y 1 que se obtiene con $a_i = [2(t_i - t_{\text{mínimo}})/(t_{\text{máximo}} - t_{\text{mínimo}})] - 1$, y k_{ij} los coeficientes de la función de covarianza. Escrita matricialmente, la función de covarianza

es: $\hat{G} = \phi K \phi'$, donde; los elementos de \hat{G} son estimaciones de (co)varianzas genéticas, ϕ y K son matrices de los coeficientes del polinomio de Legendre y coeficientes de la función de covarianza respectivamente. El orden de ϕ y K es igual al grado del polinomio usado en la función. Así, los coeficientes de la función de covarianza serán entonces^{13,47} $K = \phi^{-1} \hat{G} \phi^{-T}$.

Una descomposición del polinomio, de gran utilidad cuando se usa notación matricial, es: $\phi = M\Lambda$; los elementos de M son $m_{ij} = t_i^{(j-1)}$ y Λ es la matriz de coeficientes de los polinomios de Legendre. La función de covarianza es entonces $\hat{G} = M\Lambda K(M\Lambda)'$ y su ventaja es obtener de forma más sencilla la matriz de coeficientes del polinomio de Legendre⁴⁸.

El anterior desarrollo de la función de covarianza es valido para polinomio de orden completo, igual al número de veces que se registró la característica, no obstante, si la característica es de expresión continua, el número de veces que se puede registrar llega a ser muy grande, y la función de covarianza de orden completo produce matrices de alto orden y por lo tanto difíciles de inversión directa. Siendo entonces de importancia el desarrollo de funciones de covarianza de orden menor al número de veces que se registró la característica, orden reducido ($k < t$), sin perdida importante de ajuste¹³.

Para obtener la función de covarianza de orden reducido, es necesario escribir los elementos de \hat{G} y K , en los vectores \hat{g} y k , cuyos elementos son los de cada columna de la matriz respectiva, por ejemplo: $\hat{g}' = [G_0(1,1), \dots, G_0(t,1), G_0(1,2), \dots, G_0(t,2), \dots, G_0(t,t)]$, y expresarse en el modelo $g = Xk + e$, donde; X la matriz que relaciona los coeficientes de la función de covarianza con las (co)varianzas estimadas, y e el vector de residuales. En el modelo anterior, por la simetría de G y K , los vectores g y k , deben contener sólo los elementos correspondientes a la diagonal principal y bajo esta, así, contienen $t(t+1)/2$ y $k(k+1)/2$ elementos respectivamente. Las anteriores modificaciones implican eliminar los renglones de X correspondientes a $G(i,j)$ para $i < j$, y sumar las columnas $K(i,j)$ para $i < j$ con su columna correspondiente $K(j,i)$ ¹³.

La función de covarianza, estimada mediante cuadrados mínimos¹³ es, $\hat{k} = (X'X)^{-1}X'g$. Sin embargo, por la estructura de la matriz de covarianzas, es más apropiado el uso de cuadrados mínimos generalizados, la solución entonces es: $\hat{k} = (X'V^{-1}X)^{-1}X'V^{-1}g$ ⁴⁸. En el Anexo 2 se ejemplifica la estimación de la función de covarianza de orden completo y reducido con un caso

hipotético de producción de leche. En este se describen en detalle los cálculos de los coeficientes del polinomio de Legendre usando notación matricial⁴⁸.

2.5. La función de máxima verosimilitud de los diferentes modelos y su solución

2.5.1. Análisis univariado

Las ecuaciones resultantes al derivar las funciones de verosimilitud para estimar los componentes de varianza no son lineales, las varianzas se encuentran tanto del lado izquierdo como del lado derecho, así, su solución es más factible mediante métodos numéricos.

Los trabajos de Smith y Graser³⁴ y Graser *et al.*⁵⁸ produjeron las bases de los actuales algoritmos de uso común en la solución de la función de verosimilitud. Inicialmente el algoritmo fue propuesto para el modelo univariado y con sólo un efecto aleatorio. Posteriormente, se propuso su ampliación a más efectos aleatorios⁵⁴.

En el modelo (1) para un análisis univariado y como efecto aleatorio el valor genético aditivo, vector de observaciones con distribución NID($\mathbf{X}\beta$, \mathbf{V}), y esperanzas y varianzas: $E(\mathbf{u}) = 0$, $E(\mathbf{e})=0$, $\mathbf{R}=\mathbf{I}\sigma_e^2$, $\mathbf{V}=(\mathbf{ZAZ}'\lambda+\mathbf{I}\sigma_e^2)$, $\lambda = \sigma_a^2/\sigma_e^2$, $\hat{\sigma}_e^2 = (y'Py)/(N-r(X))$, así:

$$\text{var} \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_a^2 & 0 & \sigma_a^2\mathbf{AZ}' \\ 0 & \mathbf{R} & \mathbf{R} \\ \mathbf{ZA}\sigma_a^2 & \mathbf{R} & \mathbf{V}\sigma_e^2 \end{bmatrix}$$

Considerando las ecuaciones de los modelos mixtos para el modelo, descritas en (6), el procedimiento de Smith y Graser³⁴ y Graser *et al.*⁵⁸, requiere ampliar la matriz de coeficientes con el vector del lado derecho y la suma de cuadrados total, produciéndose:

$$\mathbf{M} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{y}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad (24)$$

El logaritmo natural de la función de verosimilitud del modelo es, $-2\log L = \ell = \text{Constante} + \log|\mathbf{R}| + \log|\mathbf{G}| + \log|\mathbf{C}| + \mathbf{y}'\mathbf{P}\mathbf{y}$, en la que los dos últimos términos se evalúan, en \mathbf{M} , absorbiendo por renglón y columna, primero a los efectos aleatorios y luego los fijos, para obtener: a) la

suma de cuadrados residual ponderada ($\mathbf{y}'\mathbf{P}\mathbf{y}$), b) el logaritmo del determinante de la matriz de coeficientes, después de aplicar las restricciones ($|\mathbf{C}|$), y, c) el rango de \mathbf{X} . El logaritmo de la matriz de coeficientes y el rango de \mathbf{X} , se obtienen del proceso de absorción y son igual a la suma de los logaritmos de los elementos de la diagonal y el número de pivotes diferentes de cero, respectivamente³⁴.

El logaritmo de $|\mathbf{R}|$ depende de su estructura, en análisis univariado, generalmente se supone que \mathbf{R} es igual a $\mathbf{I}\sigma_e^2$, y en este caso, el logaritmo de su determinante es igual a $N\log\sigma_e^2$, donde N es el número de observaciones. La matriz \mathbf{G} , para el análisis univariado y un efecto aleatorio, es $\mathbf{A}\sigma_a^2$, el logaritmo de su determinante es entonces, el número de niveles del factor aleatorio (NA) por el logaritmo de la varianza genética, más el logaritmo del determinante de la matriz \mathbf{A} ($NA\log\sigma_a^2 + \log|\mathbf{A}|$). El determinante de \mathbf{A} , no depende de los parámetros a estimar y es constante para un problema particular, así, en los métodos numéricos sólo requiere ser calculado una vez y se evalúa, por su tamaño, con procedimientos indirectos^{52,53}.

La ampliación a más efectos aleatorios es directa, por ejemplo: considerando un modelo con dos efectos aleatorios, sean el genético y el de ambiente común, el vector de efectos aleatorios es $\mathbf{u}' = (\mathbf{a}' \ \mathbf{c}')$ con \mathbf{c} como el vector adicional con número de renglones igual al de niveles del efecto (NC), $\text{var}(\mathbf{c}) = \mathbf{I}\sigma_c^2$ y covarianza entre ambos igual a cero, de forma que la matriz \mathbf{G} es diagonal con elementos $\mathbf{A}\sigma_a^2$ y $\mathbf{I}\sigma_c^2$. El logaritmo de su determinante es por lo tanto, $NA\log\sigma_a^2 + NC\log\sigma_c^2 + \log|\mathbf{A}|$ ⁵⁶.

2.5.2. Análisis multivariado

El análisis mediante modelos para varias características o multivariado, logra mayor exactitud al incluir un mayor volumen de información y la corrección parcial de los efectos de selección cuando ésta se realizó para una de las características. Sin embargo, sus desventajas son; a) aumento en los costos de cómputo, b) rápido aumento en el número de parámetros a estimar, y c) disminución en la posibilidad de lograr convergencia al aumentar en el número de parámetros a estimar^{59,60}.

Para reducir el tiempo de cómputo y aumentar la posibilidad de lograr convergencia, se usan transformaciones. Algunas transformaciones son; canónica, Cholesky y diagonalización de las ecuaciones de los modelos mixtos, ya sea a la forma tridiagonal o diagonal^{59,60,61}.

El objetivo de la transformación canónica es obtener un grupo de variables en el que las covarianzas entre ellas sean cero, ello sin perder la información contenida en las variables originales; las variables resultantes se denominan variables canónicas. La transformación canónica se aplica si las matrices de diseño son las mismas y todos los animales tienen registro en todas las características⁶².

El propósito de la transformación Cholesky es convertir la matriz \mathbf{A}^{-1} en la matriz idéntica, de forma que se ahorre tiempo al simplificar las ecuaciones como en el caso de animales no emparentados. La transformación Cholesky se aplica sobre los datos con transformación canónica, previa absorción de las ecuaciones para los efectos fijos. Las transformaciones a ecuaciones tridiagonales o diagonales se realizan con la finalidad de ahorrar tiempo, al realizar la transformación de la matriz de coeficientes a diagonal se obtiene la solución de las ecuaciones de los modelos mixtos. Los procedimientos detallados para hacer las transformaciones fueron publicados por Núñez y Rodríguez⁶⁰. Por su parte, Jensen y Mao⁵⁹ publicaron las ecuaciones para estimar los componentes de varianza de las variables transformadas.

En el análisis multivariado se supone que los residuales entre las características están correlacionados sólo si miden en el mismo animal, la matriz \mathbf{R} es entonces diagonal en bloques. El número de bloques en \mathbf{R} es igual al de animales en las observaciones. Los bloques en \mathbf{R} , son submatrices (\mathbf{R}_0) de las covarianzas residuales entre las características para cada animal, así para t características, estas son orden t . Los elementos de \mathbf{R}_0 varían según las suposiciones sobre los datos, la forma más simple es cuando las covarianzas se suponen cero, la submatriz \mathbf{R}_0 es entonces diagonal, y el determinante de \mathbf{R} es fácil de encontrar. En cambio, si las covarianzas son diferentes de cero, el problema es más complicado y se puede agrupar en a) cuando todos los animales tienen registros en las t características, y b) si los animales tienen registros perdidos⁴¹.

Si no existen registros perdidos, la matriz \mathbf{R} es diagonal en bloques donde estos son la matriz \mathbf{R}_0 tantas veces como animales se tenga, y las transformaciones antes mencionadas son aplicables si el modelo es mismo para todas características. Cuando se tiene registros faltantes, las submatrices \mathbf{R}_0 difieren por el registro faltante. Por ejemplo si se tienen dos características, los animales pueden tener registro en las dos o solo en una de ellas; en general, para t

características se esperan $2^t - 1$ submatrices \mathbf{R}_0 . El número de veces que una combinación está en \mathbf{R} es igual al número de animales que tiene dicha combinación, y la matriz \mathbf{R} es la suma directa de las matrices \mathbf{R}_0 posibles^{41,52}.

La estructura del algoritmo del modelo univariado⁵⁴ permitió su aplicación al modelo multivariado⁶³. La modificación consistió en además de \mathbf{R}_0 , definir a la matriz \mathbf{T} , de orden rt , donde r es el número de factores aleatorios en el modelo (además del error) con las covarianzas entre estos. Es decir, el algoritmo permite modelos que incluyan no sólo al efecto genético, sino además por ejemplo, el de ambiente permanente, materno o ambos. Así, $\mathbf{y}'\mathbf{P}\mathbf{y}$ y $\log|\mathbf{C}|$ se calculan como se describió anteriormente, según Smith y Graser³⁴, y el cálculo de los logaritmos de los determinantes de las matrices \mathbf{R} y \mathbf{G} , es como sigue:

1. El $\log|\mathbf{R}| = \sum_{i=1}^W \mathbf{B}_i \log |\mathbf{R}_{0i}|$, donde W es el número de combinaciones de submatrices \mathbf{R}_0 en los datos y, B el número de animales con la i -ésima combinación de características registradas.
2. El logaritmo del determinante de \mathbf{G} depende de los factores aleatorios incluidos en el modelo, así para:
 - a) Solo el genético, entonces $\mathbf{G} = \mathbf{A} * \mathbf{T}$ y $\log|\mathbf{G}| = NA \log |\mathbf{T}| + t \log|\mathbf{A}|$
 - b) El genético (\mathbf{T}_A) y el ambiente común (\mathbf{T}_C) con NC niveles, entonces $\mathbf{G} = \text{diag}\{\mathbf{A}*\mathbf{T}_A; \mathbf{D}*\mathbf{T}_C\}$ y $\log|\mathbf{G}| = NA \log|\mathbf{T}_A| + NC \log|\mathbf{T}_C| + t [\log|\mathbf{A}| + \log|\mathbf{D}|]$, donde \mathbf{D} es diagonal, comúnmente la idéntica.

Esto último permitió la disponibilidad de algoritmos para estimar componentes de varianza usando el modelo lineal general mediante REML para una gran cantidad de situaciones, algunos de estos algoritmos se encuentran en los programas MTDFREML⁴² y DFREML⁶⁴.

2.5.3. Modelo de funciones de covarianza

La generalización de las funciones de covarianza en el modelo mixto se describió anteriormente, y (18) y (19), son el modelo mixto y las ecuaciones de los modelos mixtos respectivamente. La función de máxima verosimilitud restringida del modelo de funciones de covarianza, se ha expresado de diferentes formas. Partiendo del modelo multivariado, primero se obtuvo:

$$l = -\frac{1}{2}[\text{cte.} + N \ln |\mathbf{K}_p| + N_A \ln |\mathbf{K}_a| + t \ln |\mathbf{A}| + \ln |\mathbf{C}| + (N + N_A) \ln |\Phi \Phi'| + \mathbf{y}' \mathbf{P} \mathbf{y}] \quad (25)$$

En (25), \mathbf{K}_p es la matriz de coeficientes de la función de covarianza entre los efectos ambientales permanentes y temporales, e igual a $\mathbf{P}_0 + \mathbf{R}_0$, esta última definida, en funciones de covarianza, como medidas de error y por lo general igual a $\mathbf{I}\sigma_e^2$, sin embargo, es posible la inclusión de ambas en la función de verosimilitud, así, se obtuvo la segunda expresión de la función de verosimilitud restringida²⁰:

$$l = -\frac{1}{2}[\text{cte.} + N \ln |\Phi \mathbf{K}_{P_0} \Phi'| + \text{Diag}\{\sigma_{ei}^2\} + N_A \ln |\mathbf{K}_a| + t \ln |\mathbf{A}| + \ln |\mathbf{C}| + (N + N_A) \ln |\Phi \Phi'| + \mathbf{y}' \mathbf{P} \mathbf{y}] \quad (26)$$

Las funciones de verosimilitud (25) y (26) son útiles para polinomios de orden completo y reducido. Si el orden del polinomio es reducido, el término, $(N + N_A) \ln |\Phi \Phi'|$, debe reemplazarse por $N_A \ln |\Phi_a \Phi_a'| + N \ln |\Phi_e \Phi_e'|$ ello debido a que el número de columnas en Φ es diferente al de renglones, o bien, escribir las matrices de coeficientes de tamaño t con ceros en las columnas y renglones correspondientes a los coeficientes del polinomio 'eliminados' para tener un polinomio de menor grado y maximizar sólo la parte correspondiente al grado del polinomio incluido²⁰.

Una función de verosimilitud restringida diferente a (25) y (26), se puede obtener con el modelo de regresión aleatoria usando polinomios de Legendre como covariables, el modelo es:

$$y_{ij} = F + \sum_{m=0}^{k-1} \alpha_{im} \phi_m(t_{ij}^*) + \sum_{m=0}^{k-1} \gamma_{im} \phi_m(t_{ij}^*) + \varepsilon_{ij}$$

Donde y_{ij} es la observación del i -ésimo animal en el j -ésimo tiempo, α_{im} y γ_{im} representan los coeficientes de regresión aleatoria del animal i -ésimo para el valor genético y de ambiente permanente respectivamente, k el orden del polinomio, t^* es el tiempo estandarizado según el polinomio de Legendre, y F los efectos fijos que incluyen las diferencias sistemáticas entre las observaciones (hato, día de registro por ejemplo) y los coeficientes de regresión fijos. La regresión fija se incluye con la finalidad de modelar la trayectoria promedio de la población, y las regresiones aleatorias modelan las desviaciones de los animales con respecto a ella^{18,49}.

En el anterior modelo, la covarianza entre dos registros, es:

$$\text{Cov}(y_{ij}, y_{i'j'}) = \sum_{m=0}^{k-1} \sum_{l=0}^{k-1} \phi_m(t_{ij}^*) \phi_l(t_{i'j'}^*) \text{Cov}(\alpha_{im}, \alpha_{il}) + \sum_{m=0}^{k-1} \sum_{l=0}^{k-1} \phi_m(t_{ij}^*) \phi_l(t_{i'j'}^*) \text{Cov}(\gamma_{im}, \gamma_{il}) + \text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) \quad (27)$$

Los dos primeros términos del lado derecho de (27) son las funciones de covarianza de los efectos genético y ambiental permanente. Así, el modelo de funciones de covarianza, proporciona una alternativa estratégica de estimar las funciones de covarianza^{18,48}.

La matriz de coeficientes de las ecuaciones de los modelos mixtos para el modelo de regresión aleatoria con polinomios de Legendre ampliada con el lado derecho del las mismas y la suma de cuadrados residual, usada por Smith y Graser³⁴ para los modelos univariados, es^{18,48}:

$$\mathbf{M} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{K}_a^{-1} \otimes \mathbf{A}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{W} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{K}_p^{-1} \otimes \mathbf{I}_N & \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{y}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{W} & \mathbf{y}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Las matrices \mathbf{K}_A y \mathbf{K}_P con elementos $\text{Cov}(\alpha_m, \alpha_l)$ y $\text{Cov}(\gamma_m, \gamma_l)$, denotan las matrices de coeficientes de las funciones de covarianza de los efectos, genético y de ambiente permanente respectivamente, y asumiendo; $\alpha \sim (\mathbf{0}, \mathbf{K}_A \otimes \mathbf{A})$, $\gamma \sim (\mathbf{0}, \mathbf{K}_P \otimes \mathbf{I}_N)$, α y γ no correlacionadas y para mayor simplicidad, el mismo error en todo el periodo de medición ($\text{Var}(\varepsilon) = \mathbf{R} = \mathbf{I}\sigma_e^2$). En \mathbf{M} , las matrices \mathbf{Z} y \mathbf{W} , corresponden a las matrices de diseño del los valores genéticos y de ambiente permanente, respectivamente. La Matriz \mathbf{M} es cuadrada y su tamaño es la suma de: el número de niveles de los efectos fijos, el número de total de animales por el número de coeficientes de regresión en el efecto genético, el número de animales con registro por número de coeficientes en el efecto ambiental permanente, mas uno (correspondiente al renglón, o columna, adicionado por el lado derecho del las ecuaciones).

La función de verosimilitud restringida correspondiente del sistema de ecuaciones en \mathbf{M} , es:

$$\ell = -\frac{1}{2}(N_a \log|\mathbf{K}_A| + k_A \log|\mathbf{A}| + N_p \log|\mathbf{K}_P| + \log|\mathbf{C}| + \log|\mathbf{R}| + \mathbf{y}'\mathbf{P}\mathbf{y}) \quad (28)$$

Una diferencia importante entre (25) y (26) con (28) es que las dos primeras incluyen $|\Phi\Phi'|$, de esta forma requieren de la evaluación del periodo de registro, incrementase considerablemente los requisitos de computo. En tanto, (28) es igual a la función de verosimilitud del MRA, y los determinantes que incluye son de matrice con orden igual al del polinomio en los efectos aleatorios (\mathbf{K}_A y \mathbf{K}_P) mientras \mathbf{C} , las ecuaciones de los modelos mixtos, son proporcionales al orden de polinomio en los efectos aleatorios^{12,20,23,48}.

La función de verosimilitud (28) se maximiza para las matrices de coeficientes de las funciones de covarianza usando las estrategias de búsqueda aplicadas a los modelos animal como son

los métodos; simplex, Powell (direcciones conjugadas), o casi Newton^{18,49}, con la ventaja de requerir un número considerablemente menor de parámetros ha estimar, cuando el número de coeficientes de regresión es pequeño en comparación con el de veces que se registró la característica.

2.6. Solución de las ecuaciones de los modelos mixtos

Las estimaciones de los componentes de varianza es necesaria para resolver las ecuaciones de los modelos mixtos. Entre las soluciones se encuentran las predicciones de los valores genéticos de los animales, que, para selección, son los criterios de selección. Sin embargo debido a que las ecuaciones de los modelos mixtos, son un sistema de ecuaciones con un gran número de ellas, es importante contar con una estrategia para su solución. La absorción, eliminación de Gauss, descomposición Cholesky, y los métodos iterativos son de uso común en la solución de las ecuaciones de los modelos mixtos. En la práctica, los paquetes de cómputo pueden involucrar más de una de las técnicas referidas³⁰. Una descripción de cada una de ellas se proporciona a continuación.

2.6.1. Absorción

Para reducir el número de ecuaciones de un sistema de ecuaciones es posible absorber una o más ecuaciones en las restantes. Las ecuaciones de los modelos mixtos se pueden escribir en dos grupos de ecuaciones, uno que hace referencia a los efectos fijos y el otro a los efectos aleatorios³⁰. Las ecuaciones de los modelos mixtos, (7), son entonces:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\mathbf{b}^0 + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \quad (29a)$$

$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\mathbf{b}^0 + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\mathbf{u}} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \quad (29b)$$

De la ecuación (29a), el vector de efectos fijos se puede escribir:

$$\mathbf{b}^0 = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{Z}\hat{\mathbf{u}})$$

Sustituyendo \mathbf{b}^0 en (29b), se tiene una solución para el vector de efectos aleatorios:

$$\hat{\mathbf{u}} = (\mathbf{Z}'\mathbf{P}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{P}\mathbf{y} \quad (30)$$

En este procedimiento, los efectos fijos son absorbidos en los aleatorios⁶⁵. No obstante, predecir los efectos aleatorios, en (30), se debe invertir $(\mathbf{Z}'\mathbf{P}\mathbf{Z} + \mathbf{G}^{-1})$, matriz de orden igual al número de animales (matriz muy grande), y por lo tanto su inversión directa requiere de mucho

tiempo de cómputo y es sensible a errores de redondeo, problemas que deben evitarse mediante otras técnicas de cómputo³⁰.

2.6.2. Eliminación de Gauss

Las ecuaciones de los modelos mixtos, (7), se pueden escribir $\mathbf{Cb} = \mathbf{r}$, donde la matriz \mathbf{C} es la matriz de coeficientes, \mathbf{b} es el vector cuyos elementos son los que desean estimarse e incluye los niveles de los efectos fijos y los valores genéticos de los animales y \mathbf{r} es el vector del lado derecho de las ecuaciones.

