

Modelos mixtos en el análisis de ensayos comparativos de rendimiento multiambientales de una especie perenne*

Santiago Ostengo**, María I. Cuenya** y Mónica Balzarini***

RESUMEN

Los ensayos multiambientales de mejoramiento genético de la caña de azúcar (*Saccharum* spp.) son conducidos durante diferentes edades de corte, por tratarse esta de una especie perenne. Las correlaciones temporales entre observaciones registradas en una misma parcela a través de los años, así como también las correlaciones espaciales entre parcelas experimentales vecinas y la condición de heterocedasticidad entre varianzas residuales de cada ambiente, pueden contemplarse en el contexto de los modelos lineales mixtos (MLM). El objetivo de este trabajo es comparar el desempeño de genotipos de caña de azúcar en un ensayo multiambiental bajo el marco teórico de los MLM, que contemplan posibles correlaciones temporales y espaciales entre datos de rendimientos, así como también la posible condición de heterocedasticidad. Se trabajó con ensayos comparativos de rendimiento multiambientales de caña de azúcar durante diferentes edades de corte, analizando el desempeño de 20 genotipos. Los MLM ajustados contemplaron, en la estructura de (co)varianzas de los errores, correlaciones (temporales y/o espaciales) y heterocedasticidades para toneladas de caña por hectárea (TCH) y rendimiento fabril al inicio de cosecha (RFI %). Para TCH, el mejor modelo fue el autorregresivo de orden 1 tridimensional con correlación temporal y espacial, heterocedástico a través de localidades. Para RFI % el modelo autorregresivo de orden 1 heterocedástico, que contempla solamente correlaciones temporales, fue el mejor. Ambos produjeron estimaciones más precisas (menor error estándar) del rendimiento esperado, proporcionando también cambios en el desempeño relativo de los genotipos que se comparan.

Palabras clave: ensayos multiambientales, correlación espacial, correlación temporal, *Saccharum* spp.

ABSTRACT

Mixed models in the analysis of multienvironment comparative yield trials of a perennial crop

Multienvironment trials in sugarcane (*Saccharum* spp.) breeding are conducted through different crop ages, since sugarcane is a perennial crop. The temporal correlation among data recorded for the same plot across different years, as well as the spatial correlations among data from neighboring plots and the condition of heteroscedasticity among residual variances for each environment, can be approached in the context of mixed linear models (MLM). This work aims to compare the performance of sugarcane genotypes in a multienvironment trial under the MLM framework, which contemplates that there may be temporal and spatial correlations among yield data and that the heteroscedasticity condition may apply. Twenty sugarcane genotypes were evaluated in a multienvironmental trial through different crop ages. MLM that take into account correlations (temporal and/or spatial) and heterocedasticity in the structure of (co) variances of errors were adjusted for tons of cane per hectare (TCH) and sucrose content at the beginning of harvest (SC %). For TCH, the best model was the three-dimensional first order autoregressive model, with temporal and spatial correlations and heteroscedasticity across locations. In the case of SC %, the heterocedastic first order autoregressive model, which includes only the temporal correlation, was the best. Both models provided more precise estimates (lower standard error) of the expected yield and changes in the relative performance of the compared genotypes.

Key words: multienvironment trials, spatial correlation, temporal correlation, *Saccharum* spp.

*Trabajo presentado en la XXV International Biometric Conference. Florianopolis (Brasil), diciembre de 2010.

**Sección Caña de Azúcar, EEAOC. santiagostengo@eeaoc.org.ar

***Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba - Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET).

INTRODUCCIÓN

La caña de azúcar (*Saccharum* spp.) es una especie perenne que se cosecha por al menos tres años de manera consecutiva sobre cada terreno. Por ello, los programas de mejoramiento genético de este cultivo conducen sus ensayos en distintos ambientes (multiambientales) durante las diferentes edades de corte. En estos ensayos, los datos se recolectan durante varios años consecutivos en la misma parcela de cada genotipo, lo que presupone la presencia de correlaciones temporales entre esas observaciones. Además, pueden existir correlaciones espaciales entre parcelas vecinas dentro de cada ensayo. Al ser ensayos multiambientales, las varianzas residuales heterogéneas entre ambientes también pueden estar presentes.

