

Trabajo completo

Selección de un modelo no lineal mixto de niveles múltiples para modelar el crecimiento de la soja

RECIBIDO: 16/06/2010

ACEPTADO: 15/11/2010

Garcia, M. del C. • Rapelli, C.

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística. Facultad de Ciencias Económicas y Estadística. Universidad Nacional de Rosario. Bvard Oroño 1261. (2000) Rosario, Santa Fe, Argentina.

Teléfono: 0341-4802794-int.152. E-mail: mgarcia@fcecon.unr.edu.ar

RESUMEN: Los datos obtenidos de la experimentación agrícola, frecuentemente, tienen una estructura jerárquica, pues las unidades de análisis están anidadas dentro de múltiples factores de agrupamiento y sobre esas unidades se realizan mediciones repetidas a través del tiempo. Los modelos no lineales mixtos de niveles múltiples son útiles para describir el proceso de crecimiento, pues sus efectos aleatorios corresponden a las diferentes fuentes de variabilidad. El propósito de este trabajo es modelar y seleccionar un modelo de niveles múltiples para el crecimiento de la soja. Se ajusta un modelo mixto con tres niveles anidados de efectos aleatorios y se verifica la bondad del ajuste. Los resultados muestran que el uso de efectos aleatorios en algunos de los parámetros de la función de crecimiento y una función de potencia de los valores ajustados en la variancia

intra-unidad explica la heterogeneidad de los datos.

PALABRAS CLAVE: Estudios longitudinales. Modelos no lineales mixtos de niveles múltiples. Construcción de modelos. Criterios de bondad de ajuste.

SUMMARY: Data from agricultural experiments frequently have a hierarchical or grouped structure, since units of analysis are nested within multiple grouping factors. A multi-level non-linear mixed model is useful to describe the growth process, when data are grouped according to multiple nested factors. The purpose of this paper is to present an introduction to the formulation and selection of multi-level models in the analysis of longitudinal data and illustrate their usefulness for modeling the growth of soy. A mixed model is fit

with three nested levels of random effects using criteria that allows us to directly verify the goodness of fit of the model. The results show that the use of random effects in some of the parameters of the growth function and a power function of the fitted

values in the intra-unit variance explains the heterogeneity of the data. The model also it identifies the factors that lead to a better plant growth.

KEYWORDS: Longitudinal data. Multilevel nonlinear mixed model. Building model process. Goodness-of-fit criteria.

1. Introducción

Los modelos no lineales mixtos de niveles múltiples se utilizan cuando las mediciones repetidas de la variable de interés, registradas a través del tiempo, se realizan sobre unidades agrupadas jerárquicamente. Este agrupamiento produce heterogeneidad en los datos que se puede modelar mediante la introducción de efectos aleatorios, que están asociados a las diferentes fuentes de variabilidad (grupos, unidades dentro de esos grupos) que tienen una estructura jerárquica con varios niveles.

Una etapa de la modelación estadística es la "construcción del modelo" que consiste en caracterizar a los efectos fijos y aleatorios a incluir en el mismo, así como a la obtención de un modelo parco. Durante el proceso iterativo de selección del modelo se utilizan criterios para comprobar la bondad del ajuste que se basan en la función de verosimilitud. Una de las desventajas de estos criterios es que resulta necesaria la estimación de diferentes modelos y dada la complejidad de los mismos, sumado al volumen de información disponible, aumentan la dificultad de implementación. En los últimos años surgieron estadísticas similares al coeficiente de determinación de los modelos lineales como alternativas de los anteriores. En este trabajo se presenta el proceso de construcción de un modelo no lineal de niveles múltiples para explicar la

evolución de la altura de la soja y caracterizar factores que afectan su crecimiento. Durante la etapa de selección del modelo se utilizan algunos criterios propuestos (1, 2) para comprobar la bondad del ajuste, que proveen una medida de la concordancia entre las respuestas observadas y ajustadas. Se utilizan los paquetes estadísticos R (3) y SAS (4, 5).

2. Materiales y métodos

Los datos

Una investigación relacionada con el control químico de malezas en cultivos de soja se llevó a cabo en la Facultad de Ciencias Agrarias de la Universidad Nacional de Rosario. En este estudio se combinó un arreglo espacial, con dos espaciamientos del cultivo, con dosis reducidas de herbicidas y distintas densidades de una maleza (malva). La Malva (*Anoda cristata*) es una maleza de importancia en soja y otros cultivos agrícolas de la Argentina.