Una modificación a la eliminación de Gauss es la factorización triangular, también llamada descomposición LU^{66,67,68}. Aplicada a la matriz de coeficientes, se tiene $\mathbf{C} = \mathbf{LU}$, con \mathbf{L} matriz triangular inferior y unos en la diagonal principal, y \mathbf{U} triangular superior. Las ecuaciones de los modelos mixtos entonces son:

$$\mathbf{LUb} = \mathbf{r}$$

Haciendo $\mathbf{Ub} = \mathbf{t}$, el sistema original se descompone en dos ecuaciones,

$$\mathbf{Lt} = \mathbf{r}, \text{ y}$$

$$\mathbf{Ub} = \mathbf{t}$$

Sistemas de ecuaciones que por su estructura, es fácil solucionar. Este método es conocido como sustitución hacia adelante y hacia atrás³⁰.

2.6.3. Descomposición Cholesky

Una metodología similar a la de Gauss, es la descomposición Cholesky, y tiene base en la simetría de la matriz de coeficientes. La descomposición Cholesky consiste en descomponer a la matriz de coeficientes en dos matrices triangulares $\mathbf{C} = \mathbf{TT}^T$, con \mathbf{T} triangular inferior y sus elementos, $\{t_{ij}\}$, se obtiene para $i = 1, 2, \dots, n$ y $j = 1, 2, \dots, n$, mediante la siguiente regla^{33,60,66,67,68}:

$$t_{ii} = (c_{ii} - \sum_{k=1}^{i-1} t_{ik}^2)^{1/2}$$

y

$$t_{ij} = (c_{ij} - \sum_{k=1}^{j-1} t_{ik} t_{jk}) / t_{jj}$$

El de ecuaciones, sustituyendo a \mathbf{C} por \mathbf{TT}^T , es:

$$\mathbf{b} = \mathbf{r}$$

Si $\mathbf{T}^T\mathbf{b} = \mathbf{z}$, el sistema original se descompone en dos sistemas de ecuaciones:

$$\mathbf{Tz} = \mathbf{r}, \text{ y,}$$

$$\mathbf{T}^T\mathbf{b} = \mathbf{z}$$

Solucionando primero para \mathbf{z} para luego obtener la solución para \mathbf{b} , estas son hacia delante y atrás respectivamente. El método de Cholesky no necesariamente mantiene la simetría del sistema de ecuaciones, además es demandante de memoria de cómputo y no asegura la convergencia³⁰.

2.6.4. Métodos iterativos

Entre los métodos iterativos de mayor uso en la solución de las ecuaciones de los modelos mixtos se encuentran Gauss-Seidel y Jacobi. La sobre y subrelajación, y la iteración por bloques, son una modificación a Gauss-Seidel^{60,66,67,68}.

Gauss-Seidel descompone la matriz de coeficientes en la suma de tres matrices, $\mathbf{C} = (\mathbf{L} + \mathbf{D} + \mathbf{L}')$, con \mathbf{L} triangular inferior y \mathbf{D} diagonal, el sistema de ecuaciones es entonces: $(\mathbf{L} + \mathbf{D} + \mathbf{L}')\mathbf{b} = \mathbf{r}$, o bien;

$$(\mathbf{L} + \mathbf{D})\mathbf{b} = \mathbf{r} - \mathbf{L}'\mathbf{b}.$$

El sistema de ecuaciones del esquema Gauss-Sediel, es un esquema iterativo por pasos, en el cual la última solución del vector \mathbf{b} se usa en la obtención de la siguiente. El esquema de Jacobi tiene su base en una modificación al de Gauss-Sediel, para ello se escribe el sistema de ecuaciones de Gauss-Seidel como $\mathbf{Db} = \mathbf{r} - \mathbf{L}'\mathbf{b} - \mathbf{Lb}$. Sistema que después de desarrollo algebraico, produce $\mathbf{b}^{(k+1)} = \mathbf{D}^{-1}\mathbf{r} - \mathbf{D}^{-1}(\mathbf{L}' + \mathbf{L})\mathbf{b}^k$.

La representación de ambos esquemas en forma escalar fue publicada por López³⁰ y Núñez y Rodríguez⁶⁰ y se puede encontrar en libros de análisis numérico^{66,68}. El método de Jacobi no actualiza la información más reciente en el vector de soluciones, sino que actualiza el vector completo, siendo una diferencia importante con Gauss-Seidel. El método de Jacobi no garantiza se logre convergencia⁶⁹.

La sobre y subrelajación son modificaciones de Gauss-Seidel con fines de acelerar el método para lograr convergencia y consiste en incluir el parámetro θ . Cuando el parámetro es menor a uno, el método es subrelajación, o bien sobrelajación si es mayor a uno³⁰.

La iteración por bloques es otra de las modificaciones de Gauss-Seidel y se considera de gran utilidad cuando los bloques se obtienen de forma natural como en el modelo multivariado³³.

Una dificultad de los métodos iterativos es cuando parar y aceptar la solución. El criterio más usado es un punto estandarizado, $p = (\mathbf{e}'\mathbf{e})^{0.5}/(\mathbf{r}'\mathbf{r})$, con $\mathbf{e} = \mathbf{r} - \mathbf{C}\mathbf{b}^n$, y \mathbf{b}^n la solución en la iteración n-ésima. La literatura muestra varios valores de p , sin embargo en la predicción del valor genético, se deben tomar los más pequeños^{40,42,70}.

Los métodos numéricos o iterativos se agrupan en los que usan derivadas y los libres de éstas, la elección de uno de ellos para la solución de un sistema de ecuaciones es función de: a) los requerimientos de cómputo, b) el número de iteraciones para lograr convergencia, c) facilidad de programación, d) facilidad para ser generalizado a diferentes modelos, e) sensibilidad a los valores iniciales, y f) capacidad para converger a un máximo global⁶⁰.

2.7. Comparación de modelos

La elección del modelo en el caso del TDM tiene serias limitantes si los modelos involucran diferente número de parámetros, no obstante existen pruebas con diferente confiabilidad según el modelo que se trate. Para funciones de covarianza con estimación previa de (co)varianzas¹³, dos son las pruebas para determinar el grado del polinomio de Legendre adecuado: a) iniciando con el polinomio de mayor grado posible según el orden de la matriz de covarianzas genéticas disponible, estimación completa, y posteriormente encontrar entre los polinomios de menor orden, estimación reducida, la de menor grado que no difiera estadísticamente de la estimación completa, y b) iniciar con polinomios de de orden bajo, sea k (H_0) y aumentar el grado del polinomio en una a la vez, $k+1$ (H_1), para tomar el polinomio de menor orden que no difiera del orden siguiente. Los estadísticos de prueba son: $\chi^2_{(m-p)} = (\hat{\mathbf{e}}'\hat{\mathbf{e}})' \hat{\mathbf{V}}^{-1}(\hat{\mathbf{e}}'\hat{\mathbf{e}})$, donde $m-p$ son los grados de libertad para la prueba con m y p el número de coeficientes en el polinomio de orden completo y reducido respectivamente, y $\hat{\mathbf{e}}$ el vector de residuales¹³, y $F_{p,gle} = (wSSE_{H_0} - wSSE_{H_1})/wSSE_{H_1}$, donde $wSSE$ es la suma de cuadrados del error ponderada bajo cada uno de los modelos, gle los grados de libertad del error en H_0 y $p = [(k + 1)(k + 2)/2 - (k + 1)k/2] = k + 1$,

el número de coeficientes a estimar con el modelo bajo H_1 ⁴⁸, respectivamente. La ventaja del último, consiste en iniciar con un polinomio de orden reducido para luego aumentarlo, no requiriéndose estimar matrices grandes desde el inicio.

En comparación de modelos que involucran efectos fijos, componentes de varianza, y la predicción de efectos aleatorios como el valor genético, las pruebas involucran: a) el vector de residuales, b) la función de verosimilitud y c) el número de parámetros estimados.

El método más simple de comparación sólo usa la varianza residual estimada, eligiendo el modelo con el valor más pequeño. Por la tendencia general de los modelos con más parámetros a tener menor varianza residual, el método debe ser usado sólo en la comparación entre modelos con igual número de parámetros estimados. La correlación entre el valor predicho y el observado y el porcentaje de la suma de cuadrados del error con respecto a la suma de cuadrados total, se usaron para comparar modelos con diferente número de covariables en la regresión y suposiciones sobre la matriz de (co)varianza de las observaciones. El mejor modelo además de menor varianza residual debió tener la mayor correlación⁷. El valor de la función de verosimilitud también se ha usado como criterio de ajuste¹¹.

La razón de verosimilitud, los criterios de información de Akaike, Deviance y Bayesiana, son métodos que hacen uso del valor de la función de verosimilitud, los tres últimos requieren, además, el número de parámetros estimados y el de observaciones en el modelo. La razón de verosimilitud es un método formal de comparación, y en general se compara el modelo completo, contra modelos reducidos, sin embargo esta prueba tiende a favorecer a los modelos complejos, con muchos parámetros^{19,25,56}. El criterio de información de Deviance (DIC) es asintótico y refleja bondad de ajuste y grado de parametrización. El método por ser asintótico requiere que el número de observaciones crezca más rápido que el de parámetros en el modelo, así, no es recomendable usarse en el modelo animal⁷¹.

Los criterios de información Akaike, Deviance y Bayesiana tratan de corregir por la diferencia en el número de parámetros estimados y de observaciones entre los modelos. Los valores para Akaike (AIC_k) y Bayesiana (SC_k) son, para el k-ésimo modelo, con p_k parámetros estimados y N observaciones: $AIC_k = -2\ln l + 2p_k$, y $SC_k = -2\ln l + 2p_k \log(N)$. El número de observaciones es incluido por su influencia en el valor de la función de verosimilitud⁴⁷. Una forma diferente de SC_k

consistió en tomar en cuenta los efectos fijos, así, $SC_k = \ln l - 1/2p_k n^*$; donde $n^* = N - p_f$, y p_f es el número de efectos fijo en el modelo²⁵.

La validación cruzada es de gran utilidad. Esta consiste en dividir en dos grupos las bases de datos. Los parámetros se estiman con cada modelo con un subgrupo y en el otro se usan las estimaciones, comparando los cuadrados medios del error de cada modelo⁷¹.

No obstante, aún cuando se considere que se eligió el modelo con el procedimiento estadístico adecuado, se debe considerar la consistencia biológica y si el nivel de cómputo requerido es práctico. Para ello, una vez determinado el modelo, debe verificarse su validez continuamente, así, se deben responder a las preguntas, ¿el valor genético de los animales con primera evaluación tienden a sobreestimarse? y ¿su evaluación decrece al aumentar la información sobre ellos? Si es así, el modelo es inadecuado. Responder las anteriores preguntas requiere de alto entrenamiento y dedicación en la especie evaluada, así, el personal que realice las evaluaciones debe dedicarse a pocas especies y contar con el punto de vista del ganadero, de lo contrario aumenta la posibilidad de no notar posibles errores atribuibles al modelo⁷¹.

2.8. Literatura citada

1. Ferris A. T. *Selecting for lactation curve sharp and milk yield in dairy cattle*. Ph. D. Thesis. Michigan State University. 1981.
2. Bar-Anan R., M. Ron, and G. R. Wiggans. *Association among milk yield, yield persistency, conception and culling of Israeli Holstein dairy cattle*. J. Dairy Sci. 68:382-386. 1985.
3. Sölkner J., and W. Fuchs. *A comparison of different measures of persistency with special respect to variation of test-day milk yields*. Livest. Prod. Sci. 16:305-319. 1987.
4. Henderson C. R. *General flexibility of linear model techniques for sire evaluation*. J. Dairy Sci. 57:963-972. 1974.
5. Freeman A. E. *Dairy cattle breeding*. In 6th Congress on Genetics Applied to Livestock Production, Armidale, Australia. 23:293-294. 1998.
6. Trus D., and L. G. Buttszoni. *A multiple trait approach to modeling the lactation curve*. In Proc. 4th World congress on Genetics Applied to livestock Production, Edinburgh. 14:492-495. 1994.
7. Ali T. E., and L. R. Schaeffer. *Accounting for among test day yields in dairy cows*. Can. J. Anim. Sci. 67:637-644. 1987.
8. Reents R., L. Dopp, and F. Reinhardt. *Experience with a genetic evaluation using test day data*. In proceedings of the open session of the 1997 Interbull meeting, Vienna, Austria, August 28-29. 48- 52. 1997.
9. Jamrozik J., G. J. Kistemaker, J. C. M. Dekkers, and L. R. Schaeffer. *Comparison of possible covariates used in random regression model for analyses of test day yields*. J. Dairy Sci. 80:2550-2556. 1997.
10. Tsuruta S., J. F. Keon, L. D. Van Vleck, and I. Misztal. *Bias in genetic evaluation by records of cows treated with bovine somatotropin*. J. Dairy Sci. 83:2650-2656. 2000.
11. Jakobensen J. H., P. Madsen, J. Jensen, J. Pedersen, L. G. Christensen, and D. A. Sorensen. *Genetic parameters for milk production and persistency for Danish Holstein estimated in random regression models using REML*. J. Dairy Sc. 85:1607-1616. 2002.

12. Schaeffer L. R. and J. C. M. Dekkers. *Random regression in animal models for test-day production in dairy cattle*. In 5th World Congress on Genetics Applied to Livestock Production, Guelph Canada. 18:443-446. 1994.
13. Kirkpatrick M., D. Lofsvold, and M. Bulmer. *Analysis of the inheritance, selection and evolution of growth trajectories*. Genetics 124:979-993. 1990.
14. Wiggans G. R., and M. E. Goddard, *A computational feasible test day model for genetic evaluation of yield traits in the United States*. J. Dairy Sci. 80:1795-1800. 1997.
15. Ptak E., and L. R. Schaeffer. *Use of test day yields for genetic evaluation of dairy sires and cows*. Livest. Prod. Sci. 34:23-34. 1993.
16. Wilmlink J. B. M. *Adjusted of test-day milk, fat and protein yields for age, season and stage of lactation*. Livest. Prod. Sci. 16:335-348. 1987.
17. Kirkpatrick M., W. G. Hill, and R. Thompson. *Estimating the covariance structure of traits during growth and ageing, illustrated with lactation dairy cattle*. Genet. Res. 64:57-69. 1994.
18. Meyer K. *Modeling 'repeated' records: covariance functions and random regression models to analyze animal breeding data*. 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia. 25:517-520. 1998.
19. Meyer K. *Random regressions to model phenotypic variation in monthly weights of Australian beef cows*. Livest. Prod. Sci. 65:19-38. 2000.
20. Meyer K., and W. G. Hill. *Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood*. Livest. Prod. Sci. 47:185-200. 1997.
21. Kirkpatrick M., and N. Heckman. *A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters*. J. Math. Biology. 27:429-450. 1989.
22. Jamrozik J., L. R. Schaeffer, and J. C. M. Dekkers. *Genetic evaluation of dairy cattle using test day yields and random regression model*. J. Dairy Sci. 80:1217-1226. 1997.
23. Schaeffer L. R., J. Jamrozik, G. J. Kistemaker, and B. J. Van Doormaal. *Experience with test-day model*. J. Dairy Sci. 83:1135-1144. 2000.

24. Wood P. D. P. *Algebraic model of the lactation curves in cattle*. Nature. 216:164-165. 1967.
25. Jensen J. *Genetic evaluation of dairy cattle using test-day models*. J Dairy Sci. 84:2803-2812. 2001.
26. Dekkers J. C. M., J. H. Ten Hag, and A. Weersink. *Economic aspect of persistency of lactation in dairy cattle*. Livest. Prod. Sci. 53:237-252. 1998.
27. Veerkamp R. F., J. K. Oldenbroek, H. J. Van Der Gasst, and J. H. Van Der Werf. *Genetic correlation between days until start of luteal activity and milk yield, energy balance, and live weight*. J. Dairy Sci. 83:577-583. 2000.
28. Henderson, C. R., O. Kempthorne, S. R. Searle, and C. W. Von Krosigk. *The estimation of environmental and genetic trends from records subject to culling*. Biometrics. 15:192-218. 1959.
29. Searle S. R., G. Casella, and C. E. McCulloch. *Variance Components*. John Wiley & Son. New York, USA. 232-257. 1992.
30. López V. N. *Accounting for scale effects on genetic evaluation of New Zealand dairy cattle*. Master of Agricultural Science thesis. Massey University, Palmerston North, New Zealand. 1994.
31. Henderson C. R. *Sire evaluation and genetics trends*. In: Proceedings of the animal breeding and genetics symposium in honor of Dr. Jay L. Lush. 10-41. 1973.
32. Henderson C. R. *Best linear unbiased estimation and prediction under a selection model*. Biometrics. 31:423-447. 1975.
33. Quaas R. L., R. D. Anderson, and A. R. Gilmore. *BLUP school handbook, use of mixed model for prediction and estimation of (co)variance components*. Animal genetics and breeding unit. University of New England, Australia. 1- 77. 1984.
34. Smith S. P., and H. U. Graser. *Estimating variance components in a class of mixed models by restricted maximum likelihood*. J. Dairy Sci. 69:1156-1165. 1986.
35. Becerril C. M. P. *Un modelo simple para ilustrar la estimación de parámetros y la predicción de los valores genéticos*. In 5th Meeting of International Biometric Society

Network for Central America, the Caribbean, México, Colombia, and Venezuela. Universidad Veracruzana, Xalapa, México. 423-427. 1997.

36. Henderson C. R. *Use of all relatives in intraherd prediction of breeding values and producing abilities*. J. Dairy Sci. 58:1910-1916. 1975.
37. Henderson C. R. *Use of relationships among sires to increase accuracy of sire evaluation*. J. Dairy Sci. 58:1731-1738. 1975.
38. Brotherstone S., White, I. M. S., and K. Mayer. *Genetic modeling of daily milk using orthogonal polynomials and parametric curves*. Anim. Sci. 70:407-415. 2000.
39. Gengler N., A. Tijani, G. R. Wiggans, C. P. Van Tassell, and J. C. Philpot. *Estimation of (co)variances of test day for first lactation Holstein in the United States (online only)*. J. Dairy Sci. 82:225 (abstract). 1999.
40. Blair H. T., and E. J. Pollak. *Comparison of an animal model and an equivalent reduced animal model for computational efficiency using mixed model methodology*. J. Anim. Sci. 58:1090-1096. 1984.
41. Henderson C. R., and R.L. Quaas. *Multiple trait evaluation using relatives records*. J. Anim. Sci. 43:6:1188-1197. 1976.
42. Boldman K. G., L. A. Krieses, L. D. Van Vleck, and S. D. Kachman. *A Manual for use of MTDFREML*. USDS-ARS, Clay Center, NE. 120 p. 1993.
43. Rekaya R., F. Béjar, M. J. Carbaño, and R. Alenda. *Genetic parameters for test day measurements in Spanish Holstein-Friesian*. In proceedings of the open session of the Interbull annual meeting, Prague, Czech Republic, September 7-8, 1995. 1-8. 1995.
44. Henderson C. R. Jr. *Analysis of covariance in the mixed model: Higher-level, nonhomogeneous, and random regressions*. Biometrics. 28:623-640. 1982.
45. Snedecor G. W., y W. G. Cochran. *Métodos estadísticos*. Sexta impresión. CECOSA, México. 191-195. 1979.
46. Guo Z., and H. H. Swalve. *Modelling of the lactation curves as sub-model in the evaluation of test day records*. In: Proceedings of the open session of the INTERBULL,

- annual meeting, Prague, Czech Republic, Sep. 7-8. Bulletin No 11. Topic - session IV. 1-4. 1995.
47. Schaeffer L. R. *Random regression*. Course notes, Lesson 14. CGIL, Guelph, June 28, 1997. Canada. 1-20. 1997.
48. Van Der Werf, J. H., and L. R. Schaeffer. *Random regression in animal breeding*. Course notes. CGIL, Guelph, June25-June28, 1997, Canada. pp 59. 1997.
49. Meyer K. *Estimating covariance functions for longitudinal data using random regression model*. *Genet. Sel. Evol.* 30:221-240. 1998.
50. Koenen E. P. C., and R. F. Veerkamp. *Genetic covariance functions for live weight, condition score, and dry-matter intake measured at different lactation stage of Holstein – Friesian heifers*. *Livest. Prod. Sci.* 57:67-77. 1998.
51. Olori V. E., W. G. Hill, B. J. McGuirk, and S. Brotherstone. *Estimating variance components for test day records by restricted maximum likelihood with random regression animal model*. *Livest. Prod. Sci.* 61:53-63. 1999.
52. Henderson C. R. *A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values*. *Biometrics.* 32:69-83. 1976.
53. Quaas R. L. *Computing the diagonal elements and inverse of a large numerator relationship matrix*. *Biometrics.* 32:949-953. 1976.
54. Meyer K. *Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm*. *Genet. Sel. Evol.* 21:317-340. 1989.
55. Rodríguez F. A., y R. Núñez D. *Componentes de varianza*. En: Memorias del curso intensivo, 'Introducción a la metodología de los modelos mixtos', parte III. Chihuahua, México. Julio 31-agosto 11. III-24-III38. 1995.
56. Mendenhall W., D. D. Wackerly, and R. L. Scheaffer. *Estadística matemática con aplicaciones*. Segunda edición. Grupo editorial Interamericano. 447-45. 1994.
57. Schaeffer L. R. *Maximum likelihood estimation of variance components in dairy cattle research*. *J. Dairy Sci.* 59:2146-2151. 1976.

58. Graser H. U., S.P. Smith and Tier. *A derivative free approach for estimating variance components in animal models by restricted maximum likelihood*. J. Anim. Sci. 64: 1362-1370. 1987.
59. Jensen J., and I. L. Mao. *Transformation algorithms in analysis of single trait and of multitrait with equal design matrices and one random factor per trait*. J. Anim. Sci. 66:2570-2761. 1989.
60. Núñez D. R. y F. A. Rodríguez. *Modelos que se usan actualmente en evaluaciones genéticas nacionales*. En: Memorias del curso intensivo, 'Introducción a la metodología de los modelos mixtos', parte II. Chihuahua, México. Julio 31-agosto 11. II1-II12. 1995.
61. Sosa C. F. *Diagonalización simultanea de las ecuaciones de los modelos mixtos para estimación multivariada de componentes de varianza y covarianza por máxima verosimilitud restringida*. En: Reunión Nacional de Investigación Pecuaria, Guadalajara Jal. 72 (Resumen). 1993.
62. Schaeffer L. R. *Problems in estimating covariance matrices in multitrait animal model*. In 4th World Congress on Genetics Applied to Livestock Production, Edinburg.14:441-444. 1990.
63. Meyer K. *Estimating variances and covariances for multivariate animal models by restricted maximum likelihood*. Genet. Sel. Evol. 23:67-83. 1991.
64. Meyer K. *User notes, DFREML*. Version 3.0. 1997.
65. Searle S. R. *Linear models*. John Wiley & Son. New York, USA. 266-269. 1971.
66. Gerard C. F. *Análisis numérico*. Versión en español de Jaime Luis Valls Cabrera. Alfaomega. México. 79-141. 1997.
67. León S. T. *Linear algebra with applications*. Third edition. Macmillan publishing company. New York, USA. 329-412. 1990.
68. Nieves A. y F. C. Domínguez. *Métodos numéricos aplicados a la ingeniería*. Segunda impresión. CECSA. México. 125-293. 1997.
69. Misztal I., and D. Gianola. *Indirect solution of mixed model equations*. J. Dairy Sc. 70:716-723. 1987.