El análisis de datos correlacionados, así como también la condición de heterocedasticidad, puede ser abordado, en el contexto de los modelos lineales mixtos (MLM), mediante la modelización de la estructura de varianzas y covarianzas de los términos del error (Balzarini *et al.*, 2001; Schabenberger and Pierce, 2002). La alta flexibilidad de estos modelos permite evitar los supuestos clásicos de independencia entre observaciones y varianzas residuales constantes, lo cual no era posible con los modelos tradicionalmente usados. La modelización correcta de la estructura de varianzas y covarianzas de las observaciones permite una mejor estimación de los errores estándares (Keselman *et al.*, 1998) usados en las comparaciones de medias, esperándose un incremento en la eficiencia del experimento.

Para contemplar la correlación temporal entre observaciones, registradas en distintas edades de corte, existen diversos modelos utilizados para definir la estructura de la matriz de varianza y covarianza de los errores (Núñez-Antón y Zimmerman, 2001). De estos modelos, los más utilizados son los modelos con estructura autorregresiva de orden 1 (AR1) y de simetría compuesta (CS), así como también sus versiones análogas, que contemplan varianzas residuales heterogéneas entre edades (AR1H y SCH). Además, en situaciones donde la no estacionariedad de las varianzas y de las correlaciones está presente, es de suponer que un modelo con una matriz de varianzas y covarianzas no estructurada resulte el más conveniente.

Para modelar la correlación espacial entre observaciones provenientes de parcelas vecinas, el modelo más recomendado es el de tipo potencia anisotrópica o autorregresivo de orden 1 (AR1XAR1) (Gilmour *et al.*, 1997), el cual supone que la dependencia espacial entre observaciones provenientes de dos parcelas disminuye exponencialmente, a medida que la distancia entre ellas aumenta, dependiendo también de la orientación en la cual estas distancias se calculan. Una de las ventajas de los modelos lineales mixtos es que permiten contemplar,

dentro del mismo modelo, la posible condición de heterocedasticidad de varianzas entre localidades (Casanoves *et al.*, 2005).

Este trabajo compara el desempeño de genotipos de caña de azúcar en un ensayo multiambiental, según el marco teórico de los MLM, que contemplan posibles correlaciones temporales y espaciales entre datos de rendimientos, así como también la posible condición de heterocedasticidad entre varianzas residuales obtenidas dentro de cada ambiente.

MATERIALES Y MÉTODOS

Se trabajó con la serie 2006 de Ensayos Regionales Comparativos de Variedades del Programa de Mejoramiento Genético de Caña de Azúcar de la Estación Experimental Agroindustrial Obispo Colombres (Tucumán, R. Argentina), conducidos en seis localidades representativas de la zona plantada con caña de azúcar en Tucumán. En cada parcela se evaluaron tres edades de corte: planta, soca 1 y soca 2. Todos los ensayos fueron plantados en septiembre de 2006 y cosechados en ese mismo mes en los años subsiguientes (2007, 2008 y 2009). El diseño empleado en los ensayos fue el de bloques completamente aleatorizados con tres repeticiones. La unidad experimental estuvo representada por una parcela de tres surcos de 10 metros de longitud. Se compararon 20 genotipos, cuatro de los cuales son variedades comerciales usadas como testigos. Las variables analizadas fueron: rendimiento cultural expresado en toneladas de caña por hectárea (TCH) y rendimiento fabril al inicio de cosecha (RFI %).