Entre los factores agronómicos que alteran el crecimiento de un cultivo en presencia de maleza se encuentra el espaciamiento entre filas del cultivo, que somete a la maleza a un cambio de hábitat, y la densidad de maleza existente. Se planea un ensayo compuesto por tres bloques con un espaciamiento entre filas de 35 cm y otros tres bloques con un espaciamiento de 70 cm. Se asignan aleatoriamente a cinco par-

celas dentro de cada bloque uno de los niveles de densidad de malva, 0, 2, 4, 7 y 12 plantas/m². Se efectúa la medición de la altura de las plantas en cuatro momentos del ciclo del cultivo de soja. Debido a que se está interesado en cómo las densidades de malva afectan el crecimiento en general y no para los espaciamientos específicos utilizados, los datos se tratan como si los dos espaciamientos fuesen aleatoriamente seleccionados de una población de espaciamientos. De la misma forma, para cada uno de los espaciamientos se utilizan tres bloques que se consideran una muestra aleatoria de una población de bloques con un espaciamiento específico. De esta manera, los datos resultantes están agrupados jerárquicamente dentro de niveles de múltiples factores de agrupamiento anidados (espaciamiento, bloques dentro de espaciamiento, parcelas dentro de bloques).

En este trabajo un aspecto importante a considerar es el efecto que produce la densidad de maleza en el crecimiento del cultivo, sembrado a distintos espaciamientos, debido a que la variabilidad de la respuesta puede estar afectada por los distintos niveles de los factores anidados. En consecuencia se propone un modelo estadístico para evaluar el crecimiento de la soja tomando en cuenta los factores que favorecen el crecimiento de la maleza y las diferentes unidades de heterogeneidad. Un modelo no lineal mixto de niveles múltiples se considera apropiado por tener unidades agrupadas jerárquicamente.

Métodos estadísticos

Los modelos no lineales mixtos se pueden extender para representar datos agrupados de acuerdo a múltiples factores

anidados. En este trabajo se presenta la versión para niveles múltiples del modelo de Lindstrom y Bates (6, 7), para tres factores de anidamiento, que se escribe como un modelo en dos etapas. En la primera, la respuesta para el k-ésimo nivel del tercer factor de agrupamiento del j-ésimo nivel del segundo factor dentro del i-ésimo nivel del primero en la ocasión l, Y_{ijkl} , se expresa como

$$Y_{ijkl} = f(X_{ijk}, \beta_{ijk}) + \epsilon_{ijkl}, \quad i=1, \dots, n \quad j=1, \dots, n_i \\ k=1, \dots, n_{ij} \quad l=1, \dots, n_{ijk}, \quad (1.1)$$

siendo, n el número de niveles del primer agrupamiento, n_i el número de niveles del segundo dentro del i-ésimo nivel del primer agrupamiento, n_{ij} el número de niveles del tercero dentro del nivel j del segundo dentro del i-ésimo del primer agrupamiento y n_{ijk} las ocasiones de medida. La función no lineal conocida $f(X_{ijk}, \beta_{ijk})$ relaciona el vector de respuestas con el tiempo y otras posibles covariables intra-unidad (x_{ijk}) y β_{ijk} es un vector específico del individuo que contiene los parámetros de la función no lineal. El error intra-grupo se representa por ϵ_{ijkl} .

La segunda etapa expresa al vector β_{ijk} como la suma de varias componentes, β es la componente fija o componente poblacional común a todos los sujetos, b_i , b_{ij} y b_{ijk} son las componentes específicas a cada nivel de agrupamiento,

$$\beta_{ijk} = A_{ijk} \beta + B_{ijk} b_i + B_{ijk} b_{ij} + B_{ijk} b_{ijk} \quad (1.2)$$

La matriz ($n_{ijk} \times s$) de diseño A_{ijk} incluye las covariables. Las matrices de diseño de los efectos aleatorios B_{ijk} , $B_{ij,k}$ y B_{ijk} , de respectivas dimensiones ($n_{ijk} \times v_1$), ($n_{ijk} \times v_2$) y ($n_{ijk} \times v_3$), dependen de los niveles del primer, segundo y tercer agrupamiento y posi

blemente de los valores de algunas covariables de la observación l -ésima.

Los otros elementos del modelo son,

β : Vector ($s \times 1$) de efectos fijos,

b_i : Vector ($v_1 \times 1$) de efectos aleatorios del primer nivel,

b_{ij} : Vector ($v_2 \times 1$) de efectos aleatorios del segundo nivel,

b_{ijk} : Vector ($v_3 \times 1$) de efectos aleatorios del tercer nivel,

ϵ_{ijk} : Vector ($n_{ijk} \times 1$) de errores intra-grupos.

$$\begin{aligned} \epsilon_{ijk} &\sim N_n(0, \Sigma_{ijk} = \sigma^2 I) & b_i &\sim N_v(0, \Psi_1) \\ b_{ij} &\sim N_v(0, \Psi_2) & b_{ijk} &\sim N_v(0, \Psi_3) \end{aligned}$$

El modelo se puede estimar mediante el método de máxima verosimilitud. Debido a que condicional a los efectos aleatorios $Y_{ijk} \sim N(f(\cdot), \sigma^2 I)$, la verosimilitud para Y_{ijk} se puede obtener integrando una densidad normal con respecto a la distribución de los efectos aleatorios. Pero, maximizar la función de verosimilitud resultante es complicado por la presencia de una integral multidimensional en esta función. Para una estructura simple de los efectos aleatorios, la integración se puede realizar por cuadratura Gaussiana (8), o alguna otra técnica numérica sin demasiada dificultad. Existen varias alternativas para la estimación de la verosimilitud completa de los modelos no lineales mixtos (NLMM) que están basadas en la expansión de Taylor de primer orden de la función f del modelo. La principal distinción entre esos métodos, denominados de linealización, reside en el punto alrededor del cual se hace la expansión. Ésta se puede realizar alrededor de cero, el valor esperado del vector de efectos aleatorios (9), o alrededor de alguna estimación del vector de efectos

aleatorios, usualmente llamado el mejor predictor lineal insesgado (EBLUP) (6).