70. Van Vleck L. D., and D. J. Dwyer. *Comparison of iterative procedure for solving equation for sire evaluation*. J. Dairy Sci. 68:1006-1014. 1985.
71. Schaeffer L. R. Course notes. Comparison of models. Set 17, CGIL, Guelph, September 2003. Canada. 1-5. 2003.

III. Diferentes modelos de regresión y su ajuste en producción de leche por día¹

Apodaca-Sarabia C. A.²

Resumen

Con el objetivo de comparar el ajuste a la producción diaria de leche de las ecuaciones de Ali y Schaeffer (AS), Guo y Swalve (GS), Wilmink (WL) y Word (WO) se analizaron las producciones diarias, registradas mensualmente de 902 vacas Holstein con una a seis lactancias en un establo en la región árida de México. El análisis se realizó en dos etapas, la primera para obtener las producciones predichas para cada vaca-lactancia-ecuación y su ajuste medido con; i) la correlación entre la producción predicha y observada ($r_{\hat{A}A}$), y ii) el porcentaje de la suma de cuadrados total atribuible al error (PSB). En la segunda etapa, se comparó el ajuste de las ecuaciones usando el modelo mixto con las combinaciones año-época de parto (AEP) y número de lactancia-ecuación (NLE) como fijas y vaca (V) como aleatorio, y $r_{\hat{A}A}$ y PSB las variables respuesta. La comparación de los modelos fue dentro de lactancia mediante la combinación NLE. Los efectos de AEP y NLE fueron significativos ($P < 0.0001$) en $r_{\hat{A}A}$ y PSB. La varianza V fue el 43.61 y 39.43 % de la varianza total en $r_{\hat{y}y}$ y PSB, respectivamente. En todas las ecuaciones el ajuste puede considerarse adecuado, sin embargo, AS logró mejor ajuste en cada lactancia ($P < 0.0001$). En todas las ecuaciones el menor ajuste se observó en la primera lactancia, el mayor en entre la cuarta o quinta y descendió en la sexta. La importancia estadística del número de lactancia nos indica que debe ser considerado en los modelos del día de prueba.

Palabras clave. Leche, producción, curvas, modelos, ajuste.

¹ Tesis de doctor en ciencias veterinarias. Universidad Autónoma de México. Facultad de Medicina Veterinaria

² apodacacs@correo.chapingo.mx

III - Different regression models and its fit in daily milk production¹

Apodaca-Sarabia C. A.²

Abstract

With the objective of compare the fit to daily milk production of the regression equations of Ali y Schaeffer (AS), Guo y Swalve (GS), Wilmink (WL) and Wood (WO), the daily milk production, registered monthly of 902 cows with one to six lactation of a herd from the arid region of Mexico were analyzed. The analysis was in two parts, in the first the predicted daily milk production for each cow-lactation-equation and, its fit estimated with: i) the correlation between actual and predicted daily milk yield ($r_{\hat{A}A}$), and ii) the percentage of the residual sums square in relation to total sums squares (PSB) was calculated. In the second part, a mixed model was used with $r_{\hat{y}y}$ and PSB as the dependent variables and the independent variables were; year-season (AEP) and lactation-equation (NLE) as fixed effects, and cow (V) as random effect. The fit of equation was compared within lactation. The effects, AEP and NLE were significant ($P < 0.0001$) for $r_{\hat{y}y}$ and PSB. The variance of V was the 43.61 and 39.43 % of the total variance of $r_{\hat{y}y}$ and PSB, respectively. The fit was good in all the equation, however, AS fit better than others in all lactation. In all the equations the fit was ascending until the fourth or fifth lactation and decreased in the sixth. The statistical importance of lactation number is an indicator of they inclusion, in some way, in the test day model methodology.

Keywords. Lactation, curves, models, fit.

¹ Doctoral thesis in Veterinarian Sciences. Universidad Autónoma de México. Facultad de Veterinaria

² apodacacs@correo.chapingo.mx

3.1. Introducción

El interés en las ecuaciones de regresión sobre la curva de producción diaria de leche en una lactancia, tomó importante giro con los modelos del día de registro, en los que los objetivos son predecir los valores genéticos de producción diaria y por lactancia haciendo uso más eficiente de los registros de producción diaria^{1,2,3,4}. Son muy variadas las ecuaciones de regresión usadas en la curva de producción diaria de leche y se han extendido a todas las variables relacionadas con la producción de leche. Guo y Swalve⁵ muestran una relación de ellas, sin embargo el uso de polinomios ortogonales y ecuaciones de mayor complejidad son también comunes^{6,7,8}. El ajuste de los modelos depende de un gran número de variantes, y es posible que los modelos que presenten buen ajuste en la producción diaria, no lo produzcan en los modelos de regresión aleatoria donde se incluye los cambios genéticos y de ambiente permanente de producción diaria^{9,10,11}.

La correlación entre observados y predichos, y el porcentaje que del total representa la suma de cuadrados del error son dos criterios de ajuste del modelo. Estas medidas de ajuste, enfatizan la similitud entre la producción observada y predicha durante la lactancia y no únicamente con producción de leche por lactancia¹. Así, el uso de ecuaciones en los modelos de regresión aleatoria puede permitir con facilidad estimar los valores genéticos de partes de la lactancia y de persistencia. Esta última con gran cantidad de definiciones y relaciones con el comportamiento reproductivo y salud de los animales^{12,13}. El objetivo del trabajo fue comparar los ajustes de ecuaciones de regresión y la función gama en la producción diaria de leche de una lactancia, ello con fines de comprobar su factibilidad de uso en modelos de funciones de covarianza para estimar los componentes de varianza de producción de leche por día.

3.2. Materiales y métodos

La información usada incluyó producción diaria de leche registrada cada 30 días de 902 vacas Holstein de un establo en Tlahualilo Durango, localizado a 25°54'07" de latitud norte y 103°35'09" de longitud oeste y una altitud media de 1137 msnm. El clima de la región se caracteriza por tener, 21.1°C de promedio anual de temperatura, y 239 mm de precipitación. Siendo un clima seco desértico, que corresponde a la fórmula climática BWhw'(e) según García¹⁴.

Los registros se realizaron de 1973 a 1993 e incluyeron; identificación de la vaca y de sus padres, fecha de nacimiento de la vaca, fecha y número de parto (1 a 12), y fecha de registro y producción diaria de leche (kg) con frecuencia aproximadamente mensual. La edición consistió en eliminar las lactancias con menos de 7 registros de producción y los registros mensuales posteriores al décimo mes de lactancia. En total se analizaron 2132 registros mensuales de producción diaria de leche de 902 vacas.

El análisis estadístico se realizó en dos etapas, en la primera se estimaron las ecuaciones de regresión para cada vaca y lactancia, y en la segunda se probó el ajuste de cada una de las ecuaciones y lactancias.

Ecuaciones de regresión. Se compararon cuatro ecuaciones de regresión que describen la producción diaria de leche en una lactancia. Así, para la producción por día el día de t de la lactancia (y_t), se incluyeron las ecuaciones lineales:

Ali y Schaeffer¹ (AS):

$$y_t = \beta_0 + \beta_1(t/305) + \beta_2(t/305)^2 + \beta_3 \ln(305/t) + \beta_4 [\ln(305/t)]^2 + e_t$$

Guo y Swalve⁵ (GS):

$$y_t = \beta_0 + \beta_1 t^{1/2} + \beta_2 \ln(t) + e_t$$

Y Wilmink¹⁵ (WL):

$$y_t = \beta_0 + \beta_1 t + \beta_2 e^{-0.05t} + e_t,$$

La última ecuación fue la función gama propuesta por Wood¹⁶ (WO) con la forma $y_t = at^b e^{-ct}$, ecuación que para su análisis por métodos lineales se realizó la transformación logaritmo natural, resultando, $\ln(y_t) = \ln(a) + b \ln(t) - ct + e_t$, que se puede escribir:

$$\ln(y_t) = \beta_0 + \beta_1 \ln(t) - \beta_2 t + e_t$$

En cada uno de los cuatro modelos: los β_i ($i = 0, 1, 3, 4$) son los coeficientes de regresión y e_t el error. Se relaciona, en AS, a β_3 y β_4 con el aumento inicial, y a β_1 y β_2 con la tasa de disminución posterior al pico de producción. Mientras en WO, b representa el aumento hasta la producción máxima, c la pendiente de decremento¹.

En términos matriciales las anteriores ecuaciones de regresión son: $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, donde \mathbf{y} es el vector de observaciones, producción diaria de leche o su logaritmo natural, \mathbf{e} el vector de residuales, \mathbf{X} la matriz de covariables y \mathbf{b} el vector de parámetros desconocidos que deben ser estimados. Los elementos de los renglones de \mathbf{X} y del vector \mathbf{b} , varían para cada ecuación y día de lactancia así, para AS son $[1 \ t/305 \ (t/305)^2 \ \ln(305/t) \ (\ln(305/t))^2]$, y $[b_0 \ b_1 \ b_2 \ b_3 \ b_4]'$ respectivamente. La solución, se obtuvo mediante mínimos cuadrados ya que Ali y Schaeffer¹ no encontraron diferencia significativa entre la solución con mínimos cuadrados generalizados y mínimos cuadrados, así los coeficientes de regresión estimados fueron: $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{y}$.

Para cada una de las cuatro ecuaciones de regresión se calculó la producción predicha para cada vaca y lactancia en la base de datos. Con esta información se calculó la correlación entre los observados y predichos ($r_{\hat{y}y}$) y el porcentaje de la suma de cuadrados del error con respecto a la suma de cuadrados total (PSB). Tomando los vectores de observaciones y predichos, \mathbf{y} e $\hat{\mathbf{y}}$ respectivamente, PSB es:

$$\text{PSB} = \frac{(\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})}{\mathbf{y}'\mathbf{y}} 100$$

Comparación de modelos. La segunda fase del análisis consistió en la comparación del ajuste logrado con cada una de las ecuaciones de regresión. En esta fase se generó, de la información original, una base de datos con; identificación de la vaca, año y estación de parto, número de lactancia, $r_{\hat{y}y}$ y PSB. Con fines de optimizar el uso de la información disponible y en función del reducido número de observaciones por año de parto entre 1973 y 1981, el periodo se agrupó en 1981, de igual forma, el número de observaciones de las lactancias 6 a 12 fue reducido, por que se agruparon en la lactancia 6. Estas agrupaciones implican suponer los efecto de año de parto entre 1973 y 1981 no son diferentes y que la curva de lactancia es igual para las lactancia 6 a 12.

Para comparar las ecuaciones de regresión, se generaron las combinaciones año-estación de parto (AEP) y número de lactancia-ecuación (NLE) de regresión. Lo anterior con fines de considerar los posibles cambios en el tiempo y las diferencias en ajuste de las ecuaciones atribuibles a diferentes curvas de lactancia en cada número de lactancia.

El modelo estadístico utilizado fue un modelo mixto con variables dependientes, $r_{\hat{y}y}$ y PSB, y independientes, mientras vaca (V), AEP y NLE como variables independientes. El efecto de

vaca fue considerado aleatorio, y fue incluido con fines de considerar la variabilidad de las vacas en la comparación de las ecuaciones de regresión. El modelo mixto fue:

$$y_{ijkl} = \mu + V_i + AEP_j + NLE_k + e_{ijkl}$$

Donde; y_{ijkl} fue $r_{\hat{y}y}$ o PSB en la k-ésima combinación número de lactancia-ecuación de regresión y j-ésimo combinación de año-época de parto de la i-ésima vaca, μ la media de la población, V_i el efecto de la i-ésima vaca ($i=1, 2, \dots, 902$) con media cero y varianza σ_v^2 , AEP efecto de la j-ésima combinación año-estación de parto ($j=811, 812, \dots, 931$), NLE_k efecto de la k-ésima combinación entre número de lactancia y ecuación ($k=11, 12, \dots, 64$), y e_{ijkl} el error experimental ($0, \sigma_e^2$). La comparación fue hecha para las ecuaciones de regresión en un mismo número de lactancia con las medias de cuadrados mínimos.

3.3. Resultados y discusión

El coeficiente de determinación de las ecuaciones de regresión por vaca y lactancia fluctuó entre 0.50 y 0.90, dispersión atribuible a efectos del día de registro no considerados en el modelo de regresión. Las curvas de lactancia para cada número de lactancia y ecuación de regresión se muestran en la figura 3.1. Las ecuaciones WL y WO produjeron la misma curva de lactancia, así en la figura sólo se presenta una de ellas y en el resto del capítulo se hará referencia también sólo una de ellas. Las producciones predichas siempre fueron menores con WO.

Tanto para $r_{\hat{y}y}$ como para PSB los efectos, AEP y NLE fueron altamente significativos ($P<.0001$), así, los cambios en producción diaria de leche asociados al AEP y al número de lactancia producen cambios de interés en el ajuste de las ecuaciones de regresión. La correlación entre la producción por día observa y predicha fue de 0.7122 a 0.9303 (Cuadro 3.1), mientras el PSB fluctuó entre 0.7822 y 1.9849 % (Cuadro 3.2). El mejor ajuste en todas las lactancias fue con AS, e igual que en el resto de las ecuaciones, aumento con el número de lactancia. Las estimaciones de varianza de vaca fueron de 0.01062 y 0.9222, mientras la varianza del error fue de 0.01373 y 1.4166 para $r_{\hat{y}y}$ y PSB respectivamente. Así, de la varianza total en $r_{\hat{y}y}$ y PSB, las diferencias entre las vacas causan el 43.61 % en $r_{\hat{y}y}$ y el 39.43 % en PSB.

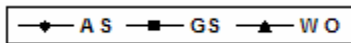
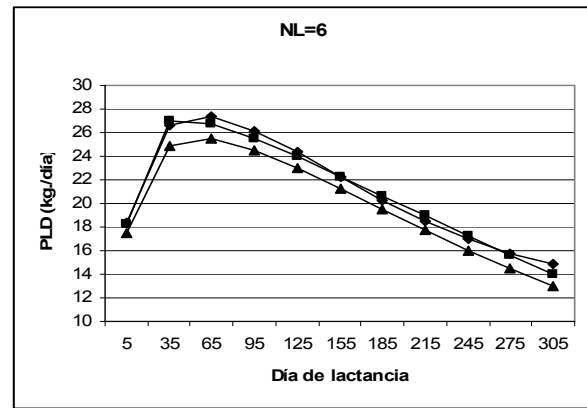
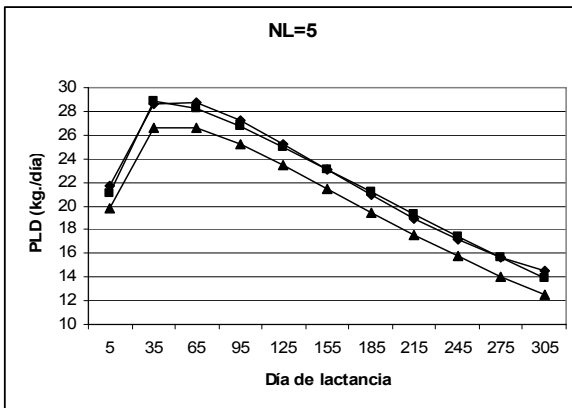
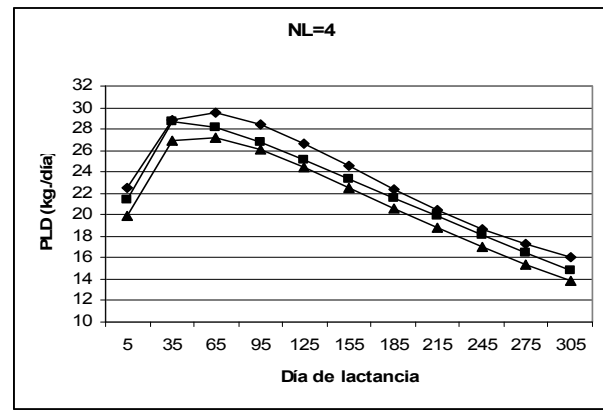
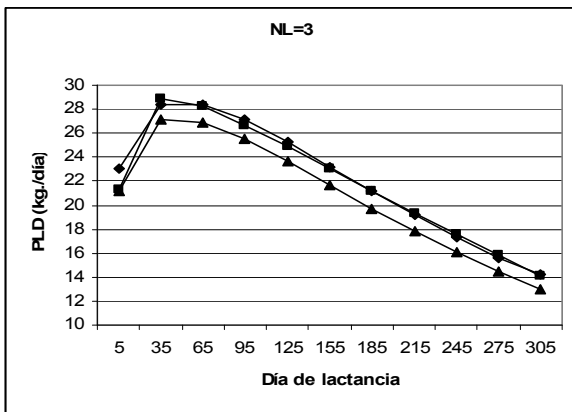
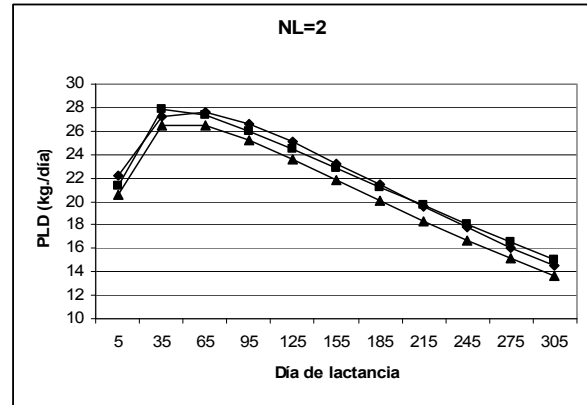
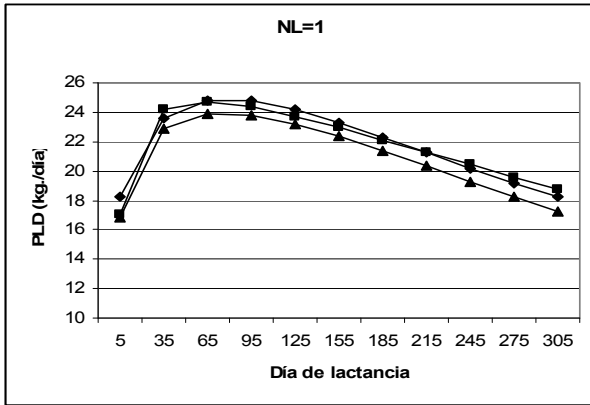


Figura 3.1. Curvas de lactancia por ecuación (Ali y Schaeffer AS; Guo y Swalve GS; Wood WO) y número de lactancia (NL).

Cuadro 3.1. Medias de mínimos cuadrados y su error estándar de correlación entre observados y predichos según la ecuación de regresión y el número de lactancia (NL).

NL	Ecuación de regresión			
	Ali y Schaeffer	Guo y Swalve	Wilmink	Wood
1	0.8492±.006778 ^a	0.7087±.006778 ^b	0.7054±.006778 ^b	0.7054±.006778 ^b
2	0.91833±.006618 ^a	0.8352±.006618 ^b	0.8318±.00618 ^b	0.8318±.006618 ^b
3	0.9192±.007505 ^a	0.8387±.007505 ^b	0.8376±.007505 ^b	0.8376±.007505 ^b
4	0.9266±.008197 ^a	0.8441±.008197 ^b	0.8438±.008197 ^b	0.8438±.008197 ^b
5	0.9269±.009901 ^a	0.8524±.009190 ^b	0.8496±.009190 ^b	0.8496±.009190 ^b
6	0.9074±.008142 ^a	0.8161±.008142 ^b	0.8164±.008142 ^b	0.8164±.008142 ^b

Diferente literal en el mismo renglón indica diferencia estadística (P<0.01)

Cuadro 3.2. Medias de mínimos cuadrados de porcentaje de la suma de cuadrados del error en relación a la suma de cuadrados total.

NL	Ecuación de regresión			
	Ali y Schaeffer	Guo y Swalve	Wilmink	Wood
1	0.8021±.0672 ^a	1.4141±.0672 ^b	1.4225±.0672 ^b	1.4225±.0672 ^b
2	0.9100±.0656 ^a	1.6636±.0656 ^b	1.6892±.0656 ^b	1.6892±.0656 ^b
3	0.9495±.0750 ^a	1.6589±.0750 ^b	1.6681±.0750 ^b	1.6681±.0750 ^b
4	0.8151±.0819 ^a	1.6648±.0819 ^b	1.6610±.0819 ^b	1.6610±.0892 ^b
5	1.0076±.0920 ^a	1.8616±.0920 ^b	1.8830±.0920 ^b	1.8830±.0920 ^b
6	1.0701±.0805 ^a	2.0712 ±.0805 ^b	2.0491±.0805 ^b	2.0491±.0805 ^b

Diferente literal en el mismo renglón indica diferencia estadística (P<0.01)

En Canadá también las ecuaciones de regresión también mostraron mejor ajuste que la función gama, trabajo en el que además, la estimación por mínimos cuadrados generalizados no produjo mejoras importantes en el ajuste de las ecuaciones de regresión¹. Sin embargo en ganado productor de carne se logró mejor ajuste cuando se realizó la estimación por métodos no lineales¹⁷. No obstante, si se buscan ecuaciones de regresión de posible uso en el modelo del día de prueba, por el volumen de información requerida, es posible que usar métodos no lineales no sea aún muy recomendable.

Usando diferentes modelos de regresión aleatoria y suposiciones sobre la estructura de covarianza de los efectos aleatorios, el mejor ajuste correspondió a las ecuaciones con mayor número de parámetros, observándose mayor ajuste cuando se usó diferente ecuación de regresión para los valores genéticos y los de ambiente permanente¹⁸. Sin embargo, aún cuando no existe consenso sobre el modelo de regresión a usar, en términos generales, los polinomios de Legendre de alto orden, cuarto o quinto, muestran mejor ajuste que las ecuaciones denominadas de regresión, la desventaja es el aumento considerable del número de parámetros en los modelos del día de prueba¹⁰. Los mismos autores observaron menor ajuste en la ecuación de GS al mostrar mayor error al inicio de la lactancia. Los resultados arriba mencionados^{10,18} se obtuvieron con diferentes modelos del día de prueba, es decir, modelos con regresiones fijas y aleatorias más el efecto fijo hato-día de registro, que define el grupo de comparación, ello es la causa principal de las diferencias con este estudio, sin embargo la aportación de los resultados de este artículo son con fines de determinar algunas ecuaciones de regresión para la regresión fija de los modelos del día de prueba.

El número de lactancia es generalmente encontrado importante en la literatura¹⁹, las diferencias no sólo son de importancia en el nivel y tendencia de producción, sino también en la varianza genética²⁰. El interés de la curva de lactancia ha cambiado en los años, inicialmente fue predecir la producción por lactancia con fines de selección buscando reducir el número de registros durante la lactancia. Sin embargo en los modelos del día de prueba, se considera que los genes responsables de la producción diaria de leche, grasa y proteína son diferentes en cada número de lactancia, así es de importante determinar el ajuste de las ecuaciones de regresión a las diferentes lactancias. Así si el modelo del día de prueba considera cada lactancia como variable correlacionada, se debe estar seguro que la ecuación de regresión en cada lactancia es la adecuada, los resultados mostraron siempre mejor ajuste con AS.