Modelos ajustados

Con el objeto de determinar la estructura más conveniente para cada situación y variable en este conjunto de datos, se compararon por separado modelos con diferentes estructuras de correlación espacial y temporal. Para contemplar la correlación temporal, la estructura autorregresiva de orden 1 mostró ser la más adecuada para ambas variables (TCH y RFI %), mientras que para considerar la correlación espacial, la estructura AR1XAR1 fue la mejor solo a nivel de TCH; es decir que a nivel de RFI %, existe frecuente independencia entre observaciones provenientes de parcelas vecinas (Ostengo, 2010). A partir de ello, para ambas variables se ajustó un modelo que contemplaba ambas correlaciones (temporal y espacial), con un modelo autorregresivo de orden 1 en tres dimensiones (tridimensional) para la estructura de varianzas y covarianzas residuales. En este modelo, las dimensiones dadas por las filas y columnas que posicionan a la parcela están asociadas a la correlación espacial anisotrópica, y la dimensión edad de corte, que indexa el dato de acuerdo al año de cosecha, está asociada a la correlación temporal. El modelo tridimensional fue comparado con un modelo nulo, en el que se considera homogeneidad de varianzas e

independencia temporal -entre datos de distintas edades- y espacial -entre parcelas de una localidad-, además de compararlo con un modelo nulo heterocedástico igual al anterior, pero con varianzas residuales heterogéneas entre localidades. Así mismo, esos modelos fueron comparados con un modelo que contempla solamente la correlación espacial, a partir de un modelo autorregresivo de orden 1 anisotrópico, y con un modelo que contempla solamente la correlación temporal, empleando un modelo autorregresivo de orden 1 heterogéneo. A excepción del modelo nulo, en todos los modelos se contempló la posible presencia de varianzas residuales heterogéneas entre localidades. La ecuación base del modelo mixto, propuesto para la evaluación simultánea de las variedades a través de localidades y edades, es:

$$y_{ijk} = \mu + G_i + L_j + E_h + B(L)_{k(j)} + GL_{(ij)} + GE_{(ih)} + LE_{(jh)} + GLE_{(ijh)} + \varepsilon_{ijk}$$

donde: y_{ijk} es el rendimiento del genotipo i , en la localidad j , edad h y bloque k ; μ es la media general; G_i es el efecto del genotipo i ; L_j es el efecto de la localidad j ; E_h es el efecto de la edad h ; $B(L)_{k(j)}$ es el efecto del bloque k dentro de la localidad j ; $GL_{(ij)}$ es el efecto de la interacción del genotipo i con la localidad j ; $GE_{(ih)}$ es el efecto de la interacción del genotipo i con la edad h ; $LE_{(jh)}$ es el efecto de la interacción de la localidad j con la edad h ; $GLE_{(ijh)}$ es el efecto de la interacción del genotipo i con la localidad j y la edad h y ε_{ijk} es el término del error asociado a la observación y_{ijk} . Los efectos G , E y GE fueron considerados como fijos, mientras que los términos correspondientes a los efectos

$B(L)$, L , GL , LE y GLE fueron considerados como aleatorios.

Todos los modelos se ajustaron usando PROC MIXED, SAS, Versión 9.1 (SAS Institute, 2003)¹. La comparación de los ajustes alcanzados con los distintos modelos se realizó utilizando el criterio de información de Akaike (AIC). Modelos con valores de AIC menores se consideraron superiores en términos de ajuste. También se utilizó para la evaluación de los modelos la prueba de cociente de verosimilitud.

RESULTADOS Y DISCUSIÓN

Los resultados de la comparación de los modelos (Tabla 1) sugieren que aquel que contempla las correlaciones temporales y espaciales (modelo 3) es el más apropiado para TCH y que este representa un mejor ajuste respecto al modelo tradicionalmente usado. Sin embargo, para RFI % el modelo que contempla solamente la correlación temporal (modelo 2) resulta suficiente para explicar la covariación subyacente entre las observaciones. Los mejores modelos para cada variable proporcionaron menores errores estándar con respecto al modelo nulo, cuando se compararon variedades a través de edades, variedades para una determinada edad y edades para una variedad en particular (Tabla 2). La disminución del error estándar no solo significa una mayor eficiencia en el experimento, sino que también produjo un incremento en el número de comparaciones con diferencias estadísticamente significativas entre pares de genotipos. Para TCH, con el modelo 0, se detectaron diferencias signi-

Tabla 1. Criterios de Akaike (AIC) y valores de (-2 Log) de verosimilitud (-2L) de modelos multiambientales ajustados para las variables toneladas de caña por hectárea (TCH) y rendimiento fabril al inicio de cosecha (RFI %), en Ensayos Regionales Comparativos de Variedades de caña de azúcar.