3. Construcción del modelo

Una etapa importante dentro de la modelación de datos con varios niveles de efectos aleatorios, es evaluar si las estructuras para los efectos fijos y aleatorios resultan adecuadas para describir los datos observados, es decir, comprobar si se han incorporado términos innecesarios en el modelo.

El término "construcción de un modelo" se refiere al proceso de caracterizar los efectos aleatorios a incluir, así como a la obtención de un modelo adecuado pero parco. El procedimiento se basa en la comparación de modelos con diferentes estructuras para la componente aleatoria y permite determinar cuáles de los parámetros del mismo serían efectos mixtos y cuáles fijos, como también modelar la dependencia de los parámetros a covariables específicas de la unidad.

La construcción de un modelo se realiza en varios pasos.

Primer paso: Determinación de los parámetros del modelo que contendrán efectos aleatorios y posteriormente la elección del tipo de estructura para la matriz de covariancias de los efectos aleatorios (Ψ).

Para ello se comienza postulando un modelo cuyos parámetros incluyen efectos aleatorios en todos los niveles de anidamiento. Luego se ajustan modelos obtenidos al eliminar algunos de los efectos aleatorios de los parámetros del primer modelo. Para verificar el ajuste de los mismos se usa el test del cociente de verosimilitud o los criterios de información de Akaike (10) o Bayesiano de Schwarz (11).

La estadística del test del cociente de verosimilitud, $U=I_0 - I_a$, se compara con una distribución de probabilidad adecuada para rechazar o no la hipótesis planteada, siendo I_0 y I_a las respectivas log-verosimilitudes de los modelos postulados bajo las hipótesis nula y alternativa.

El criterio de información de Akaike (AIC) penaliza el logaritmo de la verosimilitud ($\log\hat{L}$) restando el número de parámetros de cada modelo (s), $AIC = -2\log\hat{L} + 2$. El criterio de información bayesiano de Schwarz (BIC) es similar al anterior excepto en el valor de la penalización, $BIC = -2\log\hat{L} + s\log N$, siendo N el número unidades de análisis. En ambos casos se prefieren valores pequeños de ambos criterios.

Una vez elegido el tipo de efecto, se plantean diferentes estructuras para la matriz de covariancias de los mismos, utilizando los criterios de información para la elección entre modelos alternativos. De esta forma se obtiene un modelo que incluye términos con efectos aleatorios significativos y la matriz de covariancias seleccionada, mientras que, durante este paso del proceso, la covariancia intra sujeto se mantiene fija e igual a una de independencia.

Segundo paso: Elección de la matriz de covariancia de los errores. Para ello se plantean diferentes estructuras utilizando, nuevamente, los criterios de información para la elección de la más adecuada.

Debido a que los datos longitudinales presentan, además de la correlación, variancias heterogéneas se debe evaluar mediante el uso de gráficos la presencia de variancias desiguales. La mayoría de los enfoques gráficos se basan en los residuos estandarizados (12). Dado un conjunto de residuos se pueden aplicar los diagnósticos

de residuos usuales para regresión lineal estándar. En un modelo correctamente especificado los gráficos no mostrarán ningún patrón sistemático, con residuos fluctuando aleatoriamente alrededor del cero y variarán dentro de un rango constante.

Tercer paso: En los dos pasos anteriores, el modelo especificado incluye todas las covariables de interés. Como el objetivo que se pretende con el uso de estos modelos no sólo es caracterizar el cambio de la respuesta en el tiempo sino también los factores que influncian el cambio, se necesitan realizar pruebas de hipótesis. La significación de los parámetros se comprueba mediante los tests del cociente de verosimilitud y de Wald. Aunque este último es muy fácil de implementar puede estar afectado por la no linealidad de los parámetros. Por lo tanto en este caso se prefiere el test del cociente de verosimilitud.

Este último método requiere ajustes repetidos de modelos, por lo que algunos autores (1) propusieron estadísticas de bondad de ajuste para la evaluación de la estructura media, que se aplican directamente al modelo que se está utilizando, agilizando así la construcción del mismo.