Una ventaja adicional de los modelos del día de prueba, es la facilidad relativa de estimar los valores genéticos de producción a pico de lactancia, persistencia y de diferentes periodos de la lactancia. Cuando se tiene la función gama (WO) el calculo de la producción máxima es sólo una derivada cuyo resultado además esta disponible en la literatura^{16,19}, no obstante con la ecuación de AS, debido a que es lineal en los parámetros pero no en las covariables así, obtener el máximo requiere de métodos numéricos, haciendo el procedimiento más demandante de tiempo y memoria de computo. La importancia de la producción máxima y persistencia es ampliamente tratada en la literatura. Algunos puntos de vista son las correlaciones genéticas entre estas y producción por lactancia. Las estimaciones de las correlaciones genéticas de producción por lactancia son favorables con producción máxima, pero no con persistencia¹, lo que puede ser atribuido a la forma de medir la persistencia^{1,21}. Grossman *et al.*²² proponen un modelo con mejor definición de persistencia pero de mayor complicación y Dekkers *et al.*²³ indican su posible importancia económica.

3.4. Conclusiones

Aun cuando se encontró mejor ajuste del modelo de Ali y Schaeffer, el ajuste obtenido por las ecuaciones de Guo y Swlave, Wilmink y la función gama (Wood), indican factibilidad de uso en modelos de regresión aleatoria, sobre todo por la reducción de dos coeficientes de regresión en comparación con la primera.

El ajuste obtenido es únicamente en producción diaria de leche, es trabajo no tiene posibilidad de mostrar si el ajuste es el mismo para los valores genéticos de producción diaria de leche en la lactancia, ello debe ser probado con modelos del día de prueba.

Las diferencias en número de lactancia son de interés, como ya fue demostrado, la posibilidad de incluir en el modelo el número de lactancia como efecto fijo, ya sea sólo o anidado con las covariables para estimar las diferentes curvas puede ser de interés confines de evitar el uso del modelo multivariado en regresión aleatoria.

3.5. Literatura citada

1. Ali T. E. and L. R. Schaeffer. *Accounting for among test day yields in dairy cows*. Can. J. Anim. Sci. 67:637-644. 1987.

2. Patck E. and L. R. Schaeffer. Use of test day yields for genetic evaluation of dairy sires and cows. *Livest. Prod. Sci.* 34:23-34. 1993.
3. Meyer K. and W. G. Hill. *Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood.* *Livestock Production Science.* 47:185-200. 1997.
4. Kettunen A., E A. Mäntysaari, and J. Pösö. *Estimation of genetic parameters for daily milk yield of primiparous Ayrshire cows by random regression test-day models.* *Lives. Prod. Sci.* 66:251-261. 2000.
5. Guo, Z., and H. H. Swalve. 1995. Modeling of the lactation curves as sub-model in the evaluation of test day records. *INTERBULL*, Prage, sep. 7-8.
6. Kirkpatrick M., D. Lofsvold and M. Bulmer. *Analysis of the inheritance, selection and evolution of growth trajectories.* *Genetics.* 124:979-993. 1990.
7. Brotherstone S, I. M. S. White and K. Meyer. *Genetic modeling of dairy milk yield using orthogonal polynomials and parametric curves.* *Anim. Science.* 70:407-415. 2000.
8. Jensen, J. *Genetic evaluation of dairy cattle using test-day models.* *J. dairy Sci.* 84:2803-2812. 2001.
9. Olori V. E., W. G. Hill, B. J. McGuirk, and S. Brotherstone. *Estimating variance components for test day records by restricted maximum likelihood with random regression animal model.* *Livest. Prod. Sci.* 61:53-63. 1999.
10. López-Romero y M. J. Carabaño. *Evaluación de modelos aleatorios para el análisis de datos de producción de leche a lo largo de la lactación.* <http://www.etsia.upv.es/Acteón/docs/PAPER-new.pdf> consultado 21/1/2002.
11. Stram, D. O. and J. W. L. Lee. *Variance component testing in the longitudinal mixed model.* *Biometrics.* 50:1171-1177. 1994.
12. Sölkner J. and W. Fuchs. *A comparison of different measures of persistency with special respect to variation of test-day milk yields.* *Livestock Production Science.* 16:305-319. 1987.

13. Dekkers J. C. M., J. H. Ten Hag, and A. Weersink. *Economic aspect of persistency of lactation in dairy cattle*. Livestock Production Science. 53:237-252. 1998.
14. García de M. E. *Modificaciones al Sistema de Clasificación Climática de Köppen*. Impreso en México. 1981.
15. Wilmink, J. B. M. Adjusted of test-day milk, fat and protein yields for age, season and stage of lactation. Livest. Prod. Sci. 16:335-348. 1987.
16. Wood, P. D. P. *Algebraic model of the lactation curves in cattle*. Nature, 216: 164 -165. 1967.
17. Ramírez, V. R. *Comparación de ecuaciones y procedimientos para estimación de curvas de lactancia en Angus y Suizo Pardo y sus cruzas*. Tesis de maestría en ciencias en Producción animal, Universidad Autónoma Chapingo, Chapingo México. 105 pp. 1994.
18. Jeffrezeic, F., I. M. S. White, R. Thomson, and P. M. Visscher. *Contrasting models for lactation curve analysis*. J. Dairy Sci. 85:968-975. 2002.
19. Gutierrez, H. S. *Efectos ambientales sobre los coeficientes de la ecuación de Wood y los componentes de la curva de producción de leche en vacas Holstein*. Tesis profesional, Ingeniero Agrónomo Especialista en Zootecnia. Universidad Autónoma Chapingo, Chapingo México. 50 pp. 1994.
20. De Ross, A. P. W., A. G. F. Harbers, and J. de Jon. *Random herd curves in a test-day model for milk, fat and protein production of dairy cattle in Netherlands*. J. Dairy Sci. 87:2693-2701. 2004.
21. Ferris, A. T. *Selecting for lactation curve sharp and milk yield in dairy cattle*. A dissertation submitted to Michigan State University. Doctor of Philosophy. Department of Dairy Science. Michigan State University. 1981.
22. Grossman, M., S. M. Hartz, and W. J. Koops. *Persistency of yields lactation: a novel approach*. J. Dairy Sci. 82:2192-2197. 1999.
23. Dekkers J. C. M., J. H. Ten Hag, and A. Weersink. *Economic aspect of persistency of lactation in dairy cattle*. Livest. Prod. Sci. 53:237-252. 1998.

IV. Estimación de (co)varianzas genéticas y de ambiente permanente con diferentes modelos de funciones de covarianza¹

Apodaca-Sarabia C. A.²

Resumen

Para comparar 24 modelos de funciones de covarianza (COF) en la descripción genética de la producción diaria de leche, se analizaron 18421 producciones diarias de leche de vacas con 1 a 7 lactancias en el periodo de 1973 a 1993 de un establo del norte de México. Los 24 COF se formaron por 3 efectos fijos y 8 ecuaciones de regresión, los efectos fijos fueron: a) año de parto (AP), estación de parto (EP) y número de lactancia (NL), b) la combinación año-estación-numero de lactancia (AENL), y c) día de prueba (DR). Las ecuaciones fueron; Ali y Schaeffer (ALS), Guo y Swalve (GS), Wilmink (WL), Wood (WO) y los polinomios ortogonales de Legendre de segundo a quinto orden (PL_i , $i = 2, 3, 4, \text{ y } 5$), la regresión fija y las aleatoria fue la misma. En los modelos PL4, PL5, y WO no se logró convergencia, y en WL las (co)varianzas genéticas estimadas fueron mayores a las observadas. El mejor ajuste fue en DR-AS y DR-GS. La heredabilidad estimada, se incrementó desde el inicio de la lactancia (0.03 a 0.27), la repetibilidad (0.285 a 0.52) estimada aumentó desde el inicio de la lactancia hasta el segundo o quinto mes y disminuyó hasta el final de la lactancia. La varianza genética estimada aumento desde el inicio hasta el final de la lactancia (1.51 a 13.88 kg²) y fueron menores en los modelos con DR. Las correlaciones genéticas fueron altas, de 1 al final de la lactancia, y nunca negativas.

Palabras clave: Leche, producción, día de prueba, funciones de covarianza, parámetros genéticos.

¹ Tesis de doctor en ciencias veterinarias. Universidad Autónoma de México. Facultad de Medicina Veterinaria

² apodacacs@correo.chapingo.mx

IV. Covariance function in the estimation of genetic and environmental permanent (co)variances of milk production¹

Apodaca-Sarabia C. A.²

Abstract

With the aim to compare 24 covariance function models (COF) for describe genetically the dairy milk production, 18421 daily milk production records from cows with 1 to 7 lactations between 1973 and 1993 of a herd in the arid region of Mexico was analyzed. The COF was formed by 3 fix effects and 8 regression equations, the fix effects were: a) year (AP), season (EP), and lactation (NL), b) year-season-lactation (AENL), and c) test day (DR). The regression equation were: Ali and Schaeffer (AS), Guo y Swalve (GS), Wilmink (WL), Wood (WO), and Orthogonal polynomials of Legendre from 2 to 5 order (PLi: i= 2, 3, 4 and 5). The models with any of the three fixed effects and PL4, PL5, or WO did not converge. The models with any of the three fixed effects and WL, the genetic variances estimated were greater than phenotypic variances. The best fit were in the models DR-AS and DR-GS. The estimation of heritability for PLD was ascending from the beginning of lactation (0.03 to 0.27). The estimation of repeatability for PLD (0.285 to 0.52), increasing until the second or fifth, and decreasing the end of lactation. The estimation of genetic variance (1.51 to 13.88 kg²) was ascending from first to the tenth month of lactation. The estimated of genetic variance, with models including DR was the lowest. The genetic correlations were high and never negative.

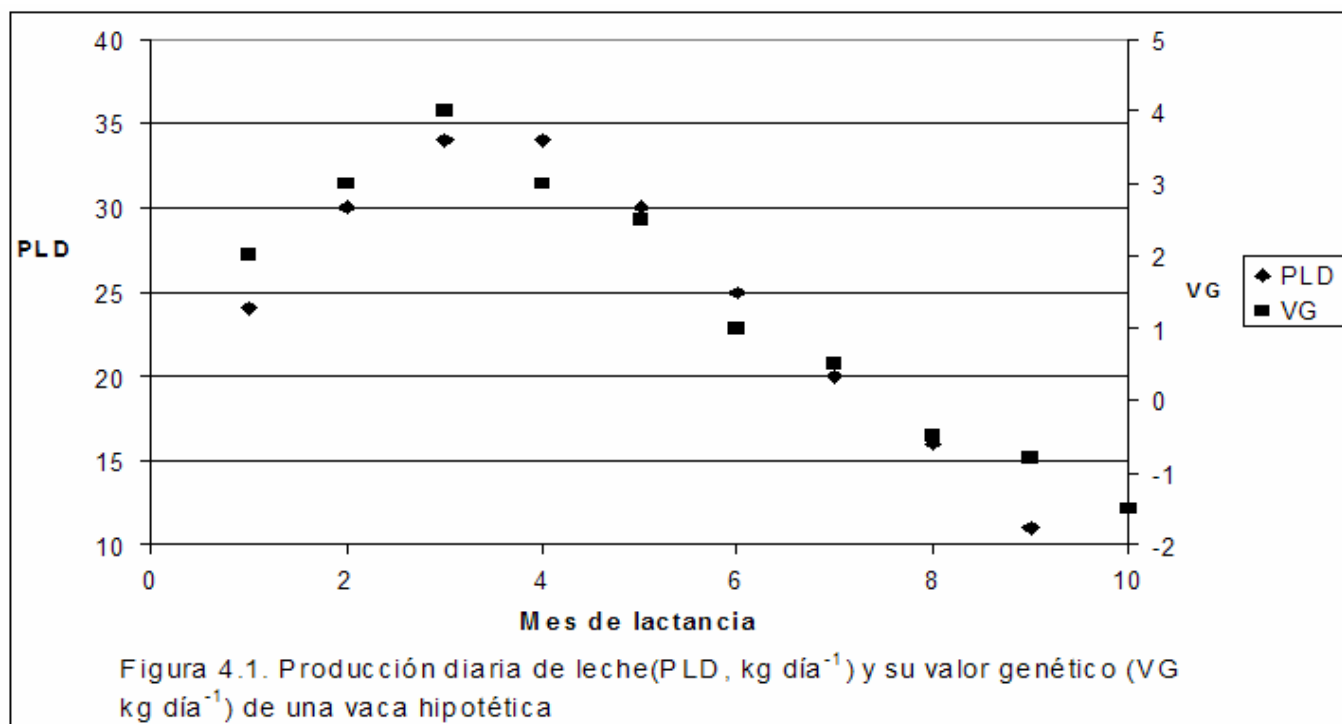
Keywords. Test-day, covariance function, milk, yields, genetics parameters.

¹ Doctoral thesis in Veterinarian Sciences. Universidad Autónoma de México. Facultad de Veterinaria

² apodacacs@correo.chapingo.mx

4.1. Introducción

Se ha demostrado que los genes responsables de la producción diaria de leche no son los mismos durante la lactancia¹, así, el valor genético, el efecto ambiental permanente y la producción diaria de leche, cambia durante la lactancia (Figura 4.1). Esta relación entre producción diaria, el valor genético y el efecto ambiental permanente de la producción con el día de lactancia es aprovechada por los modelos de regresión aleatoria y funciones de covarianza, modelos denominados como del día de prueba (TDM) y sus ventajas sobre los modelos en los que se analiza producción por lactancia, son: a) descripción genética continua de la producción diaria durante la lactancia², b) mayor exactitud en la predicción de la respuesta a la selección² y del valor genético³, c) uso más eficiente de la información^{2,4}, d) inclusión, en el modelo mixto, de los efectos ambientales sobre la producción el día de su registro², y e) eliminación de la necesidad de proyección de lactancias incompletas³. Sin embargo, en relación al modelo con producción de leche por lactancia, se incrementa el número de ecuaciones a solucionar en la predicción del valor genético de los animales^{3,5,6}.



Es grande el número de ecuaciones de regresión propuesto para producción diaria de leche, ecuaciones que también son factibles de usar en la parte aleatoria, valor genético y efecto de ambiente permanente, de los modelos de regresión aleatoria y funciones de covarianza. En el

cuadro 4.1 se muestra una relación de ecuaciones de regresión utilizadas en la predicción diaria de leche de una lactancia. El ajuste del modelo depende del efecto fijo usado para definir el grupo contemporáneo y de la ecuación de regresión usada en la parte fija y aleatoria^{3,7,8}. Las diferencias en ajuste se reflejan en las estimaciones de los componentes de varianza y en las predicciones de los valores genéticos y de los efectos ambientales permanentes^{9,10,11,12}. El objetivo del presente estudio fue comparar diferentes modelos de funciones de covarianza y estimar las (co)varianzas genética y de efecto permanente de producción diaria de leche durante la lactancia.

Cuadro 4.1. Ecuaciones de uso en modelos de regresión aleatoria para producción diaria de leche en una lactancia.

Ecuación	Autor
a) $\ln(y_t^*) = b_0 + b_1 t$	Jenkins y Ferrel (1984)
b) $y_t^{-1} = b_0 + b_1 t^{-1} + b_2 t$	Nelder (1966)
c) $y_t = b_0 t^{b_1} e^{-b_2 t}$	Wood(1967)
d) $y_t = b_0 + b_1 t + b_2 e^{-0.05t}$	Wilmink (1987)
e) $y_t = b_0 + b_1 \sqrt{t} + b_2 \ln t$	Guo y Swalve (1995)
f) $\ln(y_t) = b_0 + b_1 \ln t + b_2 t + b_3 \sqrt{t}$	Badner y Anderson (1985)
g) $y_t^{-1} = b_0 + b_1 t^{-1} + b_2 t + b_3 t^2 + b_4 t^3$	Nelder (1966)
h) $y_t = b_0 + b_1 t^2 + b_2 t^3 + b_3 t^4$	
i) $y_t = b_0 + b_1 t^2 + b_2 t^3 + b_3 \ln t$	Guo y Swalve (1995)
j) $y_t = b_0 + b_1 \sqrt{t} + b_2 \ln t + b_3 t^4$	
k) $y_t = b_0 + b_1 (t/305) + b_2 (t/305)^2 + b_4 \ln(305/t) + b_4 \ln^2(305/t)$	Ali y Schaeffer (1987)
l) $y_t = b_0 + b_1 t + b_2 \sin(0.01)t^2 + b_3 \sin(0.01)t^3 + b_4 e^{-0.055t}$	Guo y Swalve (1995)

* producción de leche el día t de la lactancia.

4.2. Materiales y métodos

El análisis consistió en evaluar diferentes modelos de funciones de covarianza para estimar las (co)varianzas genéticas y ambientales permanentes de producción diaria de leche. Los modelos difirieron tanto en el efecto efectos fijos como en la regresión fija y aleatoria usada en la predicción del valor genético y el ambiental permanente.

Regresiones fijas y aleatorias. Para describir el efecto promedio del día de lactancia sobre la producción diaria de leche, las desviaciones causadas por el valor genético de cada animal y las desviaciones atribuibles al los efectos ambientales permanentes, se consideraron ocho ecuaciones de regresión, estas fueron las propuestas por: Ali y Schaeffer (AL), Guo y Swalve (GL), Wilmlink (WL), Wood (WO) y polinomios ortogonales de Legendre de segundo a quinto orden (PL_i, i= 2, 3, 4, 5). Las ecuaciones de regresión tienen mínimo 3 y máximo 6 coeficientes de regresión.

La función gamma incompleta propuesta por WO¹³ con la forma; $y_t = at^b e^{-ct}$, no es lineal, así, para estimar los coeficientes de regresión con el procedimiento lineal es necesario obtener su logaritmo natural, resultando: $\ln(y_t) = \ln(a) + b \ln(t) - ct + e_t$. Otros dos modelos con tres coeficientes de regresión, fueron los propuestos por WL¹⁴, con ecuación $y_t = \beta_0 + \beta_1 t + \beta_2 e^{-0.05t} + e_t$, y GS¹⁵, definida por; $y_t = \beta_0 + \beta_1 t^{1/2} + \beta_2 \ln(t) + e_t$. La ventaja de los dos últimos sobre la ecuación de WO, es que son ecuaciones lineales y no requieren transformación para ser analizados on métodos lineales. El modelo de regresión con 5 coeficientes de AS¹, representa la producción por día el día t de la lactancia con, $y_t = \beta_0 + \beta_1(t/305) + \beta_2(t/305)^2 + \beta_3 \ln(305/t) + \beta_4 [\ln(305/t)]^2 + e_t$. En las ecuaciones anteriores, y_t es la producción por día el día t de la lactancia, a, b, c, y β_i son los coeficientes de regresión y e_t es el residual.

Para los polinomios ortogonales de Legendre (PL) se requiere transformar el día de lactancia (t) a la variable x, definida en el intervalo [-1, 1] y calculada con $x_i = [2(t_i - t_{\text{mínimo}})/(t_{\text{máximo}} - t_{\text{mínimo}})] - 1$. El polinomio de Legendre (PL) es:

$$\phi_j(X) = \frac{1}{2^j} \sqrt{\frac{2j+1}{2}} \sum_{m=0}^{j/2} (-1)^m \binom{j}{m} \binom{2j-2m}{j} x^{j-2m}$$

Con fines de simplificar su cálculo, se define:

$$P_0(x)=1, P_1(x)=x, P_{n-1} = \frac{1}{n+1} ((2n+1)XP_n(x)-nP_{n-1}(x)), \text{ y } \phi_n(x) = \sqrt{\frac{2n+1}{2}} P_n(x)$$

El PL escrito matricialmente es: $\Phi = \mathbf{M}\Lambda$. Donde los elementos de \mathbf{M} son $m_{ij} = t_i^{*(j-1)}$ y Λ es la matriz de los coeficientes obtenidos en esquema iterativo del PL antes mencionado. Así, considerando la producción diaria de leche en los días; 15, 75, 135 y 195. Los valores de x son: -1, -1/3, 1/3, y 1 respectivamente. Y los coeficientes de la función:

$$\Phi = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & -1/3 & 1/9 & -1/27 \\ 1 & 1/3 & 1/9 & 1/27 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.7071 & 0 & 0 & 0 \\ 0 & 1.224745 & 0 & 0 \\ -0.7906 & 0 & 2.371708 & 0 \\ 0 & -2.806243 & 0 & 4.677072 \end{bmatrix} = \begin{bmatrix} 0.7071 & -1.2247 & 0.5811 & -1.8708 \\ 0.7071 & -0.4082 & -0.6382 & 0.7622 \\ 0.7071 & 0.4082 & -0.6382 & -0.7622 \\ 0.7071 & 1.2247 & 0.5811 & 1.8708 \end{bmatrix}$$

Dado que el PL de grado máximo, para producción diaria de leche, es muy grande, se probó hasta el polinomio de quinto grado, iniciando con el polinomio más reducido, segundo grado, e incrementado un grado a la vez hasta el de quinto orden (PL_i, i=2, 3, 4, y 5). En caso de no obtener solución en un grado menor al quinto los restantes no se probaron.

Los modelos de regresión, antes mencionados, en términos matriciales son: $\mathbf{z} = \mathbf{X}\beta + \mathbf{e}$, donde: \mathbf{z} es el vector de observaciones, producciones diarias de leche durante la lactancia o su logaritmo natural según el modelo, \mathbf{X} es la matriz de covariables, β el vector de coeficientes a estimar, y \mathbf{e} el vector de residuales. Los elementos de β y \mathbf{X} varían para cada ecuación de regresión, el cuadro 4.2 muestra los elementos de β y la forma general de los renglones de la matriz de diseño (\mathbf{X}). En cuadro 4.2 para los polinomios se muestra la función, sin embargo, en el polinomio cúbico (PL3) β es $[\beta_0 \beta_1 \beta_2 \beta_3]'$, y renglón de \mathbf{X} para el día 15 es: [0.7071 -1.2247 0.5811 -1.8708]. Esto son los elementos de las matrices de diseño en la parte; a) fija para cuantificar el efecto promedio del día de lactancia en la producción diaria de leche y b) aleatoria para predecir los valores genéticos y de ambiente permanente de todos los animales y animales con registro, respectivamente.

Efectos fijos. Los efectos fijos incluidos en los modelos corresponden a los efectos que en la literatura se han considerado de mayor importancia en la producción de leche, tales como año de parto (AP), estación de parto (EP), número de lactancia (NL), y día de prueba (DR). El AP, EP y NP se incluyeron como tres efectos fijos (AEPNL) y como la combinación de los tres

(AENL). La combinación AENL y el DR hacen, con facilidad, el grupo de contemporáneos, la inclusión de los tres efectos por separado reduce el número de clases en los efectos fijos.

Debido a que la variable de respuesta es producción diaria de leche durante la lactancia, es necesaria la inclusión de la regresión fija, con fines de estimar la curva promedio de la información disponible, en ella se uso la misma que para los efectos aleatorios.