Modelo*	TCH		RFI %	
	AIC	-2 L	AIC	-2 L
Modelo 0	8065,1	8053,1	2746,9	2736,9
Modelo 0 (H)	8036,1	8014,1	2738,4	2718,4
Modelo 1	7930,0	7884,0	2736,0	2692,0
Modelo 2	7942,0	7884,0	<u>2722,7</u>	2668,7*
Modelo 3	<u>7905,4</u>	7847,4**	2728,7	2680,7

*Modelos mixtos multiambientales a través de localidades y edades de corte para las variables TCH y RFI %. Modelo 0: considera observaciones espaciales y temporales independientes y varianzas residuales homogéneas entre localidades. Modelo 0 (H): considera observaciones espaciales y temporales independientes y varianzas residuales heterogéneas entre localidades. Modelo 1: considera solamente correlación espacial con un modelo autorregresivo de orden 1 anisotrópico para la estructura de varianzas y covarianzas residuales y varianzas residuales heterogéneas entre localidades. Modelo 2: considera solamente correlación temporal con un modelo autorregresivo de orden 1 heterogéneo para la estructura de varianzas y covarianzas residuales y varianzas residuales heterogéneas entre localidades. Modelo 3: considera la correlación temporal y espacial con un modelo autorregresivo de orden 1 en tres dimensiones (edad de corte, columna y fila) para la estructura de varianzas y covarianzas residuales y varianzas residuales heterogéneas entre localidades.

Valores subrayados indican el mínimo AIC (mejor modelo).

**Mejor ajuste de acuerdo a la prueba de cociente de verosimilitud, a un nivel de significación del 5%, usando como referente cada uno de los modelos restantes.

¹Las rutinas que se utilizaron en SAS Versión 9.1, para ajustar los modelos, figuran en el Anexo 1.

Tabla 2. Valores del error estándar promedio de la diferencia entre dos medias de: (1) variedades a través de edades, (2) variedades para cada edad y (3) edades para cada variedad, obtenidos de dos modelos mixtos para las variables TCH y RFI %.

Modelo †	TCH			RFI %		
	1	2	3	1	2	3
Modelo 0	3,62	4,74	7,06	0,194	0,324	0,584
Modelo 2				0,182	0,301	0,570
Modelo 3	3,27	4,32	6,75			

† Modelos mixtos a través de localidades y edades de corte para las variables TCH y RFI %. Modelo 0: considera varianzas residuales homogéneas entre localidades y observaciones espaciales y temporales independientes. Modelo 2: considera solamente correlación temporal con un modelo autorregresivo de orden 1 heterogéneo para la estructura de varianzas y covarianzas residuales y varianzas residuales heterogéneas entre localidades. Modelo 3: considera correlación temporal y espacial con un modelo autorregresivo de orden 1 en tres dimensiones (edad de corte, columna y fila) para la estructura de varianzas y covarianzas residuales y varianzas residuales heterogéneas entre localidades.

ficativas en el 46% de las comparaciones, mientras que con el modelo seleccionado, este valor se incrementó a un 51%. En el caso de RFI %, el aumento fue del 58% al 63%. Además, los modelos seleccionados produjeron cambios en cuanto al ordenamiento de los genotipos, según valores medios predichos a través de localidades y edades de corte para ambas variables. Para TCH, seis genotipos cambiaron su orden con el modelo de mejor ajuste, mientras que para RFI % esta situación se dio en ocho genotipos.