Las dos medidas alternativas son:

1.- R^2 , es una modificación del coeficiente de determinación de los modelos lineales. La estadística se define como,

$$R^2(V) = 1 - \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} d_{ijk}^2(V)}{\sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} d_{ijk0}^2(V)} = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} W_{ijk} R^2_{ijk}(V)}{\sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} W_{ijk}}$$

dónde,

$$R^2_{ijk}(V) = 1 - \frac{d^2_{ijk}(V)}{d^2_{ijk0}(V)},$$

$$d^2_{ijk} = (Y_{ijk} - Y_{ijk})' V^{-1} (Y_{ijk} - Y_{ijk}) \quad y$$

$$d^2_{ijk0}(V) = (Y_{ijk} - Y_{ijk0})' V^{-1} (Y_{ijk} - Y_{ijk0}),$$

siendo, $V = \sum_{ijk} W_{ijk} = d_{ijk0}(V)$, Y_{ijk} el vector de las respuestas de la unidad ijk e Y_{ijk} y $d_{ijk}^2(V)$ e Y_{ijk0} y $d_{ijk0}^2(V_{ijk'})$ los vectores de los valores predichos y las distancias de Mahalanobis bajo los modelos a evaluar y de comparación (modelo nulo), respectivamente

Para implementar esta definición de R^2 se necesita especificar un modelo nulo apropiado también como los valores de V . Generalmente se elige una estructura simple para el modelo nulo. Para V se puede

optar elegir la matriz de covariancias del modelo nulo, para que la medida de bondad de ajuste se pueda usar para comparar los distintos modelos a evaluar.

2. Coeficiente de concordancia (r_c).

La expresión de este coeficiente, que puede ser usado como una medida alternativa a la anterior, es,

siendo, $J_{n_{ijk}}$ es un vector de unos, $\bar{y} =$

$$r_c = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ijk}} (\mathbf{Y}_{ijk} - \hat{\mathbf{Y}}_{ijk})^T (\mathbf{Y}_{ijk} - \hat{\mathbf{Y}}_{ijk})}{\sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ijk}} (\mathbf{Y}_{ijk} - \bar{y} \mathbf{J}_{n_{ijk}})^T (\mathbf{Y}_{ijk} - \bar{y} \mathbf{J}_{n_{ijk}}) + \sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ijk}} (\hat{\mathbf{Y}}_{ijk} - \bar{y} \mathbf{J}_{n_{ijk}})^T (\hat{\mathbf{Y}}_{ijk} - \bar{y} \mathbf{J}_{n_{ijk}}) + N(\bar{y} - \bar{\bar{y}})^2}$$

Este coeficiente da una medida de la concordancia entre los valores observados y predichos y se interpreta como

Ajuste perfecto valores cercanos a 1

Falta de ajuste valores menores o iguales a 0.

Debido a que ambos coeficientes se incrementan a medida que se sobreparametrizan los modelos sus valores se pueden ajustar por el número de parámetros de la siguiente manera,

$$R^2 = 1 - k(1 - R^2),$$

$$r_{c,a} = 1 - k(1 - r_c).$$

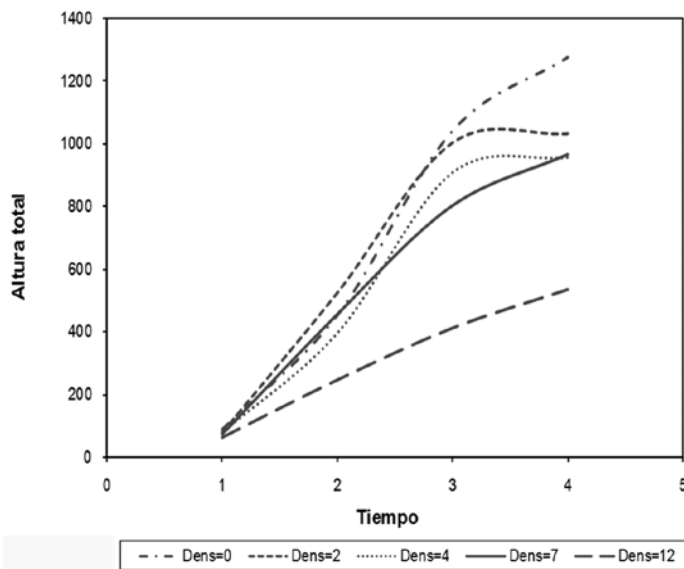
siendo, $k = N/(N-s)$ y s el número de parámetros de efectos fijos.

La principal ventaja de usar estas medidas como bondad de ajuste es que no se requiere la especificación de la función de verosimilitud y resultan muy útiles para datos continuos.

4.- Aplicación

Se evalúa la altura de la soja, en cuatro momentos de su crecimiento, sembrada en bloques con distinto espaciamiento (35 y 70 cm) entre las filas y en combinación con distintas densidades de una maleza (malva), pero sin el uso de herbicida. La evolución del crecimiento promedio de la soja para distintas densidades de siembra se representan en el siguiente gráfico:

Gráfico 4.1: Altura promedio de la soja para distintas densidades de malva



Los datos describen un patrón de crecimiento no lineal de tipo sigmoideo. Después de realizar análisis exploratorios y de residuos se elige la función no lineal de Gompertz para representar los datos.