Cuadro 4.2. Elementos del vector de parámetros desconocidos (β) y de los renglones de la matriz de covariables (X) según la ecuación de regresión.

Ecuación	β	Renglón de X , Z , o W
Wood (1967)	$[\ln(a) \ b \ c]'$	$[1 \ \ln(t) \ -t]$
Ali y Schaeffer (1987)	$[\beta_0 \ \beta_1 \ \beta_2 \ \beta_3 \ \beta_4]'$	$[1 \ (t/305) \ (t/305)^2 \ \ln(305/t) \ [\ln(305/t)]^2]$
Wilmink (1987)	$[\beta_0 \ \beta_1 \ \beta_2]'$	$[1 \ t \ e^{-0.05t}]$
Guo y Swalve (1997)	$[\beta_0 \ \beta_1 \ \beta_2]'$	$[1 \ t^{1/2} \ \ln(t)]$
Polinomio de Legendre	Según el polinomio	$\phi_n(x)^a$

^a función de la cual se obtendrá el valor del coeficiente en las covariables en el modelo.

Base de datos. Para probar los modelos en la estimación de las covarianzas genéticas y de ambiente permanente de producción diaria de leche, se uso la siguiente base de datos. La información consistió de 28852 registros mensuales de producción diaria de leche de un establo localizado en Tlahualilo Durango, 25°54'07" de latitud norte y 103°35'09" de longitud oeste, con 1137 msnm, 21.1°C, y 239 mm de altitud y promedio anual de temperatura y precipitación, respectivamente. El clima es seco desértico y corresponde a la formula climática BWhw'(e) ¹⁶.

Los registros para cada vaca y lactancia incluyeron identificación de la vaca y sus progenitores, fecha de nacimiento y del parto, número de lactancia y, fecha y producción de leche por día (kg) registrada aproximadamente cada mes. Las lactancias, fueron de 1 a 14 y de la 7 en adelante se agruparon en la 7. Los registros comprendieron de 1973 a 1993, agrupando en 1981 las lactancia entre 1973 y 1981. Estas agrupaciones se realizaron con la finalidad de aumentar el número de registros en cada uno de los niveles de los efectos aprovechando así la información disponible e implica suponer que los cambios en producción diaria de leche durante la lactancia

son los mismos de la séptima a la catorceava lactancia y no existe diferencia entre las vacas con parto entre 1973 y 1981. Las épocas de partos se definieron de acuerdo a las estaciones del año y, por último, el día de prueba se definió agrupando a las vacas con registro de producción de leche el mismo día y año correspondiente, 148 días de registro.

La edición consistió en eliminar: a) las lactancias con menos de seis registros de producción mensual de leche, b) vacas con identificación errónea de sus padres, y c) producciones mensuales extremas, fuera del rango entre el y cuatro desviaciones estándar fenotípicas sobre a bajo el promedio de producción diaria de leche del mes de lactancia respectivo, resultando 18421 observaciones.

Cuadro 4.3. Composición de los modelos de funciones de covarianza y su abreviatura respectiva.

Regresión fija y aleatoria	Efecto fijo		
	AP, EP y NL ^a	AEPNL ^b	DR ^c
Ali y Schaeffer	APEPNL-AS	AEPNL-AS	DR-AS
Guo y Swalve	APEPNL-GS	AEPNL-GS	DR-GS
Wilmink	APEPNL-WL	AEPNL-WL	DR-WL
Wood	APEPNL-WO	AEPNL-WO	DR-WO
Orden del polinomio de Legendre			
Segundo	APEPNL-PL2	AEPNL -PL2	DR-PL2
Tercero	APEPNL-PL3	AEPNL -PL3	DR -PL3
Cuarto	APEPNL-PL4	AEPNL -PL4	DR -PL4
Quinto	APEPNL-PL5	AEPNL -PL5	DR -PL5

^aAP, EP y NL= Año de parto (AP), época de parto (EP) y número de lactancia (NL). ^bAEPNL = Combinación de AP, EP y NL. ^cDR = Día de prueba.

Modelos estadísticos. Se compararon 24 modelos de funciones de covarianza definidos por la combinación de las ecuaciones de regresión y los efectos fijos antes señalados, 8 y 3

respectivamente. En cada modelo, la ecuación de regresión fija, la del valor genético y del efecto de ambiente permanente, fue la misma. El cuadro 4.3 muestra las combinaciones de los efectos fijos y la regresión fija y aleatoria que originaron los 24 modelos. Los modelos se abreviaron con la simbología mencionada para los efectos fijos y regresión, separadas por un guión, por ejemplo el modelo con DR como efecto fijo y AS en la regresión fija y aleatoria es “DR-AS”.

En general, los modelos estadísticos usados corresponden al modelo mixto: $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wp} + \mathbf{e}$. Donde; \mathbf{y} , \mathbf{b} , \mathbf{u} , \mathbf{p} y \mathbf{e} son vectores de observaciones, desconocido de efectos fijos, desconocido de los coeficientes de regresión aleatoria del valor genético de cada animal, desconocido de coeficientes de regresión aleatoria del efecto de ambiente permanente de cada animal con registros, y de residuales respectivamente. En tanto \mathbf{X} , \mathbf{Z} y \mathbf{W} , son matrices conocidas que relacionan las observaciones con los; efectos fijos, coeficientes de regresión aleatoria del valor genético de cada animal y, coeficientes de regresión aleatoria de los efectos de ambiente permanente de cada animal con registros respectivamente. Las esperanzas y varianzas de los modelos son:

$$\mathbf{E} \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{Xb} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{ y } \text{var} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A} \otimes \mathbf{G}_0 & 0 & 0 \\ 0 & \mathbf{I} \otimes \mathbf{P}_0 & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{pmatrix} = \begin{pmatrix} \mathbf{A} \otimes \mathbf{K}_G & 0 & 0 \\ 0 & \mathbf{I} \otimes \mathbf{K}_P & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{pmatrix}$$

Las igualdades en las varianzas se originan de: $\text{Var}(\mathbf{Zu}) = \mathbf{Z}\text{Var}(\mathbf{u})\mathbf{Z}' = \mathbf{Z}\mathbf{K}_G\mathbf{Z}' = \mathbf{G}_0$, y, $\text{Var}(\mathbf{Wp}) = \mathbf{W}\text{Var}(\mathbf{p})\mathbf{W}' = \mathbf{W}\mathbf{K}_P\mathbf{W}' = \mathbf{P}_0$, para \mathbf{K}_G y \mathbf{K}_P , las matrices de coeficientes de las funciones de covarianza genética aditiva y de ambiente permanente, respectivamente⁴. Las matrices \mathbf{G}_0 y \mathbf{P}_0 son de (co)varianza genética y de ambiental permanente respectivamente. Las ecuaciones de los modelos mixtos:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{K}_G^{-1} \otimes \mathbf{A}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{I} \otimes \mathbf{K}_P^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

El logaritmo natural de la función de máxima verosimilitud restringida de los 24 modelos es, para N_a y N_p el número niveles del valor genético y de ambiente permanente respectivamente, \mathbf{C} la matriz de coeficientes de las ecuaciones del modelo mixto, y, $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$:

$$\ell = -\frac{1}{2}(N_a \log|\mathbf{K}_G| + k_A \log|\mathbf{A}| + N_p \log|\mathbf{K}_p| + \log|\mathbf{C}| + \log|\mathbf{R}| + \mathbf{y}'\mathbf{P}\mathbf{y})$$

El orden de \mathbf{K}_G y \mathbf{K}_P es igual al número de coeficientes de regresión en la ecuación usada en el valor genético aditivo y de ambiente permanente. Así, son de 3x3 para las ecuaciones de WO, WL, GS y PL2, de 4x4 en PL3, de 5x5 en AS y PL4, y de 6x6 en PL5. Mientras el número de parámetros a estimar en cada una de ellas, por efecto aleatorio es de 6, 10, 15 y 21 respectivamente.

La comparación de los modelos se realizó tomando en cuenta diferentes criterios: el valor de la función de máximo verosimilitud restringida y la gráfica de la producción diaria predicha y promedio durante la lactancia. La razón de verosimilitud se uso para los modelos con polinomios de Legendre, anidados. Los polinomios de Legendre se estimaron del segundo orden al de quinto, y en el momento que uno de ellos no tuviera solución, a partir de este los restantes no se estimaron. El análisis se realizó mediante el programa DFREML (Derivative Free Restricted Maximum Likelihood)¹⁷.

4.3. Resultados y discusión

Las medias y desviaciones estándar de producción diaria de leche por mes de lactancia, se presentan en el cuadro 4.4. La máxima producción de leche se estimó en el segundo mes de lactancia. La desviación estándar fue descendente, de 7.79 a 5.93 la mayor en el primer mes de lactancia. La producción diaria de leche es similar a las estimadas en Brasil^{18,19} y superiores a estimaciones en Alemania²⁰. Sin embargo, las desviaciones estándar no sólo son mayores y más uniformes durante la lactancia a las estimadas en Brasil y Alemania. Las estimaciones de la desviación estándar fenotípica de Araujo *et al.*¹⁹ aunque similares a las de este estudio, cambian durante la lactancia siendo menor en el primer mes de lactancia, mayor en el sexto mes y después disminuyó gradualmente.

En todos los modelos con las ecuaciones WL, WO y PL4 no se encontró solución adecuada. En los modelos con la ecuación de WL, las varianzas estimadas fueron muy grandes, mientras en los modelos con la ecuación de WO, y el PL4 no se logró convergencia. De esta forma los modelos con el PL5 no fueron analizados.

Cuadro 4.4. Número de observaciones (Obs.) media y desviación estándar (DS) de producción diaria de leche por mes de lactancia.

Mes	1	2	3	4	5	6	7	8	9	10
Obs.	2015	1950	1955	1938	1940	1939	1874	1798	1655	1536
Media	25.07	27.93	27.30	25.91	24.21	22.69	20.77	19.22	17.53	13.20
DS	7.79	7.61	7.57	7.26	7.07	6.97	6.66	6.39	6.24	5.93

Los modelos con mejor ajuste según el valor de la función de máxima verosimilitud restringida (ℓ) y diferencia entre observados y predichos fueron los modelos con DR como efecto fijo (Cuadro 4.5 y Figura 4.2). No obstante, existen diferencias entre los criterios, menores valores de ℓ correspondieron a los modelos DR-PL2, DR-PL3 y DR-GS, mientras por la figura 4.2 los mejores modelos fueron con DR-AS y Dr-GS. La figura 4.2 muestra subestimación al inicio, al pico y final de la lactancia de los modelos en DR-PL2 y DR-PL3, y en los primeros 5 días de la lactancia en DR-GS. El modelo DR-AS, aún cuando por ℓ fue el quinto mejor modelo (Cuadro 4.5), mostró menor sesgo en toda la lactancia (Figura 4.2).

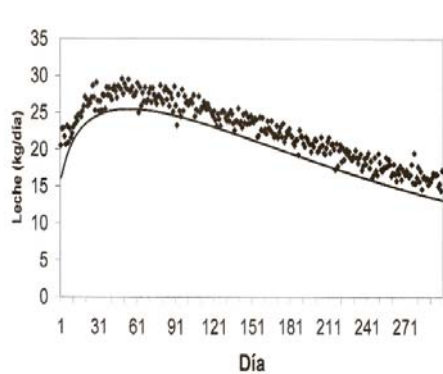
Cuadro 4.5. Logaritmo natural de la función de verosimilitud restringida para los modelos que se logró convergencia y el orden de ajuste (n) por el valor de la función de máxima verosimilitud.

Regresión	AP, EP y NL ^a	AEPNL ^b	DR ^c
Polinomio Cuadrático	-42180.0794 (4)	-41088.4085 (9)	-42531.9716 (1)
Polinomio Cúbico	-41956.3761 (6)	-40832.9273 (11)	-42244.6071 (2)
Ali y Schaeffer	-41804.9278 (8)	-40683.4594 (12)	-42073.0956 (5)
Guo y Swalve	-41923.3000 (7)	-40870.9167 (10)	-42184.8993 (3)

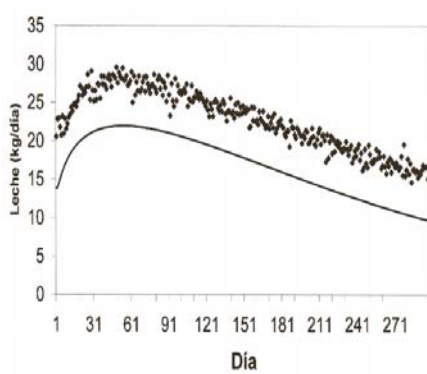
^aAP, EP y NL= Año de parto, época de parto y número de lactancia (NL). ^bAEPL = Combinación de AP, EP y NL. ^cDR = Día de prueba.

El ajuste de los modelos con AS y WL combinándolas entre la regresión fija y aleatoria no fue diferente, resultado de interés por la diferencia en el número de coeficientes de regresión entre las ecuaciones²¹. Similarmente, la comparación de polinomios de tercer a quinto grado, AS y WL, entre los polinomios, se estimó mejor ajuste a mayor orden, sin embargo, AS ajustó mejor que PL5 y WL mejor que PL3 y PL4. La combinación de polinomios de diferentes órdenes en la regresión fija y aleatoria no mejoro el ajuste del modelo²². Estos resultados son similares a los obtenidos en este estudio en los que las ecuaciones de AS y GS con 5 y 3 coeficientes de regresión mostraron buen ajuste.

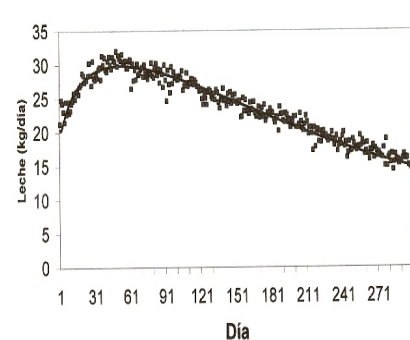
Ali y Schaeffer



AP, EP y NL^a-AS

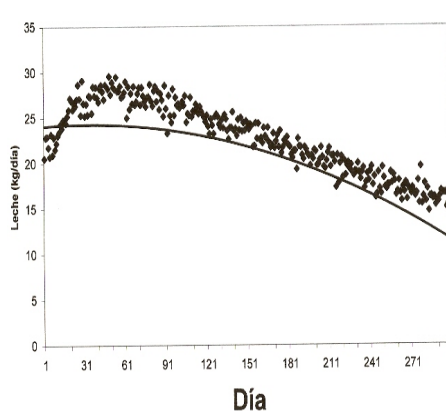


AEPNL^b-AS

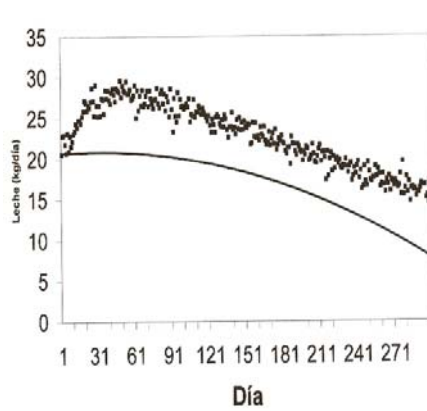


DR^c-AS

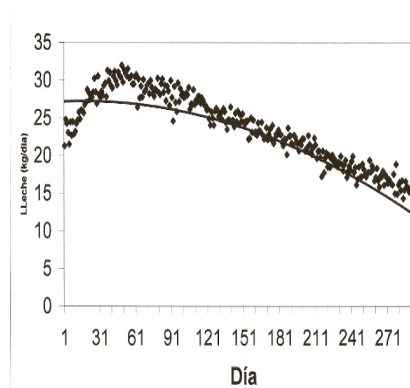
Polinomio de Legendre Cuadrático



AP, EP y NL^a-PL2



AEPNL^b-PL2



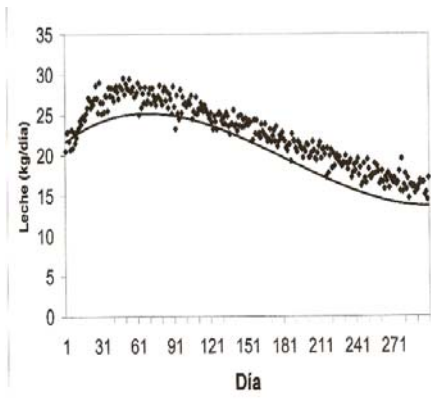
DR^c-PL2

— Predichos + Observado (promedio)

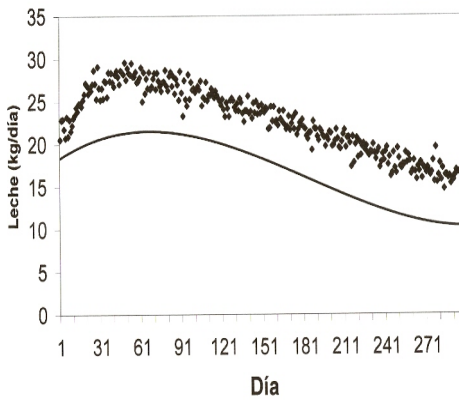
Efectos fijos: ^aAP, EP y NL= Año de parto (AP), época de parto (EP) y número de lactancia (NL). ^bAEPNL = Combinación de AP, EP y NL. ^cDR = Día de prueba.

Figura 4.2. Promedio de producción diaria de leche observada y producción predicha con diferentes modelos de funciones de covarianza.

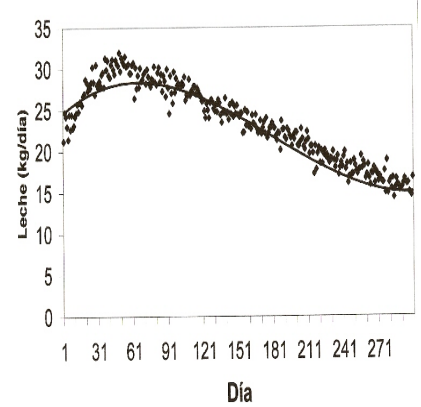
Polinomio de Legendre Cúbico



AP, EP y NL^a-PL3

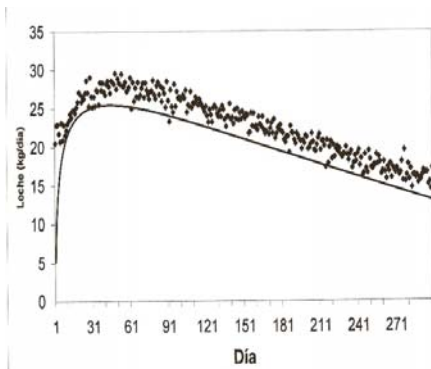


AEPNL^b-PL3

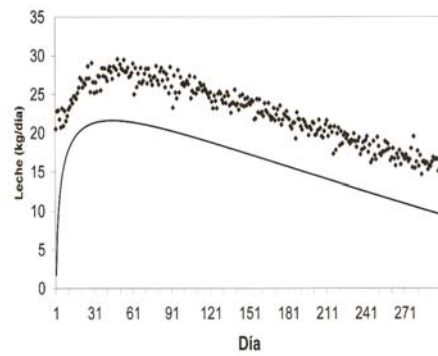


DR^c-PL3

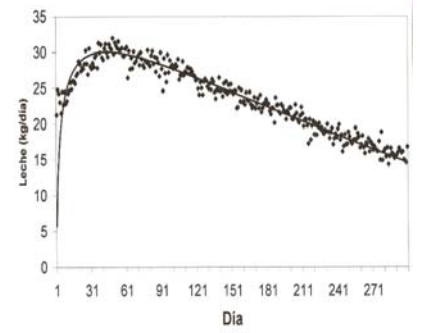
Guo y Swalve



AP, EP y NL^a-GS



AEPNL^b-GS



DR^c-GS

— Predichos + Observado (promedio)

Efectos fijos: ^aAP, EP y NL= Año de parto (AP), época de parto (EP) y número de lactancia (NL). ^bAEPNL = Combinación de AP, EP y NL. ^cDR = Día de prueba.

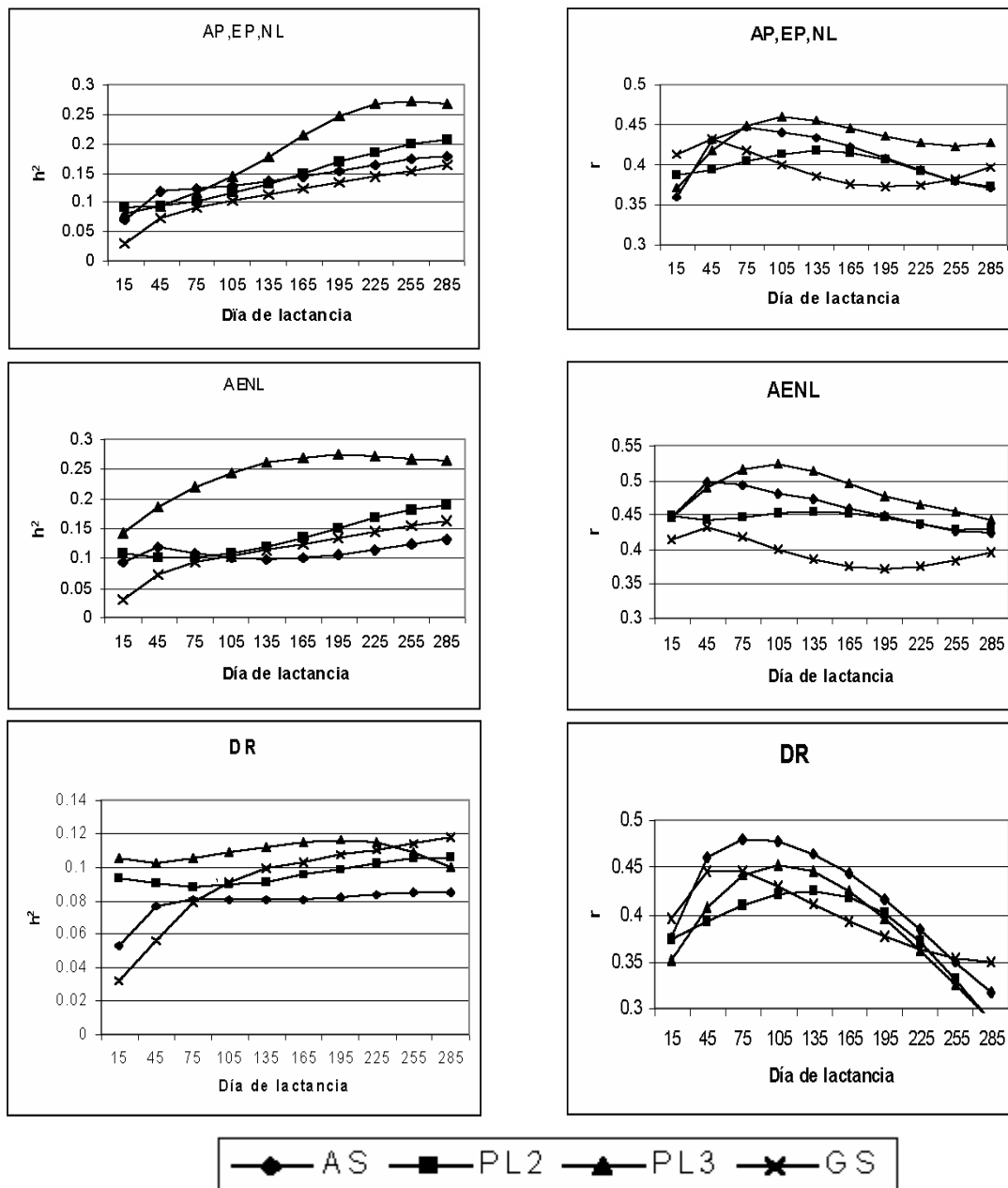
Figura 4.2 (Continuación). Promedio de producción diaria de leche observada y producción predicha con diferentes modelos de funciones de covarianza.