CONCLUSIONES

El marco teórico de los modelos lineales mixtos proporcionó la posibilidad de un mejor ajuste de los datos con respecto al que se hubiese obtenido bajo los supuestos clásicos de homogeneidad de varianzas a través de ensayos e independencia temporal y espacial de las observaciones, proveyendo un menor error estándar para la comparaciones de medias y, por ende, una mayor eficiencia en el experimento. El cambio de modelo permitió detectar diferencias estadísticamente significativas entre genotipos, no evidenciadas por el modelo clásico, así como también producir alteraciones en la posición relativa de jerarquía de sus desempeños, lo que condujo a cambios en la interpretación de los datos analizados. El trabajar con un modelo que se ajusta a los datos significativamente mejor que otro(s) implica otorgar una mayor confiabilidad a las conclusiones del desempeño relativo de las variedades que se comparan.

BIBLIOGRAFÍA CITADA

Balzarini, M.; S. B. Milligan and M. S. Kang. 2001. Best linear unbiased prediction: a mixed model approach in

multi environment trials. En: Kang, M. S. (ed.), Crop improvement challenges in the twenty-first century. Food Products Press, Binghamton, NY, USA, pp. 102-113.

Casanoves, F.; R. Macchiavelli and M. Balzarini. 2005. Error variation in multienvironment peanut trials: within-trial spatial correlation and between-trial heterogeneity. Crop Sci. 45: 1927-1933.

Gilmour, A. R.; B. R. Cullis and A. P. Verbyla. 1997. Accounting for natural and extraneous variation in the analysis of field experiments. J. Agr. Biol. Envir. S. 2: 269-273.

Keselman, H.; J. Algina; R. Kowalchuck and R. Wolfinger. 1998. A Comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. Communication in Statistics: simulation and computation 27: 591-604.

Núñez-Antón, V. y D. L. Zimmerman. 2001. Modelización de datos longitudinales con estructuras de covarianza no estacionarias: modelos de coeficientes aleatorios frente a modelos alternativos. Quaderns d'Estadística i Investigació Operativa (Qüestió) 25 (2): 225-262.

Ostengo, S. 2010. Análisis de ensayos regionales comparativos de variedades de caña de azúcar vía nuevos procedimientos biométricos. Tesis Master of Science inédita. Instituto Agronómico Mediterráneo de Zaragoza, Universitat de Lleida, Zaragoza, España.

SAS Institute. 2003. SAS/STAT release 9.1. SAS Inst., Cary, NC, USA.

Schabenberger, O. and F. Pierce. 2002. Contemporary statistical models for the plant and soil sciences. CRC Press, Boca Raton, Florida, USA.

ANEXO 1

Modelo 0

```
proc mixed data=base06;  
class variedad localidad repet edad;  
model y=variedad edad variedad*edad;  
random repet(localidad) localidad localidad*variedad localidad*edad localidad*variedad*edad;  
lsmeans variedad variedad*edad/pdiff;  
run;
```

Modelo 0 (H)

```
proc mixed data=base06;  
class variedad localidad repet edad;  
model y=variedad edad variedad*edad;  
random repet(localidad) localidad localidad*variedad localidad*edad localidad*variedad*edad;  
repeated/ group=localidad;  
lsmeans variedad variedad*edad/pdiff;  
run;
```

Modelo 1

```
proc mixed data=base06;  
class variedad localidad repet edad;  
model y=variedad edad variedad*edad;  
random repet(localidad) localidad localidad*variedad localidad*edad localidad*variedad*edad;  
repeated/subject=localidad type=sp(powa)(col fila) group=localidad;  
lsmeans variedad variedad*edad/pdiff;  
run;
```

Modelo 2

```
proc mixed data=base06;  
class variedad localidad repet edad parcela;  
model y=variedad edad variedad*edad;  
random repet(localidad) localidad localidad*variedad localidad*edad localidad*variedad*edad;  
repeated/sub=parcela type=arh(1) group=localidad;  
lsmeans variedad variedad*edad/pdiff;  
run;
```

Modelo 3

```
proc mixed data=base06;  
class variedad localidad repet edad parcela;  
model y=variedad edad variedad*edad;  
random repet(localidad) localidad localidad*variedad localidad*edad localidad*variedad*edad;  
repeated/subject=localidad type=sp(powa)(col fila ed)group=localidad;  
lsmeans variedad variedad*edad/pdiff;  
run;
```