En una primera instancia se considera un modelo que incluye efectos aleatorios para las tres fuentes de variabilidad.

El ajuste de las curvas se realiza con el software estadístico R, mientras que las estadísticas de bondad de ajuste se calculan utilizando el procedimiento IML de SAS (4, 5).

El modelo utilizado para Y_{ijkl} , la altura de la soja de la k -ésima parcela (densidad) en el bloque j -ésimo del espaciamiento i en

la ocasión de medida l -ésima con $i=1,2$, $j=1,\dots,3$, $k=1,\dots,5$ y $l=1,\dots,4$, tiene la expresión:

$$Y_{ijkl} = \beta_{0ijk} \exp(-\beta_{1ijk} \exp(-\beta_{2ijk} t)) + \varepsilon_{ijk}t$$

$$\begin{aligned} \beta_{0ijk} &= \beta_0 + \beta_{01} D_2 + \beta_{02} D_4 + \beta_{03} D_7 + \beta_{04} D_{12} + \\ & b_{0i} + b_{0j} + b_{0ijk} \\ \beta_{1ijk} &= \beta_1 + \beta_{11} D_2 + \beta_{12} D_4 + \beta_{13} D_7 + \beta_{14} D_{12} + \\ & b_{1i} + b_{1j} + b_{1ijk} \\ \beta_{2ijk} &= \beta_2 + \beta_{21} D_2 + \beta_{22} D_4 + \beta_{23} D_7 + \beta_{24} D_{12} + \\ & b_{2i} + b_{2j} + b_{2ijk} \end{aligned} \tag{4.1}$$

Se supone que los tres parámetros del modelo dependen linealmente de las variables ficticias utilizadas para identificar los 5 niveles del factor tratamiento,

$$D_2 = \begin{cases} 1 & \text{si la densidad es 2} \\ 0 & \text{en otro caso} \end{cases}$$

$$D_4 = \begin{cases} 1 & \text{si la densidad es 4} \\ 0 & \text{en otro caso} \end{cases}$$

$$D_7 = \begin{cases} 1 & \text{si la densidad es 7} \\ 0 & \text{en otro caso} \end{cases}$$

$$D_{12} = \begin{cases} 1 & \text{si la densidad es 12} \\ 0 & \text{en otro caso} \end{cases}$$

y que,

$$\mathbf{b}_i = \begin{pmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{pmatrix} \stackrel{iid}{\sim} N(\mathbf{0}, \Psi_1), \quad \mathbf{b}_{ij} = \begin{pmatrix} b_{0ij} \\ b_{1ij} \\ b_{2ij} \end{pmatrix} \stackrel{iid}{\sim} N(\mathbf{0}, \Psi_2),$$

$$\mathbf{b}_{ijk} = \begin{pmatrix} b_{0ijk} \\ b_{1ijk} \\ b_{2ijk} \end{pmatrix} \stackrel{iid}{\sim} N(\mathbf{0}, \Psi_3) \text{ y } \boldsymbol{\varepsilon}_{ijk} = \begin{pmatrix} \varepsilon_{ijk1} \\ \vdots \\ \varepsilon_{ijk4} \end{pmatrix} \stackrel{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_{ijk})$$

Además, se supone que Ψ_1 , Ψ_2 y Ψ_3 son matrices de covariancias con estructura diagonal y $\sum_{ijk} = \sigma^2 I$. Los vectores aleatorios indican b_i efecto espaciamento, b_{ij} efecto bloque dentro de espaciamento y b_{ijk} efecto densidad dentro de bloque.

Los parámetros del modelo $\beta = (\beta_{0ijk}, \beta_{1ijk}, \beta_{2ijk})'$ representan:

β_{0ijk} la asíntota de la curva,

β_{1ijk} el inicio de crecimiento poblacional de la curva,

β_{2ijk} la tasa de crecimiento poblacional.

El modelo (4.1), que será llamado modelo 1, supone que los errores son independientes y con variancia homogénea: $\sum_{ijk} = \sigma^2 I_4$, y considera tres efectos aleatorios para los tres parámetros. A continuación se presenta el proceso de construcción del modelo.

Especificación de los efectos aleatorios

Para los efectos aleatorios se impone una matriz de covariancias sin estructura. Se plantean diferentes modelos, denominando el modelo 1 con efectos aleatorios en todos los parámetros, como completo y los otros (modelos 2, 3, 4, 5 y 6) reducidos.

Se usa la prueba de la razón de verosimilitud (TRV) o algún criterio de información para determinar cuáles parámetros tienen asociados una componente aleatoria.