Entre las diferentes formas de reducir el número de parámetros a estimar, se encuentran suponer las varianzas genéticas (V_A) y ambiental permanente (V_P) de producción diaria de leche (PLD) no cambian durante la lactancia^{12,21,23}, comparando los modelos MV, AS y polinomios ortogonales suponiendo V_P de PLD es la misma o cambia gradualmente durante la lactancia, el mejor ajuste se logró supliendo cambios en la V_P , ajustando mejor el PL3¹². De igual forma, los TDM mostraron mejor ajuste que el MR, y entre los TDM ajustaron mejor los polinomios de Legendre de alto orden sobre todo en valor genético, teniendo el inconveniente del elevado número de parámetros a estimar²³. Este resultado puede ser atribuido, en parte, a que en general las pruebas como el valor de ℓ y la razón de verosimilitud favorecen a los modelos con más parámetros²⁴ así, es común, como en el presente estudio, encontrar superiores a modelos con 5 o 6 coeficientes de regresión, sin embargo, los resultados sugieren un buen ajuste con GS, ecuación con 3 coeficientes de regresión.

Los cambios en V_A y V_P de PLD son de gran magnitud, así, suponer son las mismas durante la lactancia, produce sesgos importantes en la predicción de la producción de leche, de esta forma, el MR no es adecuado. Los MV además de la desventaja del número de parámetros, si se desea describir las (co)varianzas en cada día de lactancia, se deben interpolar entre los días incluidos en el modelo, produciéndose un error adicional a la estimación de los componentes de varianza. El ajuste de los TDM es mayor suponiendo diferente varianza residual en las etapas de la lactancia^{22,23,25}, así el ajuste con GS podría ser equivalente al estimado con AS.

Las estimaciones de heredabilidad de producción diaria de leche durante la lactancia, mostraron grandes diferencias entre los modelos, sin embargo, en general, fue menor al inicio de la lactancia (0.03) y mayor al final (0.27), las mayores estimaciones correspondieron a AEPNL-PL3, y las menores a los modelos con DR, el modelo con peor y los de mejor ajuste (Figura 4.3). Las estimaciones de heredabilidad son muy sensibles al modelo mixto usado²⁶, con los modelos de regresión aleatoria (MRA) multivariado la heredabilidad estimada de PLD (h^2_{PLD}) fue mayor entre el quinto y sexto mes de lactancia, observando mayores estimaciones con MRA^{7,11}, igual tendencia en h^2_{PLD} (0.18, 0.33 y 0.26 al inicio, quinto mes y final) fue estimada, sin embargo, fueron más altas cuando la producción se registro a intervalos irregulares en comparación con intervalos regulares²⁰. Estimaciones de h^2_{PLD} en modelo con el polinomio de tercer orden fueron de 0.12, 0.25 y 0.15 al primer, séptimo u octavo, y doceavo

mes de lactancia en hatos de alta y promedio producción, mientras en hatos con baja producción fue menor al inicio, 0.12 y mayor al final, 0.15²⁷.



Efecto fijo: AP,EP,NL= año de parto, época de parto y número de lactancia; AENL= combinación de los tres anteriores; DR= día de prueba. AS= Ali y Schaeffer; PLi= polinomio de Legendre, i= 2, 3, 4, 5; GS= Guo y Swalve.

Figura 4.3. Heredabilidad (h^2) y repetibilidad (r) estimadas con diferentes modelos de funciones de covarianza.

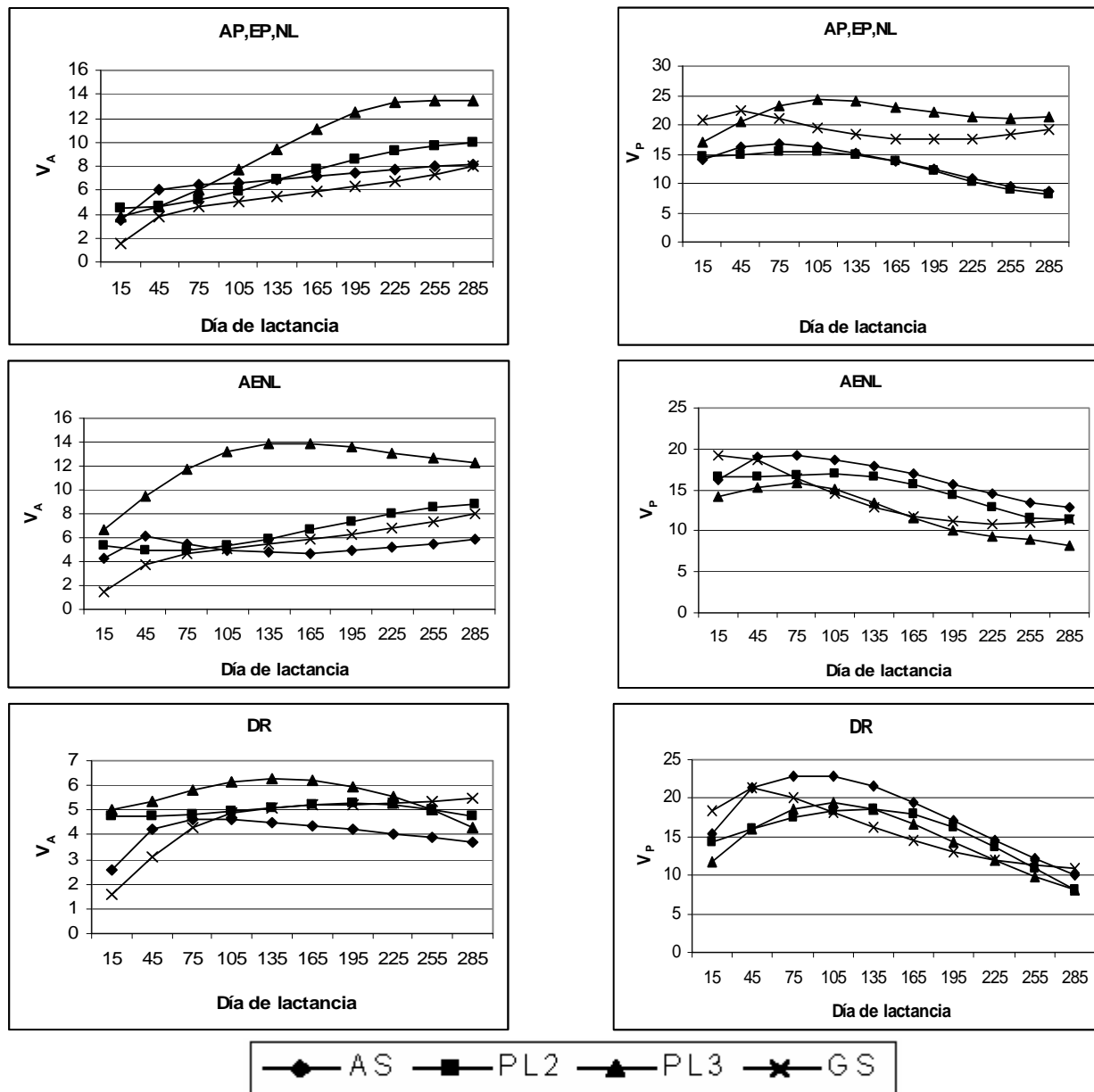
Con la ecuación de AS en el efecto genético y suponiendo constante el efecto de ambiente permanente, la h^2_{PLD} fue de 0.59 al inicio y posteriormente fluctuó alrededor de 0.4²⁸, sin embargo con el polinomio de Legendre, la h^2_{PLD} fue también mayor en la mitad de la lactancia²⁹, o bien, con polinomios, sólo difirieron al inicio de la lactancia, observándose aumento hasta el final de la lactancia, 0.20 a 0.45²⁵, o fue constante, 0.38, con polinomio lineal³⁰. Las mayores estimaciones al inicio y final de la lactancia, con los modelos AS y WL, se interpretó como posible inadecuación de estas ecuaciones para modelar las (co)varianzas, aún cuando lo sean adecuados para la curva de lactancia²³.

Las estimaciones de repetibilidad de producción diaria de leche (r_{PLD}) variaron entre 0.285 y 0.52, mostrando incremento hasta el segundo o quinto mes y posterior descenso hasta el final de la lactancia, donde se registraron los valores más bajos. Sin embargo, a diferencia de h^2_{PLD} , los modelos con el mejor ajuste no produjeron las estimaciones más bajas (Figura 4.3). Suponiendo el efecto permanente es el mismo en toda la lactancia, se estimó en 5.79 kg.², la varianza de efectos ambiental permanente de PLD (V_P), valor muy inferior a los del presente estudio²⁸. Estimaciones mayores y sin efecto de la aplicación de somatotropina fueron realizadas en Holstein, observándose r_{PLD} de 0.66, 0.61 y 0.69 en el primer, quinto y décimo mes de lactancia respectivamente³¹.

Las V_A estimadas fueron influenciada por el modelo, las varianzas fueron menores inicio de la lactancia (1.51 kg.²) y mayor al final, (13.88 kg.²). Los modelos con DR-AS y DR-GS, produjeron las estimaciones de varianza genética más bajas (Figura 4.4). Las estimaciones son menores a las de la literatura^{12,21,23,25,30}. Son de interés los cambios en las estimaciones de las varianzas por diferentes autores, mismos que, según el modelo, varían de mayor al inicio y menor al final, hasta incrementos desde el primer mes de lactancia^{12,21,23,30}. Las estimaciones de V_A son influidas por el número de lactancia y para lograr mejor estimación es posible la inclusión de la curva de lactancia en hato como aleatorio, así las estimaciones de heredabilidad son menores a las estimadas sin el efecto de hato, manteniendo la misma tendencia durante la lactancia³¹. Igual tendencia en la varianza genética, pero mayores valores y sin influencia del uso de somatotropina fueron estimadas en Holstein³².

Las estimaciones de V_P cambian según la ecuación de regresión usada en el modelo mixto y, a juzgar por las medidas de ajuste de los modelo, los polinomios de mayor orden son más

adecuados, estas mostraron valores más altos, siendo mayores al inicio y final de la lactancia, y similar a las estimadas con AS²³ en este estudio. Olori *et al.*²⁵ estimaron valores similares de varianza de ambiente permanente siendo más constantes durante la lactancia.



Efecto fijo: AP,EP,NL= año de parto, época de parto y número de lactancia; AENL= combinación de los tres anteriores; DR= día de prueba. AS= Ali y Schaeffer; PL_i= polinomio de Legendre, *i*= orden; GS= Guo y Swalve.

Figura 4.4. Varianzas (kg²) genéticas (V_A) y permanentes (V_P) estimadas con diferentes modelos de funciones de covarianza.

Las estimaciones de las correlaciones genéticas fueron de 0.99 a 1.0 entre meses contiguos, destacando la correlación genética de 1 en los últimos tres meses de la lactancia, y de 0.47 a 0.88 entre el primero y décimo mes (Figura 4.5). El modelo DR-GS mostró las mayores variaciones al inicio de la lactancia, confirmando el mayor sesgo observado en el modelo, ello, como fue antes mencionado, puede indicar que DR-GS es inadecuado para modelar las (co)varianzas genéticas²³. A diferencia de la literatura el modelo DR-AS, mostró cambios más graduales, confirmando su menor sesgo en toda la lactancia (Figura 4.2).

Las correlaciones genéticas fueron siempre positivas, sin embargo correlaciones negativas entre los primeros días con el tercer y hasta el décimo mes de lactancia han sido estimadas²². Correlaciones similares fueron estimadas con el polinomio de Legendre de cuarto orden, con 5 coeficientes de regresión, confirmando la superioridad de las ecuaciones con mayor número de parámetros^{21,22,23,25,26}.

Correlaciones genéticas cercanas a uno en los últimos tres meses de lactancia han sido también estimadas^{12,36}. Estas correlaciones y las estimaciones de heredabilidad, también para el final de la lactancia, son indicadores de: a) reducción, relativa, de los efectos no genéticos sobre la producción diaria de leche y, b) los genes involucrados en la producción diaria son los mismos en los últimos tres meses de la lactancia. Esto último, desde un punto de vista fisiológico, es atribuido a la disminución de requerimientos para producción al final de la lactancia.

Correlaciones genéticas negativas entre producción por día al inicio de la lactancia y en los días 7, 14 hasta el 315, fueron negativas cuando se incluyó la ecuación de WL en los modelos de regresión aleatoria, mientras AS produjo correlaciones genéticas negativas entre los días 7 con 14, 28, 42, 63, 91 y 315²². Las estimaciones de las correlaciones genéticas fueron siempre positivas en los TDM con polinomios de Legendre. El antagonismo generado por las correlaciones genéticas negativas es de gran importancia, pues complica los esquemas de selección y dificulta el posible cambio en la forma de la curva de lactancia. Así, es de mayor interés contar con el modelo adecuado.

Efecto fijo

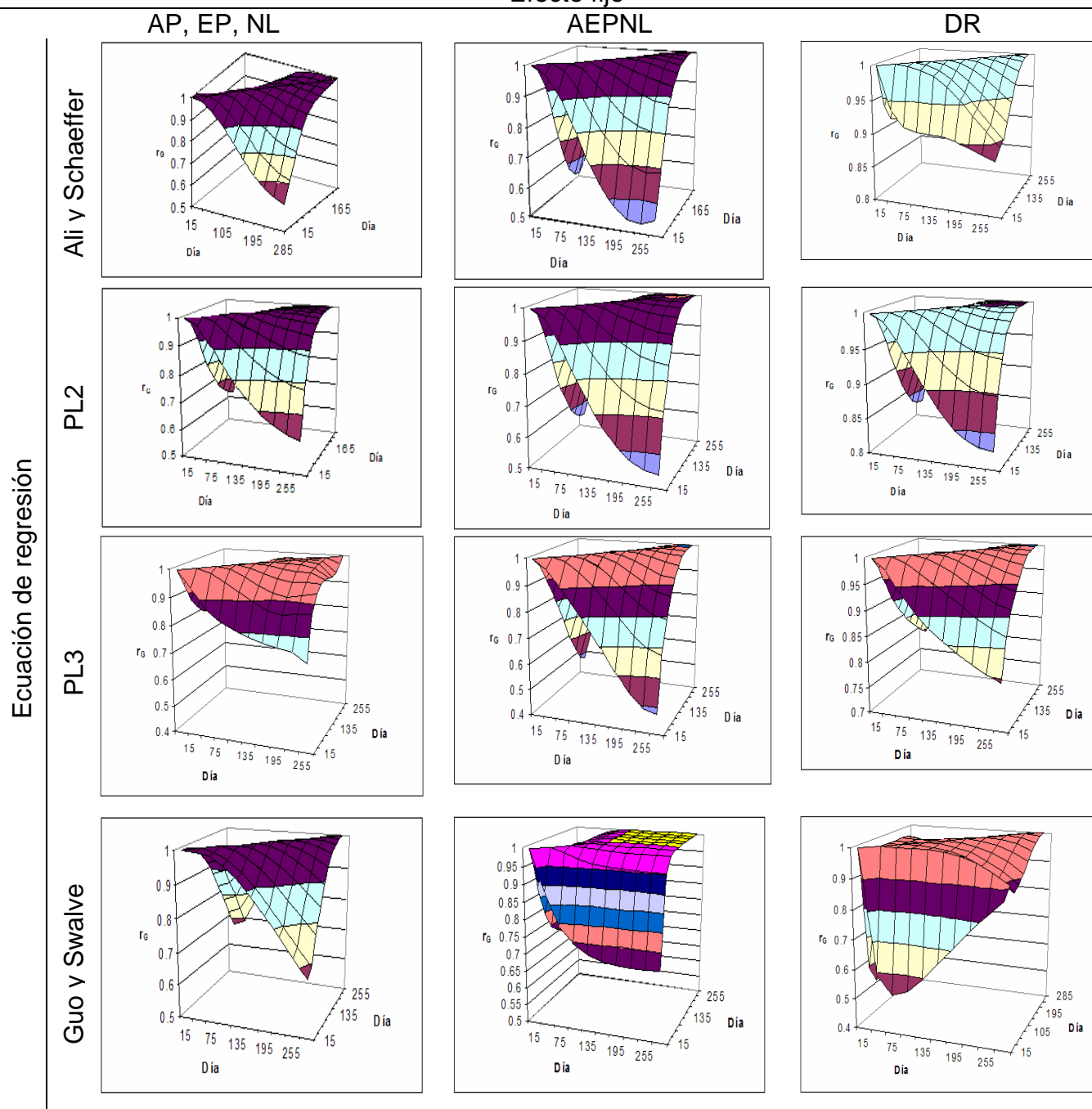


Figura 4.5. Correlaciones genéticas estimadas con diferentes modelos de funciones de covarianza.

Efecto fijo: AP,EP,NL= año de parto, época de parto y número de lactancia; AENL= combinación de los tres anteriores; DR= día de prueba. AS= Ali y Schaeffer; PLi= polinomio de Legendre,i= orden; GS= Guo y Swalve. PL2= polinomio cuadrático; PL3= polinomio cúbico.

4.4. Conclusiones

Las diferencias en el ajuste de los modelos con las estimadas en la literatura posiblemente se atribuyan a la suposición de la misma varianza residual durante la lactancia. Mayores varianzas residuales al inicio de la lactancia se observaron en los modelos DR-GS y los que incluyeron año de parto, época de parto y número de lactancia como efecto fijo. Y aún cuando los resultados evidenciaron esta falta de ajuste sólo en DR-GS, la falta de convergencia (PL4 y WO) y las grandes varianzas estimadas (WL), pueden ser causadas por la suposición de varianza residual homogénea. Además, debe considerarse el uso de ecuaciones que mejoren el ajuste sobre todo cuando la persistencia es muy alta³².

Las estimaciones de h^2_{PLD} durante la lactancia han sido muy variables, este estudio no fue la excepción. Algunas de causas de las diferencias son: a) la regresión aleatoria en el modelo, b) la inclusión o no de los efectos de ambiente permanente y la suposición sobre la estructura de sus covarianzas, y c) el efecto fijo en el TDM. Así, es difícil lograr consenso sobre el modelo adecuado. Si las diferencias entre los hatos son considerables y la reproducción juegue un papel importante en la producción³³, debe buscarse la forma de su inclusión en el TDM, existen algunas propuestas de cómo deben ser considerados en la literatura³⁴.

4.5. Literatura citada

1. Ali T. E. and L. R. Schaeffer. Accounting for among test day yields in dairy cows. *Can. J. Anim. Sci.* 67:637-644. 1987.
2. Meyer K. and W. G. Hill. Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood. *Livestock Production Science.* 47:185-200. 1997.
3. Rekaya R., M. J. Carbaño, and M. A. Toro. Use of test day yields for the genetic evaluation of production traits in Holstein – Friesian cattle. *Livest. Prod. Sci.* 1999; 57:203-217. 1999.
4. Kirkpatrick M., D. Lofsvold and M. Bulmer. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics.* 124:979-993. 1990.
5. Patck E. and L. R. Schaeffer. Use of test day yields for genetic evaluation of dairy sires and cows. *Livest. Prod. Sci.* 34:23-34. 1993.
6. Gengler N., A. Tijani, G. R. Wiggans, and J. C. Philpot. 1997. Estimation of (co)variance components of test day for U. S. Holstein. In: *Proceeding of the 1997 Interbull meeting, Vienna, Austria, August 28-29, 1997: 53-58.*
7. Meyer K, H. U. Graser and K. Hammond. Estimates of genetic parameters for first lactation test day production of Australian black and white cows. *Livest. Prod. Sci.* 21:177-199. 1989.
8. Pander B. L., W. G. Hill and R. Thompson. Genetic parameters of test day records of British Holstein-Friesian heifers. *Anim. Prod.* 55:11-21. 1992.
9. Gadini, Claudia H., Jeffrey F. Keown and L. Dale Van Vleck. Genetic Parameters of test day milk. Fat and Protein Yields. In: *6th Congress on Genetics Applied to Livestock Production, Armidale, Australia. 23:303. 1998.*
10. Goulart M. S., M. A. Ramalho de Freitas and C. H. Gadini. Genetic parameters of test-day milk yield of Holstein cows. In: *6th Congress on Genetics Applied to Livestock Production. Armidale, Australia. 23:311. 1998.*
11. Kettunen A., E. A. Mäntysaari, I. Strandén, J. Pösö and M. Lidauer. Estimation of genetic parameters for first lactation test day milk production using random regression models. In:

- 6th World Congress on Genetics Applied to Livestock Production. Armidale, Australia. 23:307-310. 1998.
12. Kettunen A., E. A. Mäntysaari, and J. Pösö. Estimation of genetic parameters for daily milk yield of primiparous Ayrshire cows by random regression test-day models. *Livest. Prod. Sci.* 66:251-261. 2000.
13. Wood P. D. P. Algebraic model of the lactation curves in cattle. *Nature*, 216: 164 -165. 1967.
14. Wilink, J. B. M. Adjusted of test-day milk, fat and protein yields for age, season and stage of lactation. *Livest. Prod. Sci.* 16:335-348. 1987.
15. Guo, Z., and H. H. Swalve. 1995. Modelling of the lactation curves as sub-model in the evaluation of test day records. *INTERBULL*, Prague, sep. 7-8.
16. García de M. E. Modificaciones al Sistema de Clasificación Climática de Köppen. Impreso en México. 1981.
17. Meyer K. DFREML, User notes. Version 3.0 1997.
18. Machado, S. G., M. A. R. Freitas and C. H. Gadini. Genetic parameters of test day milk yields of Holstein cows. *Genetics and Molecular Biology.* 22:383-.86. 1999
19. Araujo-Cobuci, J., R. F. Euclides, P. Sávio-Lopez, C. N. Costa, R. de Almeida-Torres and C. Silva P. Estimation of genetic parameters for test-day milk yield in Holstein cows using random regression model. *Genetics and Molecular Biology.* 28:75-83 .2005
20. Swalve H. H. The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. *J. Dairy Sci.* 78:929-938. 1995.
21. Jamrozik J., G. J. Kistemaker, J. C. M. Dekkers, and L. R. Schaeffer. Comparison of possible covariates use in random regression model for analyses of test day yields. *J. Dairy Sci.* 80:2550-2556. 1997.
22. Brotherstone S, I. M. S. White and K. Meyer. Genetic modeling of dairy milk yield using orthogonal polynomials and parametric curves. *Anim. Science.* 70:407-415. 2000.
23. López-Romero y M. J. Carabaño. Evaluación de modelos aleatorios para el análisis de datos de producción de leche a lo largo de la lactación. <http://www.etsia.upv.es/Acteón/docs/PAPER-new.pdf> consultado 21/1/2002.