Tabla 4.1. Comparación de modelos con diferentes términos de efectos aleatorios

Modelo	b_{0i}	b_{1i}	b_{2i}	b_{0ij}	b_{1ij}	b_{2ij}	b_{0ijk}	b_{1ijk}	b_{2ijk}	AIC	-2L	TRV	p-value
Modelo 1	*	*	*	*	*	*	*	*	*	1418.4	1368.4		
Modelo 2	*	*	*	*			*			1410.4	1368.4	0.0016	0.9980
Modelo 3	*	*	*	*						1409.2	1369.2	0.7897	0.9777
Modelo 4	*		*	*			*			1408.4	1368.4	0.0016	0.9998
Modelo 5	*		*				*			1409.3	1371.3	2.9223	0.8185
Modelo 6			*	*			*			1406.4	1368.4	0.0016	0.9998

Los tres parámetros del modelo 1 se consideran mixtos, con efectos aleatorios en los niveles espaciamento, bloque y densidad. Como los otros modelos están anidados dentro del primero se puede usar la prueba del cociente de verosimilitud o los criterios de información, para comprobar si la reduc-

ción de los parámetros produce algún cambio significativo en el modelo. La elección recae sobre el modelo 6 que posee efecto espaciamento aleatorio en la tasa de crecimiento, mientras que la asíntota posee efectos aleatorios bloque dentro de espaciamento y densidad dentro de bloque.

Especificación de la estructura de variancias y covariancias de los errores intra individuos

El gráfico 4.2 muestra los residuos transformados del modelo 6. En el mismo se observa que las variancias no son homogéneas.

La tabla 4.2 presenta los criterios que permiten evaluar diferentes estructuras de covariancias para el modelo 6. Los modelos 7 a 11 modifican sólo la matriz de covariancias intra unidad del modelo 6.

Gráfico 4.2. Residuos versus valores predichos para el Modelo 6

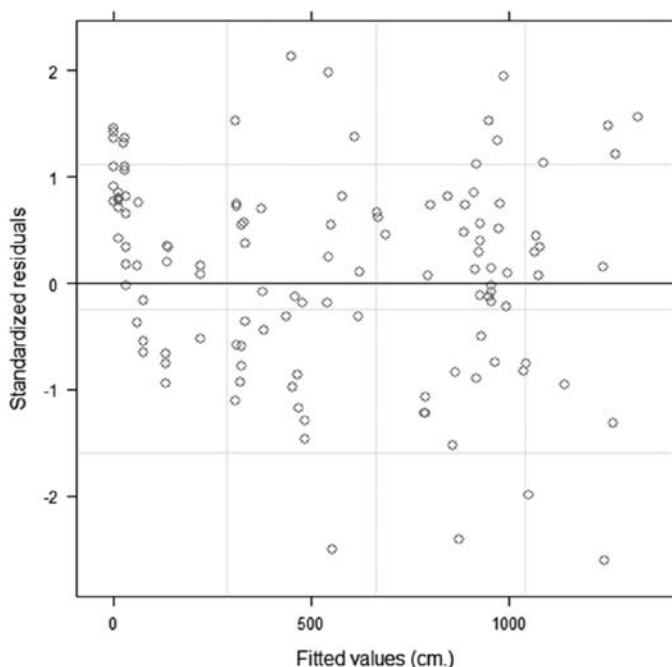


Tabla 4.2. Comparación de modelos con diferentes estructuras de covariancias para los errores intra-individuo

Modelo	Σ	AIC	BIC
Modelo 6	Independencia con variancias homogéneas	1406.4	1459.3
Modelo 7	Autorregresiva de orden 1	1400.6	1456.4
Modelo 8	Función de potencia del tiempo	1381.6	1437.3
Modelo 9	Función de potencia de los valores ajustados	1365.8	1421.5
Modelo 10	Función exponencial del tiempo	1397.9	1453.6

El uso de una función de potencia de los valores ajustados para la variancia de los errores intra unidad (modelo 9) mejora el desempeño del modelo 6.

Comparación entre efectos de las densidades

Durante las dos etapas anteriores los efectos de la densidad fueron incluidos en todos los parámetros como efectos fijos. Debido a que uno de los objetivos es determinar las densidades de malva que afec-

tan el crecimiento de la soja en el tiempo se prueba la significación de los efectos.

La tabla siguiente presenta el modelo completo (modelo 9) y varios modelos reducidos que surgen de llevar las hipótesis planteadas al modelo.

Tabla 4.3. Comparación de modelos con efectos fijos

Modelo	Hipótesis	AIC	-2L	TRV	p-value	R ²	r _c
Modelo 9		1365.8	1325.8			0.9447	0.9761
Modelo 11	$\beta_{01} = \beta_{02} = \beta_{03} = \beta_{04} = 0$	1380.8	1348.8	23.0526	<.0001	0.9211	0.9656
Modelo 11.1	$\beta_{01} = \beta_{02} = \beta_{03} = 0$	1395.3	1361.3	35.4906	<.0001	0.9443	0.9755
Modelo 11.2	$\beta_{01} = \beta_{02} = \beta_{03} = 0$	1378.3	1344.3	18.5028	0.0003	0.9243	0.9670
Modelo 11.3	$\beta_{01} = \beta_{02} = \beta_{03} = 0$	1378.6	1344.6	18.8379	0.0003	0.9245	0.9672
Modelo 11.4	$\beta_{02} = \beta_{03} = \beta_{04} = 0$	1381.7	1347.7	21.9491	0.0001	0.9226	0.9663
Modelo 12	$\beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = 0$	1394.4	1362.4	36.6392	<.0001	0.9440	0.9765
Modelo 12.1	$\beta_{11} = \beta_{12} = \beta_{13} = 0$	1363.3	1329.3	3.4845	0.3228	0.9452	0.9763
Modelo 12.2	$\beta_{11} = \beta_{12} = \beta_{13} = 0$	1396.4	1362.4	36.5751	<.0001	0.9434	0.9762
Modelo 12.3	$\beta_{11} = \beta_{12} = \beta_{13} = 0$	1396.4	1362.4	36.6579	<.0001	0.9366	0.9737
Modelo 12.4	$\beta_{12} = \beta_{13} = \beta_{14} = 0$	1391.2	1357.2	31.3733	<.0001	0.9395	0.9749
Modelo 13	$\beta_{21} = \beta_{22} = \beta_{23} = \beta_{24} = 0$	1390.1	1358.1	32.3358	<.0001	0.5337	0.8124
Modelo 13.1	$\beta_{21} = \beta_{22} = \beta_{23} = 0$	1366.2	1332.2	6.4030	0.0936	0.9415	0.9747
Modelo 13.2	$\beta_{21} = \beta_{22} = \beta_{23} = 0$	1392.1	1358.1	32.3351	<.0001	0.9504	0.9784
Modelo 13.3	$\beta_{21} = \beta_{22} = \beta_{23} = 0$	1392.1	1358.0	32.2649	<.0001	0.9341	0.9727
Modelo 13.4	$\beta_{12} = \beta_{13} = \beta_{14} = 0$	1385.7	1351.7	25.9192	<.0001	0.9408	0.9754
Modelo 14	$\beta_{11} = \beta_{12} = \beta_{13} = \beta_{21} = \beta_{22} = \beta_{23} = 0$	1361.5	1333.5	7.6789	0.2626	0.9411	0.9745

El modelo 11 permite probar si el crecimiento límite de la soja es el mismo para todas las densidades de malva. Como esta hipótesis se rechaza ($p < 0.0001$) se plantean los modelos 11.1, 11.2, 11.3 y 11.4 llegando a la misma conclusión. De la misma manera se procede con las otras hipótesis, obteniéndose que el inicio y la tasa de crecimiento de las curvas no presentan diferencias significativas para las densidades 2, 4 y 7.

Las conclusiones anteriores, que se obtienen utilizando el test de la razón de verosimilitud, coinciden si se consideran los

criterios de información. Se observa que los mismos presentan el menor valor para el modelo 14, que resulta el modelo elegido.

La tabla 4.3 permite ilustrar, además, el uso de las estadísticas de bondad de ajuste presentadas. Existe una disminución de los valores de los coeficientes R^2_a y $r_{c,a}$ cuando se eliminan efectos significativos del modelo (modelos 11, 12 y 13).

Los coeficientes R^2_a y $r_{c,a}$ de los modelos resultantes de eliminar los efectos no significativos (modelo 12.1, 13.1 y 14) no difieren de los obtenidos para el modelo 9.

Del proceso de construcción del modelo de los efectos aleatorios son, resulta el siguiente modelo estimado,

$$\hat{Y}_{ijkl} = \hat{\beta}_{0ijk} \exp(-\hat{\beta}_{1ijk} \exp(\hat{\beta}_{2ijk} t)),$$

$$\begin{aligned} \hat{\psi}_1 &= 0.0874 \\ \hat{\psi}_2 &= 0.1522, \\ \hat{\psi}_3 &= 0.2570 \end{aligned}$$

donde,

$$\begin{aligned} \hat{\beta}_{0ijk} &= 1352.21 - 86.01D_2 - 223.98D_4 - 354.09D_7 \\ &+ 636.69D_{12} + \hat{b}_{0ij} + \hat{b}_{0ijk} \\ \hat{\beta}_{1ijk} &= 8.59 - 4.08D_{12} \\ \hat{\beta}_{2ijk} &= 1.11 - 0.78D_{12} + \hat{b}_{2i} \end{aligned}$$