24. Stram, D. O. and J. W. L. Lee. Variance component testing in the longitudinal mixed model. *Biometrics*. 50:1171-1177. 1994.
25. Olori V. E., W. G. Hill, B. J. McGuirk, and S. Brotherstone. Estimating variance components for test day records by restricted maximum likelihood with random regression animal model. *Livest. Prod. Sci.* 61:53-63. 1999.
26. Jensen, J. Genetic evaluation of dairy cattle using test-day models. *J. dairy Sci.* 84:2803-2812. 2001.
27. Bormann, J., G. R. Wiggins, T. Druet, and N. Gengler. Within-Herd effects of age at test day and lactation stage on test-day yields. *J. Dairy Sc.* 86:3765-3774. 2003.
28. Jamrozik, J. and L. R. Schaeffer. Estimates of genetic parameters for a test day model with random regression for yield traits of first lactation Holstein. *J. Dairy Sci.* 80:762-770. 1997
29. Veerkamp, R. F. and M. E. Goddard. Covariance functions across herd production levels for test day records on milk, fat, and protein yields. *J. Dairy Sci.* 81:1690-1701. 1998.
30. Pool, M. H., L. L. G. Janss, and T. H. E. Neuwissen. Genetic parameters of Legendre polynomials for first parity lactation curves. *J. Dairy Sci.* 83:2640-2649. 2000.
31. De Ross, A. P. W., A. G. F. Harbers, and J. de Jon. Random herd curves in a test-day model for milk, fat and protein production of dairy cattle in Netherlands. *J. Dairy Sci.* 87:2693-2701. 2004.
32. Tsuruta, S., J. F. Keown, L. D. Van Vleck, and I. Misztal. Bias in genetic evaluation by records of cows treated with somatotropin. *J. Dairy Sc.* 83:2650-2656. 2000.
33. Grossman, M., S. M. Hartz, and W. J. Koops. Persistency of yields lactation: a novel approach. *J. Dairy Sci.* 82:2192-2197. 1999.
34. Lopez, H, L. D. Satter, and M. C. Wiltbank. Relationship between level of milk production and estrous behavior of lactation dairy cows. *Anim. Repro. Sci.* 81:209-223. 2004
35. Trus D. and L. G. Buttazzoni. A multiple trait approach to modeling the lactation curve. In *Proc. 4th World congress on Genetics Applied to livestock Production*, Edinburg. 14:492-495. 1994.

36. Lui, Z., F. Reinhardt and R. Reents. Estimating parameters of a random regression test day model for first three lactation milk production traits using the covariance function approach. In: Proceeding of the 2000 Interbull Meeting, Bled, Slovenia. May 14 -15, 2000. 74-78.

V. Conclusiones generales

Además de la regresión fija, el efecto DR, mostró remover con mayor eficiencia los efectos ambientales temporales sistemáticos, sin embargo, cuando se está considerando producciones de diferentes lactancias, una alternativa de interés es la combinación DR-número de lactancia.

En general los resultados coinciden con la literatura dejando claro que las ecuaciones con más parámetros tienen mejor ajuste, sin embargo el ajuste obtenido con la ecuación de Guo y Swalve es sólo cuestionable en los primeros días de lactancia. Ello posiblemente se solucione suponiendo heterogeneidad de varianzas, siendo mayor la varianza residual al inicio de la lactancia.

La suposición de varianza residual homogénea, puede ser la explicación de la diferencia en la tendencia de las estimaciones de heredabilidad y las de la literatura, sin embargo es también de interés la gran variabilidad en estimaciones de los parámetros genéticos atribuibles a las ecuaciones de regresión usadas en el modelo encontrada en la literatura.

Con fines de considerar la variación causada por el número de lactancia en los modelos de regresión aleatoria o funciones de covarianza, es conveniente, probar la combinación día de prueba-número de lactancia, esta define con mayor énfasis biológico el grupo de comparación y requiere de menos parámetros en la función de covarianza que el modelo multivariado. Sin embargo puede reducir el número de observaciones por celda, con su consecuente efecto sobre la exactitud de los valores genéticos.

Considerar los valores genéticos y la varianza residual son los mismos en las diferentes lactancias y la varianza residual es homogénea durante la lactancia, son, posiblemente, las razones de mayor peso para explicar las diferencias entre los parámetros genéticos estimados y los de la literatura. Para este estudio, sin embargo el reducido número de observaciones no permitió suponer el valor genético de los animales también cambia entre lactancias.

Quizás la opción de mayor interés es, suponer que el valor genético de producción diaria de leche cambia tanto en la misma como en diferentes lactancias, así, el modelo es adecuado es multivariado, lo que aumenta, significativamente, el número de parámetros a estimar. Así el efecto de número de lactancia deja de ser de interés como efecto fijo en el modelo.

Son varias las formas de usar el TDM, y aún cuando existe incertidumbre sobre el modelo adecuado, para algunos países o regiones, la corrección previa de los registros de producción por día por los efectos fijos importantes y posterior estimación de los parámetros genéticos de la producción por lactancia y predicción de los valores genéticos es una forma de hacer uso de sus ventajas.

El uso de los valores genéticos de partes de la lactancia, producción máxima y persistencia por ejemplo, debe ser sujeto a mayor estudio, sobretodo, en relación a sus beneficios, el más estudiado es su relación con el comportamiento reproductivo. Sin embargo la metodología del TDM puede aportar más información.

Para mejorar la comparación de los modelos es necesario, incluir un mayor número de observaciones para así poder usar pruebas con mayor potencia y la validación cruzada.

El modelo multivariado, no sólo es impractico para estimar la estructura de covarianza de producción diaria de leche, sino también, por la correlación genética de uno al final de la lactancia, la convergencia resultará difícil de lograr.

A2.1. Anexo 1. Regresión aleatoria

Como ejemplo del modelo de regresión aleatoria, supóngase se tiene la producción de leche por día de cuatro vacas. Usando la ecuación de Guo y Swalve¹⁴ en la regresión fija y aleatoria, los efectos genético de los animales. En la información, las vacas 3 y 4 son hijas del toro 1, y, las vacas 5 y 6 hijas del toro 2. El Cuadro A1.1 muestra las observaciones.

Cuadro A1.1. Producción diaria de leche de cuatro vacas.

Vaca	Días en Leche (DL)	Raíz de DL	Logaritmo natural de (DL)	Producción de leche (kg/día)
3	5	2.24	1.61	10
3	45	6.71	3.81	25
3	80	8.94	4.38	30
3	110	10.49	4.70	28
3	140	11.83	4.94	25
4	25	5.00	3.22	19
4	55	7.42	4.01	35
4	90	9.49	4.50	28
4	120	10.95	4.79	22
4	150	12.25	5.01	18
5	18	4.24	2.89	27
5	52	7.21	3.95	35
5	85	9.22	4.44	31
5	115	10.72	4.74	25
6	15	3.87	2.71	26
6	45	6.71	3.81	37
6	75	8.66	4.32	32
6	105	10.25	4.65	28
6	135	11.62	4.91	26

El modelo de regresión aleatoria es:

$$y_{ij} = b_0 + b_1 t^{(0.5)} + b_2 \ln t + \alpha_{0j} + \alpha_{1j} t^{(0.5)} + \alpha_{2j} \ln t + e_{ij}$$

donde : y_{tj} es la producción de leche por día el día t de la lactancia del j -ésimo animal, b_i ($i=0,1,2$) son los coeficientes de regresión que cuantifican el efecto del día de lactancia sobre la producción diaria de leche en todo el conjunto de datos, α_{ij} ($i = 0,1,2$) son los coeficientes de regresión del valor genético de producción de leche por día el día t de la lactancia del j -ésimo animal ($j=1,2,\dots,6$), y e_{tj} el error.

Los primeros tres coeficientes de regresión, los b_i , corresponden a la regresión fija y describe la curva promedio del hato. Los siguientes tres, los α_{ij} , corresponden a la regresión aleatoria que describe las desviaciones, con respecto a la curva general, causadas por los valores genéticos de cada uno de los animales.

Matricialmente el modelo es:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

Donde, \mathbf{y} , \mathbf{b} , \mathbf{u} y \mathbf{e} , son vectores de observaciones, de coeficientes desconocidos de regresión fija, de coeficientes de regresión aleatoria (también desconocidos), de los valores genéticos de cada animal, y de residuales; \mathbf{X} y \mathbf{Z} son matrices con elementos conocidos que relacionan las observaciones con los coeficientes de regresión fija y aleatoria, respectivamente.

Las esperanzas y varianzas del modelo son:

$$E \begin{bmatrix} y \\ u \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \end{bmatrix}$$

y

$$\text{Var} \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} G \otimes A & 0 \\ 0 & \sigma_e^2 \end{bmatrix}$$

Los vectores y matrices son, entonces:

$$\mathbf{y}' = (10 \ 25 \ 30 \ 28 \ 25 \ 19 \ 35 \ 28 \ 22 \ 18 \ 27 \ 35 \ 31 \ 25 \ 26 \ 37 \ 32 \ 28 \ 26),$$

$$\mathbf{b}' = (b_0 \ b_1 \ b_2),$$

$$\mathbf{a}' = (\alpha_{01} \ \alpha_{11} \ \alpha_{12} \ \alpha_{02} \ \alpha_{12} \ \alpha_{22} \ \alpha_{03} \ \alpha_{13} \ \alpha_{23} \ \alpha_{04} \ \alpha_{14} \ \alpha_{24} \ \alpha_{05} \ \alpha_{15} \ \alpha_{25} \ \alpha_{06} \ \alpha_{16} \ \alpha_{26}),$$

$$\mathbf{X} = \begin{bmatrix} 1 & 224 & 1.61 \\ 1 & 671 & 3.81 \\ 1 & 894 & 4.38 \\ 1 & 1049 & 4.70 \\ 1 & 1183 & 4.94 \\ 1 & 5.00 & 3.22 \\ 1 & 7.42 & 4.01 \\ 1 & 9.49 & 4.50 \\ 1 & 1095 & 4.79 \\ 1 & 1225 & 5.01 \\ 1 & 424 & 2.89 \\ 1 & 7.21 & 3.95 \\ 1 & 9.22 & 4.44 \\ 1 & 1072 & 4.74 \\ 1 & 3.87 & 2.71 \\ 1 & 6.71 & 3.81 \\ 1 & 8.66 & 4.32 \\ 1 & 1025 & 4.65 \\ 1 & 1162 & 4.90 \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 224 & 1.67 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 671 & 3.81 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 894 & 4.38 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1049 & 4.70 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1183 & 4.94 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 5.00 & 3.22 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 7.42 & 4.01 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 9.49 & 4.50 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1095 & 4.79 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1225 & 5.01 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 424 & 2.89 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 7.21 & 3.95 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 9.22 & 4.44 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1072 & 4.75 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 3.87 & 2.71 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 6.71 & 3.1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 8.66 & 4.62 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1025 & 4.65 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1162 & 4.90 \end{bmatrix}$$

La matriz del numerador del coeficiente de parentesco y de (co)varianzas genéticas entre los coeficientes de regresión del valor genético son:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1 & 1/4 \\ 0 & 1/2 & 0 & 0 & 1/4 & 1 \end{bmatrix}$$

y

$$\mathbf{G}_0 = \begin{bmatrix} 4.5 & 2.3 & 1.0 \\ 2.3 & 3.8 & 1.9 \\ 1.0 & 1.9 & 4.3 \end{bmatrix}$$

Con inversas:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.67 & 0 & -0.67 & -0.67 & 0 & 0 \\ 0 & 1.67 & 0 & 0 & -0.67 & -0.67 \\ -0.67 & 0 & 1.33 & 0 & 0 & 0 \\ -0.67 & 0 & 0 & 1.33 & 0 & 0 \\ 0 & -0.67 & 0 & 0 & 1.33 & 0 \\ 0 & 0.67 & 0 & 0 & 0 & 1.33 \end{bmatrix}$$

y

$$\mathbf{G}_0^{-1} = \begin{bmatrix} 0.3225 & -0.2024 & 0.0144 \\ -0.2024 & 0.4648 & -0.1583 \\ 0.0144 & -0.1583 & 0.2992 \end{bmatrix}$$

La matriz de residuales es la matriz idéntica (de orden igual al número de observaciones) multiplicada por la varianza del error, $\mathbf{R}_0 = \mathbf{I}_{19}\sigma_e^2$. Su inversa por lo tanto, es la idéntica por el inverso de la varianza del error. Así, suponiendo para el ejemplo $\sigma_e^2 = 5.0$, la inversa de la matriz de residuales, es, entonces: $\mathbf{R}_0^{-1} = \mathbf{I}_{19}(1/5)$.

Las ecuaciones de los modelos mixtos:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

Donde:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \begin{bmatrix} 3.80 & 31.56 & 15.48 \\ 31.56 & 293.00 & 137.82 \\ 15.48 & 137.82 & 65.98 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} = \begin{bmatrix} 101.40 \\ 853.08 \\ 420.15 \end{bmatrix}$$

$$ZR^{-1}Z + G_0^{-1} \otimes A^{-1} =$$

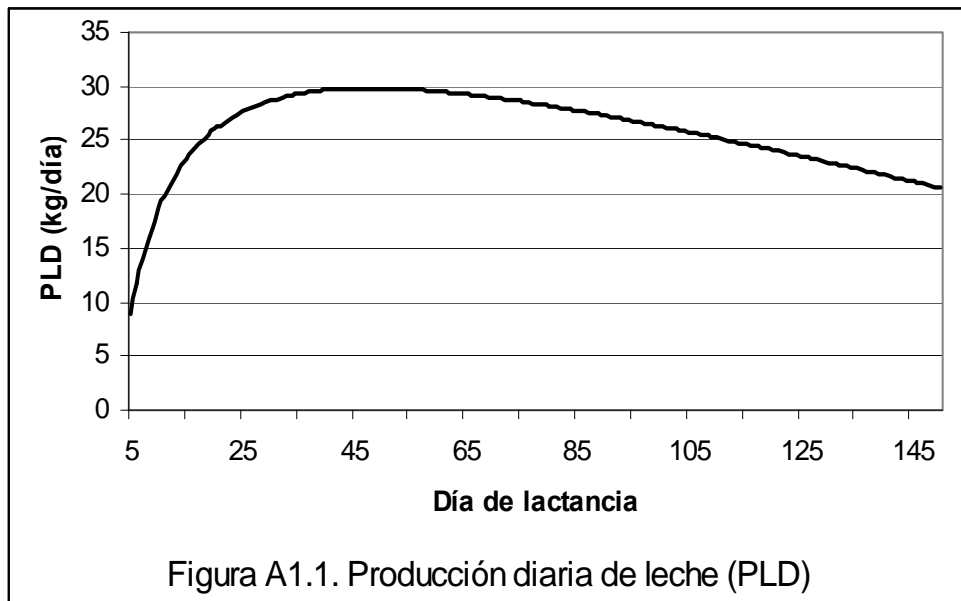
0.54	0.00	-0.22	-0.22	0.00	0.00	-0.34	0.00	0.13	0.13	0.00	0.00	0.02	0.00	-0.01	-0.01	0.00	0.00
0.00	0.54	0.00	0.00	-0.22	-0.22	0.00	-0.34	0.00	0.00	0.13	0.13	0.00	0.02	0.00	0.00	-0.01	-0.01
-0.22	0.00	0.43	0.00	0.00	0.00	0.13	0.00	-0.27	0.00	0.00	0.00	-0.01	0.00	0.02	0.00	0.00	0.00
-0.22	0.00	0.00	0.43	0.00	0.00	0.13	0.00	0.00	-0.27	0.00	0.00	-0.01	0.00	0.00	0.02	0.00	0.00
0.00	-0.22	0.00	0.00	0.43	0.00	0.00	0.13	0.00	0.00	-0.27	0.00	0.00	-0.01	0.00	0.00	0.02	0.00
0.00	-0.22	0.00	0.00	0.00	0.43	0.00	0.13	0.00	0.00	0.00	-0.27	0.00	-0.01	0.00	0.00	0.00	0.02
-0.34	0.00	0.13	0.13	0.00	0.00	1.77	8.04	3.58	-0.31	0.00	0.00	-0.27	0.00	0.11	0.11	0.00	0.00
0.00	-0.34	0.00	0.00	0.13	0.13	8.04	7677	3522	0.00	-0.31	-0.31	0.00	-0.26	0.00	0.00	0.11	0.11
0.13	0.00	-0.27	0.00	0.00	0.00	3.58	3522	1718	0.00	0.00	0.00	0.11	0.00	-0.21	0.00	0.00	0.00
0.13	0.00	0.00	-0.27	0.00	0.00	-0.31	0.00	0.00	1.62	9.02	4.30	0.11	0.00	0.00	-0.21	0.00	0.00
0.00	0.13	0.00	0.00	-0.27	0.00	0.00	-0.31	0.00	9.02	8662	4046	0.00	0.11	0.00	0.00	-0.21	0.00
0.00	0.13	0.00	0.00	0.00	-0.27	0.00	-0.31	0.00	4.30	4046	1956	0.00	0.11	0.00	0.00	0.00	-0.21
0.02	0.00	-0.01	-0.01	0.00	0.00	-0.27	0.00	0.11	0.11	0.00	0.00	1.30	6.28	3.01	-0.20	0.00	0.00
0.00	0.02	0.00	0.00	-0.01	-0.01	0.00	-0.27	0.00	0.00	0.11	0.11	6.28	545	2652	0.00	-0.20	-0.11
-0.01	0.00	0.02	0.00	0.00	0.00	0.11	0.00	-0.21	0.00	0.00	0.00	3.01	2652	1364	0.00	0.00	0.00
-0.01	0.00	0.00	0.02	0.00	0.00	0.11	0.00	0.00	-0.21	0.00	0.00	-0.20	0.00	0.00	1.40	8.22	4.08
0.00	-0.01	0.00	0.00	0.02	0.00	0.00	0.11	0.00	0.00	-0.21	0.00	0.00	-0.11	0.00	8.22	7540	3561
0.00	-0.01	0.00	0.00	0.00	0.02	0.00	0.11	0.00	0.00	0.00	-0.21	0.00	-0.11	0.00	4.08	3562	1764

$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} = (\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z})' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1.00 & 8.04 & 3.89 \\ 8.04 & 76.00 & 35.22 \\ 3.89 & 35.22 & 16.56 \\ 1.00 & 9.02 & 4.30 \\ 9.02 & 88.00 & 40.46 \\ 4.30 & 40.46 & 18.94 \\ 0.80 & 6.28 & 3.21 \\ 6.28 & 54.00 & 26.52 \\ 3.21 & 26.52 & 13.24 \\ 1.00 & 8.22 & 4.08 \\ 8.22 & 75.00 & 35.62 \\ 4.08 & 35.62 & 17.24 \end{bmatrix} \quad \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} = \begin{bmatrix} 0.00 \\ 0.00 \\ 0.00 \\ 0.00 \\ 0.00 \\ 0.00 \\ 23.60 \\ 209.57 \\ 99.58 \\ 24.40 \\ 216.33 \\ 104.59 \\ 23.60 \\ 184.17 \\ 94.54 \\ 29.80 \\ 243.01 \\ 121.45 \end{bmatrix}$$

La solución:

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \alpha_{01} \\ \alpha_{11} \\ \alpha_{21} \\ \alpha_{02} \\ \alpha_{12} \\ \alpha_{22} \\ \alpha_{03} \\ \alpha_{13} \\ \alpha_{23} \\ \alpha_{04} \\ \alpha_{14} \\ \alpha_{24} \\ \alpha_{05} \\ \alpha_{15} \\ \alpha_{25} \\ \alpha_{06} \\ \alpha_{16} \\ \alpha_{26} \end{bmatrix} = \begin{bmatrix} -13.38 \\ -6.69 \\ 23.14 \\ -0.45 \\ 0.41 \\ -0.64 \\ -0.63 \\ 0.02 \\ -0.09 \\ -0.65 \\ 0.65 \\ -0.97 \\ -0.96 \\ 0.01 \\ -0.07 \\ 0.80 \\ -0.01 \\ 0.62 \\ 0.64 \\ -0.16 \\ 1.04 \end{bmatrix} \quad \begin{array}{l} \text{Regresión fija} \\ \text{Regresiones aleatorias para predecir el} \\ \text{valor genético de producción por día en} \\ \text{cada día de lactancia de cada uno de los} \\ \text{6 animales.} \end{array}$$

En el vector de soluciones, los b_i , primeros tres elementos del vector de soluciones, corresponden a la regresión fija, los 18 restantes corresponden a las ecuaciones de predicción de los valores genéticos de producción diaria de leche para cada uno de los 6 animales en la base de datos. La figura A1.1 muestra la curva de producción diaria de leche entre los 5 y 150 días de lactancia que se observaron en los registros. La curva promedio del hato mostró producción diaria, a los 5 y 150 días de 8.89 y 20.60 kg. día⁻¹, respectivamente. El tiempo a máxima producción, la producción máxima y la acumulada en los 146 días de lactancia (del día 5 al 150), fueron: 48 días, 29.83 kg. Día⁻¹ y 3777.9 kg. respectivamente (Cuadro A1.2).



Los valores genéticos de un animal cualquiera, ya sea para un día o el acumulado en un periodo de lactancia, se obtienen con facilidad de la ecuación de predicción del animal respectivo. Tomando al animal 6, la ecuación de predicción es:

$$\hat{A}_{6,t} = 0.6419 - 0.1563t^{0.5} + 1.0429 \ln t$$

El valor genético el día 48 de la lactancia, día de máxima producción de la información es:

$$\hat{A}_{6,48} = 0.6919 - 0.1563(48)^{0.5} + 1.0459 \ln(48) = 3.60 \text{ kg. día}^{-1}$$

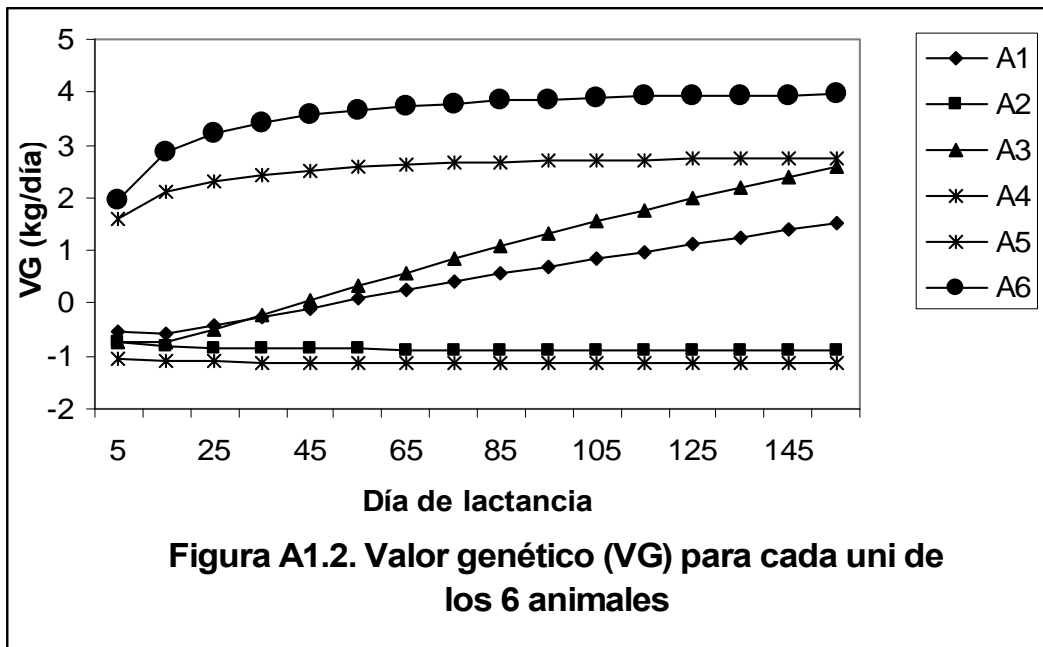
El valor genético acumulado entre los días 5 y 150 de la lactancia, es la suma de los valores genéticos predichos en ese periodo, entonces:

$$\hat{A}_{6, \sum_{t=5}^{150} t} = \sum_{t=5}^{150} (0.6919 - 0.1563 t^{0.5} + 1.0459 \ln(t))$$

$$\hat{A}_{6, \sum_{t=5}^{150} t} = 146(0.6919) - 0.1563 \sum_{t=5}^{150} t^{0.5} + 1.0459 \sum_{t=5}^{150} \ln t$$

$$= 101.0174 - 0.1563(1224.5278) + 1.0459(601.8420) = 529.99 \text{ kg.}$$

La figura A1.2 muestra los valores genéticos de los 6 animales para cada uno de los 146 días de la lactancia. Es importante notar que los animales 5 y 6 tienen valores genéticos positivos en los 146 días de lactancia estudiados. El valor genético de los animales 2 y 4 fue siempre negativo con poco cambio en los 146 días, mientras los animales 1 y 3 tienen valores genéticos negativos hasta aproximadamente el pico de lactancia (48 días) y posteriormente son positivos.



La producción de leche para el promedio del hato y los valores genéticos de cada uno de los animales en los días 5 y 48 de la lactancia, y la acumulada entre el día 5 y 150 de la lactancia se muestran en el cuadro A1.2. El valor genético de producción en 146 días de los animales 2 y 4 fue negativo y es causado por valores genéticos de producción por día negativos durante los 146 días en animales. El valor genético, también de producción acumulada en 146 días, de los animales 1, 3, 5 y 6 fue positivo, sin embargo, los animales 1 y 3 tuvieron valores genéticos de producción diaria negativos los primeros 48 días y posteriormente positivos, mientras en animales 5 y 6 fue siempre positivo.

Cuadro A1.2. Ecuaciones de regresión de producción diaria de leche y valor genético de producción diaria de leche de los seis animales y producción y valor genético en diferentes días de lactancia y acumulada entre el día 5 y 150.

Coeficiente de regresión	General	Animal						
		1	2	3	4	5	6	
0	-13.38	-0.45	-0.63	-0.65	-0.96	0.80	0.64	
1	-6.69	0.41	0.02	0.66	0.01	-0.09	-0.16	
2	23.14	-0.64	-0.09	-0.97	-0.07	0.62	1.04	
Día	Producción	Valor Genético						
	-----kg. día ⁻¹ -----							
5	8.89	-0.54	-0.74	-0.75	-1.03	1.58	1.97	
48	29.83	-0.03	-0.87	0.13	-1.13	2.54	3.60	
150	20.60	1.45	-0.89	2.50	-1.13	2.74	3.95	
$\sum_{T=5}^{150} T$	-----kg.-----							
	3777.90	60.19	-126.78	122.50	-164.14	373.08	529.99	

A2.2. Anexo 2. Predicción de la estructura de covarianza con funciones de covarianza según Kirkpatrick y colaboradores

La estimación de componentes de (co)varianza con funciones de covarianza y polinomios de Legendre, es más fácil de entender si se interpreta como extrapolación de la estructura de (co)varianza de una característica de expresión continua, para ello es indispensable estimar primero parte de la estructura de covarianza, $\hat{\mathbf{G}}$. La mejor forma de realizar esta extrapolación es mediante una ecuación de predicción, ecuación denominada por Kirkpatrick *et al.*³², función de covarianza en la que usaron polinomios ortogonales de Legendre.

Para dos registros en los tiempos, t_i y t_j , la función de covarianza se escribe:

$$\text{Cov}(y_i, y_j) = f(t_i, t_j) = \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} \phi(t_i) \phi(t_j) h_{ij} = \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} \phi(a_i) \phi(a_j) \tau_{ij}.$$

Donde k es el orden de \mathbf{G} y a_i y a_j el tiempo estandarizado con valores en el intervalo entre -1 y 1 y obtenido mediante:

$$a_i = \frac{2(t_i - t_{\text{mínimo}})}{t_{\text{máximo}} - t_{\text{mínimo}}} - 1$$

Matricialmente, la función de covarianza es, $\hat{\mathbf{G}} = \boldsymbol{\phi} \mathbf{K} \boldsymbol{\phi}'$, donde, $\boldsymbol{\phi}$ y \mathbf{K} son matrices de los coeficientes del polinomio de Legendre y de la función de covarianza respectivamente. Sistema que, al expresar $\boldsymbol{\phi} = \mathbf{M} \boldsymbol{\Lambda}$, hace mas simple la obtención del polinomio de Legendre, sobre todo en operaciones con matrices. Los elementos de \mathbf{M} son $m_{ij} = a_i^{(j-1)}$ y $\boldsymbol{\Lambda}$ es la matriz de coeficientes obtenidos en esquema iterativo del polinomio de Legendre. Así, se produce $\hat{\mathbf{G}} = \mathbf{M} \boldsymbol{\Lambda} \mathbf{K} \boldsymbol{\Lambda}' \mathbf{M}'$. En ambos sistemas la matriz de coeficientes es; $\mathbf{K} = \boldsymbol{\phi}^{-1} \hat{\mathbf{G}} \boldsymbol{\phi}^{-T} = \mathbf{M}^{-1} \boldsymbol{\Lambda}^{-1} \mathbf{K} \boldsymbol{\Lambda}^{-1} \mathbf{M}^{-1}$.

Para calcular el polinomio de Legendre, primero se define $\mathbf{P}_0(x) = 1$, y $\mathbf{P}_1(x) = x$. Entonces en general el polinomio de $n+1$ se describe por la ecuación:

$$\mathbf{P}_{n+1} = \frac{1}{n+1} ((2n+1)X\mathbf{P}_n(x) - n\mathbf{P}_{n-1}(x))$$

Los resultados de las ecuaciones anteriores se normalizan con:

$$\phi_n(x) = \sqrt{\frac{2n+1}{2}} \mathbf{P}_n(x)$$

Para realizar los cálculos adecuados con fines de estimar la función de covarianza, suponga se tiene las estimaciones de las (co)varianzas genéticas de producción diaria de leche en los días (t_i) 15, 75, 135, y 195:

$$\hat{\mathbf{G}} = \begin{bmatrix} 1.514 & 2.620 & 2.680 & 2.578 \\ 2.620 & 4.632 & 4.902 & 4.880 \\ 2.680 & 4.902 & 5.474 & 5.735 \\ 2.578 & 4.880 & 5.735 & 6.275 \end{bmatrix}$$

El polinomio que predice las (co)varianzas de la característica en el tiempo de su expresión, puede ser de orden completo o reducido. Es de orden completo cuando el orden de \mathbf{G} y el del polinomio son iguales, y de orden reducido si el orden del polinomio es menor al de la matriz \mathbf{G} . El orden del polinomio usado en la función de covarianza es de gran importancia en características que, como la producción de leche por día, se expresan en un periodo de tiempo grande. Así, dado que usar polinomio de orden completo es prohibitivo, se debe, entonces, buscar de entre los polinomios de orden reducido el de menor orden cuyo ajuste no difiera estadísticamente del obtenido con el polinomio de orden completo.

a. Orden completo

En los datos el valor mínimo y máximo de día de lactancia fueron 15 y 195 respectivamente. Los valores estandarizados de día de lactancia para los días en el ejemplo, son:

t_i	15	75	135	195
a_i	-1	-1/3	1/3	1

El polinomio de Legendre de grado máximo que se puede obtener con la información disponible es el tercero, polinomio cúbico. Así, los coeficientes del polinomio cúbico son:

$$\mathbf{P}_0(x) = 1,$$

$$\phi_0(x) = \sqrt{\frac{2(0)+1}{2}} \mathbf{P}_0(x) = 0.707107,$$

$$\mathbf{P}_1(x) = x,$$

$$\phi_1(x) = \sqrt{\frac{2(1)+1}{2}} \mathbf{P}_1(x) = \sqrt{\frac{3}{2}} x = 1.224745x,$$

$$\mathbf{P}_2(x) = \frac{1}{2} ((2(1) + 1)x \mathbf{P}_1(x) - 2\mathbf{P}_{1-1}(x)) = 0.5(3x(x) - 1(1)) = 0.5(3x^2 - 1) = \frac{3}{2}x^2 - \frac{1}{2},$$

$$\phi_2(x) = \sqrt{\frac{2(2)+1}{2}} \mathbf{P}_2(x) = \sqrt{\frac{5}{2}} (1.5x^2 - 0.5) = 2.371708x^2 - 0.790569,$$

$$\begin{aligned} \mathbf{P}_3(x) &= \frac{1}{3} ((2(2)+1)x\mathbf{P}_2(x) - \mathbf{P}_{2-1}(x)) = \frac{1}{3} (5x(\frac{3}{2}x^2 - \frac{1}{2}) - 2x) = \frac{1}{3} (\frac{15}{2}x^3 - \frac{5}{2}x - 2x) \\ &= \frac{15}{6}x^3 - \frac{9}{6}x, \end{aligned}$$

$$\phi_3(x) = \sqrt{\frac{2(3)+1}{2}} \mathbf{P}_3(x) = \sqrt{\frac{7}{2}} (\frac{15}{6}x^3 - \frac{9}{6}x) = 4.677072x^3 - 2.806243x.$$

Las ecuaciones resultantes de las funciones son:

$$\phi_0(x) = 0.707107 + 0x + 0x^2 + 0x^3$$

$$\phi_1(x) = 0 + 1.224745x + 0x^2 + 0x^3$$

$$\phi_2(x) = -0.790569 + 2.371708x^2 + 0x^3$$

$$\phi_3(x) = 0 - 2.806243x + 0x^2 + 4.677072x^3$$

Los coeficientes para cada una de las $\phi_i(x)$ producen la matriz:

$$\mathbf{\Lambda} = \begin{bmatrix} 0.707107 & 0 & 0 & 0 \\ 0 & 1.224745 & 0 & 0 \\ -0.790565 & 0 & 2.371708 & 0 \\ 0 & -2.806243 & 0 & 4.677072 \end{bmatrix}.$$

Los elementos de \mathbf{M} , $\{m_{ij}\}$, para $i = 1, 2, 3, \text{ y } 4$ y $j = 1, 2, 3, \text{ y } 4$, son:

i	j			
	1	2	3	4
1	$a_1^{1-1} = -1^0 = 1$	$= a_1^{2-1} = -1^1 = -1$	$a_1^{3-1} = -1^2 = 1$	$a_1^{4-1} = -1^3 = -1$
2	$a_2^{1-1} = -(1/3)^0 = 1$	$a_2^{2-1} = (-1/3)^1 = -1/3$	$a_2^{3-1} = (-1/3)^2 = 1/9$	$a_2^{4-1} = (-1/3)^3 = -1/27$
3	$a_3^{1-1} = 1/3^0 = 1$	$a_3^{2-1} = 1/3^1 = 1/3$	$a_3^{3-1} = 1/3^2 = 1/9$	$a_3^{4-1} = (1/3)^3 = 1/27$
4	$a_4^{1-1} = 1^0 = 1$	$a_4^{2-1} = 1^1 = 1$	$a_4^{3-1} = 1^2 = 1$	$a_4^{4-1} = 1^3 = 1$

Por lo tanto:

$$\mathbf{M} = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & -1/3 & 1/9 & -1/27 \\ 1 & 1/3 & 1/9 & 1/27 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

La matriz de coeficientes del polinomio de tercer orden, 4 coeficientes, es:

$$\Phi = \mathbf{M}\Lambda = \begin{bmatrix} 0.7071 & -1.2247 & 0.5811 & -1.8708 \\ 0.7071 & -0.4082 & -0.6382 & 0.7622 \\ 0.7071 & 0.4082 & -0.6382 & -0.7622 \\ 0.7071 & 1.2247 & 0.5811 & 1.8708 \end{bmatrix}$$

La matriz de coeficientes de la función de covarianza:

$$\mathbf{K} = \Phi^{-1} \mathbf{G} \Phi^{-T} = \begin{bmatrix} 8.0045 & 1.1409 & -1.0029 & 0.2028 \\ 1.1409 & 0.4212 & -0.0345 & -0.0007 \\ -1.0029 & -0.0345 & 0.1724 & -0.0381 \\ 0.2028 & -0.0007 & -0.0381 & 0.0084 \end{bmatrix} y$$

la función de covarianza:

$$\mathbf{F} = \Lambda \mathbf{K} \Lambda' = \begin{bmatrix} 5.2312 & 0.5345 & -1.1596 & 0.8116 \\ 0.5345 & 0.7026 & 0.0888 & -0.1141 \\ -1.1596 & 0.0888 & 0.3243 & -0.2446 \\ 0.8116 & -0.1141 & -0.2446 & 0.1838 \end{bmatrix}$$

La predicción de la matriz **G** inicial es:

$$\hat{\mathbf{G}}_3 = \phi \mathbf{K} \phi' = \mathbf{M} \mathbf{F} \mathbf{M}' = \begin{bmatrix} 1.514 & 2.620 & 2.680 & 2.578 \\ 2.620 & 4.632 & 4.902 & 4.880 \\ 2.680 & 4.902 & 5.474 & 5.735 \\ 2.578 & 4.880 & 5.735 & 6.275 \end{bmatrix}$$

Para estimar las (co)varianza entre días dentro del intervalo medido y que no están en la información inicial, 45 y 105 por ejemplo, es necesario obtener los elementos correspondientes en **M**, previa estimación de a_{45} y a_{105} . Estos son:

$$a_{45} = \frac{2(45-15)}{195-15} - 1 = -\frac{2}{3} \text{ y } a_{105} = \frac{2(105-15)}{195-15} - 1 = 0.$$

En tanto los elementos de **M**.

$$\begin{aligned} \mathbf{m}_{45} &= (a_{45}^{1-1} a_{45}^{2-1} a_{45}^{3-1} a_{45}^{4-1}) \\ &= (-2/3^0 (-2/3)^1 (-2/3)^2 (-2/3)^3) \\ &= (1 \ -2/3 \ 4/9 \ -8/27). \end{aligned}$$

$$\begin{aligned} \mathbf{m}_{105} &= (a_{105}^{1-1} a_{105}^{2-1} a_{105}^{3-1} a_{105}^{4-1}) \\ &= (0^0 \ 0^1 \ 0^2 \ 0^3) \\ &= (1 \ 0 \ 0 \ 0). \end{aligned}$$

Las (co)varianzas entre ellos, será más simple si se hace una matriz de dos por cuatro tal que:

$$\mathbf{M}_{45,105} = \begin{bmatrix} \mathbf{m}_{45} \\ \mathbf{m}_{105} \end{bmatrix} = \begin{bmatrix} 1 & -2/3 & 4/9 & -8/27 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Así, se tiene $\mathbf{CV} = \mathbf{M}_{45,105} (\mathbf{F}) \mathbf{M}_{45,105}'$

$$\mathbf{CV} = \begin{bmatrix} \sigma_{45}^2 & \sigma_{45,105} \\ \sigma_{45,105} & \sigma_{105}^2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -2/3 & 4/9 & -8/27 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 5.2312 & 0.5345 & -1.1596 & 0.8116 \\ 0.5345 & 0.7026 & 0.0888 & -0.1141 \\ -1.1596 & 0.0888 & 0.3243 & -0.2446 \\ 0.8116 & -0.1141 & -0.2446 & 0.1838 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & -2/3 \\ 0 & 4/9 \\ 0 & -8/27 \end{bmatrix}$$

$$= \begin{bmatrix} 3.3660 & 4.1190 \\ 4.1190 & 5.2312 \end{bmatrix}$$

b. Orden reducido

Como fue antes mencionado, si el grado de la función de covarianza es menor al rango de \hat{G} , la estimación se conoce como de orden reducido, y es de gran interés cuando el orden de la matriz de covarianzas genéticas es muy grande. Así, se debe encontrar de entre las funciones de covarianza la de menor número de coeficientes y el mejor ajuste posible. La decisión sobre el grado de la función de covarianza adecuada, radica en aquella de menor grado y error no diferente estadísticamente de las de grado mayor.

Si k es el orden de \hat{G} , 3 para el ejemplo, podemos escribir:

$$\mathbf{G} = \mathbf{N}\mathbf{F}_r\mathbf{N}'$$

Donde \mathbf{N} es rectangular con k renglones y $n < k$ columnas y \mathbf{F}_r es la matriz de función de covarianza de menor orden reducido, menor a 3. Entonces, el sistema se escribe, $\mathbf{N}'\mathbf{G}\mathbf{N} = \mathbf{N}'\mathbf{N}\mathbf{F}_r\mathbf{N}'\mathbf{N}$, y la matriz de coeficientes es $\mathbf{F}_r = (\mathbf{N}'\mathbf{N})^{-1}\mathbf{N}'\mathbf{G}\mathbf{N}(\mathbf{N}'\mathbf{N})^{-1}$.

El ajuste de diferentes polinomios se prueba comparando las sumas de cuadrados del error para cada uno. Para ello se expresa la matriz \hat{G} inicial y predicha con los polinomios de cada orden, completo y reducido, en un vector columna, así, $\hat{g}' = (g_{11} \ g_{12} \dots \ g_{1t} \ g_{21} \ g_{22} \dots \ g_{tt})$, y el error es la resta entre las (co)varianzas iniciales y las predichas en cada polinomios. Por ultimo la para $\mathbf{e}'\mathbf{e}$, la suma de cuadrados del error, el estadístico de prueba es:

$$F_C = \frac{(\hat{\mathbf{e}}'\hat{\mathbf{e}}_n - \hat{\mathbf{e}}'\hat{\mathbf{e}}_k) / ((\mathbf{k}(\mathbf{k}+1)/2 - \mathbf{n}(\mathbf{n}+1)/2)}{\mathbf{e}'\mathbf{e} / (\mathbf{k}(\mathbf{k}+1)/2)}$$

Tomando, para el ejemplo, polinomios de orden 2 y 1, (polinomio cuadrático y lineal respectivamente) se tiene:

$$\mathbf{N}_2 = \begin{bmatrix} 1 & -1 & 1 \\ 1 & -1/3 & 1/9 \\ 1 & 1/3 & 1/9 \\ 1 & 1 & 1 \end{bmatrix}, \mathbf{N}_1 = \begin{bmatrix} 1 & -1 \\ 1 & -1/3 \\ 1 & 1/3 \\ 1 & 1 \end{bmatrix}, \mathbf{F}_{r2} = \begin{bmatrix} 5.2312 & 1.2739 & -1.1596 \\ 1.2739 & 0.6473 & -0.1341 \\ -1.1596 & -0.1341 & 0.3243 \end{bmatrix}, \text{ y}$$

$$\mathbf{F}_{r1} = \begin{bmatrix} 4.0428 & 1.1994 \\ 1.1994 & 0.6473 \end{bmatrix}$$

Por último las estimaciones de las (co)varianzas $\hat{\mathbf{G}}_2 = \mathbf{N}_2 \mathbf{F}_{r2} \mathbf{N}_2'$ y $\hat{\mathbf{G}}_1 = \mathbf{N}_1 \mathbf{F}_{r1} \mathbf{N}_1'$:

$$\hat{\mathbf{G}}_2 = \begin{bmatrix} 1.6039 & 2.5555 & 2.8839 & 2.5889 \\ 2.5555 & 4.2101 & 4.9056 & 4.6421 \\ 2.8839 & 4.9056 & 5.8888 & 5.8335 \\ 2.5889 & 4.6421 & 5.8335 & 6.1633 \end{bmatrix} \text{ y } \hat{\mathbf{G}}_1 = \begin{bmatrix} 2.2912 & 2.6593 & 3.0274 & 3.3955 \\ 2.6593 & 3.3151 & 3.9709 & 4.6267 \\ 3.0274 & 3.9709 & 4.9144 & 5.8578 \\ 3.3955 & 4.6267 & 5.8578 & 7.0890 \end{bmatrix}.$$

Los vectores de (co)varianzas iniciales y estimadas con el polinomio cúbico y cuadrático, $\hat{\mathbf{g}}$, $\hat{\mathbf{g}}_2$ y $\hat{\mathbf{g}}_1$ respectivamente:

$$\hat{\mathbf{g}} = \begin{bmatrix} 1.514 \\ 2.620 \\ 2.680 \\ 2.578 \\ 2.620 \\ 4.632 \\ 4.902 \\ 4.880 \\ 2.680 \\ 4.902 \\ 5.474 \\ 5.735 \\ 2.578 \\ 4.880 \\ 5.735 \\ 6.275 \end{bmatrix}, \hat{\mathbf{g}}_2 = \begin{bmatrix} 1.6039 \\ 2.5555 \\ 2.8839 \\ 2.5889 \\ 2.5555 \\ 4.2101 \\ 4.9056 \\ 4.6421 \\ 2.8839 \\ 4.9056 \\ 5.8888 \\ 5.8335 \\ 2.5889 \\ 4.6421 \\ 5.8335 \\ 6.1633 \end{bmatrix}, \text{ y } \hat{\mathbf{g}}_1 = \begin{bmatrix} 2.2912 \\ 2.6593 \\ 3.0274 \\ 3.3955 \\ 2.6593 \\ 3.3151 \\ 3.9709 \\ 4.6267 \\ 3.0274 \\ 3.9709 \\ 4.9144 \\ 5.8578 \\ 3.3955 \\ 4.6267 \\ 5.8578 \\ 7.0890 \end{bmatrix}$$

Los vectores de error son $\hat{\mathbf{e}}_1 = \hat{\mathbf{g}} - \hat{\mathbf{g}}_1$, y $\hat{\mathbf{e}}_2 = \hat{\mathbf{g}} - \hat{\mathbf{g}}_2$ donde el subíndice indica el orden de la matriz del polinomio, en tanto las sumas de cuadrados para los polinomios de grado 1 y 2 ($\hat{\mathbf{e}}_1' \hat{\mathbf{e}}_1$

y $\hat{\epsilon}_2 \hat{\epsilon}_2$), son, 6.7877 y 0.595 respectivamente. El valor de la prueba, para el polinomio de orden 2 (el de menor error) es, tomando $\sigma_{\epsilon_2}^2 = 0.5167/6 = 0.0861$, la prueba entonces:

$$F = \frac{(6.7877 - 0.595)/(6 - 3)}{0.0861} = 23.9748$$

El valor de F tabulada con 3 y 6 grados de libertad para el numerador y denominador respectivamente y $\alpha = 0.05$, es de 8.94. La diferencia es entonces significativa, implicando se obtiene mejor ajuste con el polinomio de cúbico.