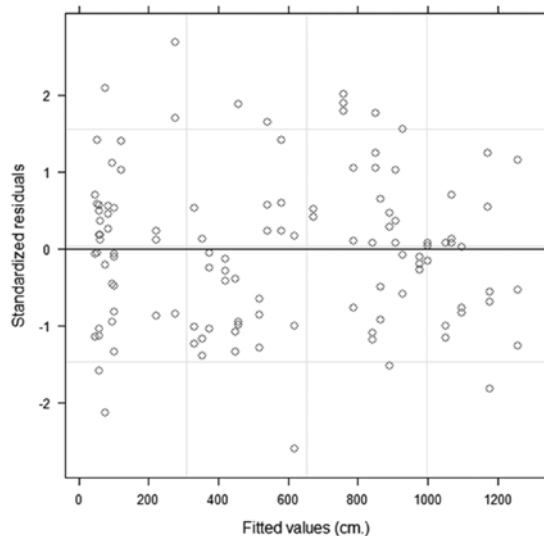
La estimación de b_{0i} provee el valor del crecimiento máximo para cada espaciamiento, mientras que la de b_{0ijk} lo suministra para cada densidad dentro de cada bloque. La estimación \hat{b}_{2i} representa las diferencias en la tasa de crecimiento de la soja producidas por el espaciamiento utilizado. Las estimaciones de las componentes de variancia

y las estimaciones de los parámetros de la matriz de covariancias intra individuo,

$$\begin{aligned} \hat{\Sigma}_{ijk} &= \text{Vâr}(\epsilon_{ijk}) = \hat{\sigma}^2 (\hat{Y}_{ijk})^{2\delta} \\ \hat{\sigma}^2 &= 0.7738 \\ \delta &= 0.7379 \end{aligned}$$

En el gráfico 4.3 se observa que los residuos están aleatoriamente distribuidos alrededor de cero y presentan variancias constantes. Se puede concluir que el modelo presenta un ajuste adecuado del patrón de cambio de la respuesta en el tiempo.

Gráfico 4.3. Residuos versus valores predichos para el Modelo 14



5. Discusión

Un modelo no lineal mixto de niveles múltiples es un modelo en el que los efectos aleatorios incluidos en él están asociados a

las unidades de heterogeneidad (espaciamientos, bloques y parcelas) que tienen una estructura jerárquica con niveles múltiples.

En este trabajo se utiliza ese enfoque para modelar la evolución de la altura de la soja, en presencia de una maleza. Se evalúan si diferentes densidades de malva afectan el crecimiento del cultivo, sembrada en bloques con distintos espaciamientos entre las filas. La inclusión de efectos aleatorios específicos en el modelo permite ajustar curvas con diferentes formas para las parcelas, bloques y espaciamiento.

El uso comparativo de criterios tradicionales y estadísticas que permiten comprobar la bondad del ajuste en forma directa conducen a la selección de los mismos factores que afectan el cambio de la respuesta en el tiempo.

Del ajuste de las curvas y su comparación se puede concluir que:

- El uso de efectos aleatorios permite distintos comportamientos de las curvas individuales para cada nivel de agrupamiento:
 - La tasa de crecimiento de las curvas es diferente para cada nivel de espaciamiento.
 - El límite de crecimiento es diferente tanto para las parcelas como para los bloques dentro de cada espaciamiento.
- Una densidad alta (12 plantas/m²) de maleza influencia el inicio y la velocidad del crecimiento de la soja.
- La soja alcanza una altura mayor en presencia de baja densidad de malva. A medida que la densidad aumenta la soja logra un menor crecimiento.

Agradecimientos

Agradecemos al Dr. Eduardo Puricelli que generosamente aportó los datos para poder aplicar la metodología presentada en este trabajo.

Bibliografía

1. Vonesh, E. F., Chinchilli, V. M. and Pu, K., 1996. Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics* **52**: 572-587.
2. Vonesh, E. F.; Chinchilli, V. M., 1997. "Linear and Nonlinear Models for the Repeated Measurements". Marcel Dekker.(NY, USA).
3. R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
4. Littell, R.C.; Milliken, G.A.; Stroup, W.W.; Wolfinger, R.D., 1996. "SAS® System for Mixed Models". SAS Institute Inc. (Cary, NC, USA)
5. SAS Institute Inc. 2004. SAS/STAT Software: Versión 9.1. Cary, NC: SAS Institute Inc.
6. Lindstrom, M.J.; Bates, D.M., 1990. Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics*. **46**: 673-687.
7. Pinheiro, J.C. and Bates, D.M., 1995. Model building for nonlinear mixed-effects models. *Technical Report*, 91, Department of Biostatistics, University of Wisconsin-Madison.
8. Davidian, M. and Gallant, A. R. 1992. Smooth nonparametric maximum likelihood estimation for population pharmacokinetic, with application to quinidine. *J. of Pharmacokinetics and Biopharmaceutics*. **20**: 529-556.
9. Sheiner, L. B. y Beal, S. L., 1980. Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: Routine clinical pharmacokinetic data. *J. of Pharmacokinetics and Biopharmaceutics*. **8**: 553-571.
10. Akaike, H., 1972. A new look at the statistical model identification. *IEEE Trans. on Automatic Control*. **19**: 716-723.
11. Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics*. **6**: 461-464.
12. Garcia, M. del C. y Rapelli, C., 2008. Uso de criterios de bondad de ajuste para seleccionar un modelo no lineal. *Revista FABICIB*. **12**: 161-172.