

Uma nova abordagem na avaliação da interacção genótipo×ambiente em espécies lenhosas de propagação vegetativa: o caso de clones de videira

Elsa Gonçalves

Secção de Matemática/DCEB e LEAF, Instituto Superior de Agronomia, Universidade de Lisboa, *elsagoncalves@isa.ulisboa.pt*

Antero Martins

LEAF, Instituto Superior de Agronomia, Universidade de Lisboa, *anteromart@isa.ulisboa.pt*

Palavras-chave: Dependência nos erros aleatórios; Medidas repetidas; Modelo autoregressivo de primeira ordem; Modelo de simetria composta; Modelos mistos.

Resumo: Neste trabalho estuda-se a interacção genótipo×ambiente ($G \times E$) numa espécie perene. Neste contexto, fazem-se avaliações num dado local na mesma unidade experimental ao longo de anos consecutivos. Ou seja, os erros aleatórios associados a duas quaisquer observações na mesma unidade experimental não são independentes. Ajustam-se, aos dados de rendimento, modelos mistos com matrizes de covariâncias do vector dos erros aleatórios distintas, como a matriz de simetria composta (CS) e autorregressiva de primeira ordem (AR1), de modo a caracterizar esse fenómeno. Verifica-se que os modelos com matrizes CS e AR1 revelam um melhor ajustamento face ao modelo que admite erros aleatórios independentes, garantindo assim o estudo mais preciso da interacção $G \times E$.

1 Introdução

A avaliação da interacção genótipo×ambiente ($G \times E$) é um objetivo incontornável de qualquer programa de melhoramento de plan-

tas para obtenção de variedades geneticamente homogéneas (todos os indivíduos geneticamente iguais). Para compreender o comportamento de um genótipo em ambientes distintos é essencial que a avaliação das características alvo seja feita no maior número de ambientes possível. Em melhoramento de espécies anuais e perenes herbáceas a avaliação da interacção é uma prática corrente [6, 9, 2], dada a reduzida área e duração dos ensaios. Sobre este tema, as principais técnicas de estudo deste fenómeno são abordadas em Gonçalves e Martins [4]. Relativamente a espécies lenhosas de propagação vegetativa, em Portugal tem sido feito um esforço coerente para o estudo da interacção $G \times E$ em clones de videira. As várias técnicas de interpretação do fenómeno estudadas têm sido aplicadas a dados de rendimento (kg de uva/planta) destacando-se: representação gráfica da ordenação dos clones quanto a diferentes características nos diversos ambientes, cálculo do coeficiente de variação do rendimento de um genótipo nos distintos ambientes, análise de regressão dos valores do rendimento de um genótipo sobre os índices ambientais [5], ajustamento de modelos mistos multivariados e construção de *biplots* [4, 3]. No entanto, tratando-se de uma planta perene (ciclo de vida plurianual), outras abordagens poderão ser ainda exploradas.

Neste tipo de espécie, a instalação de ensaios multi-locais é um processo demorado e de custos elevados, pelo que, na prática, o número de ensaios instalados para o efeito raramente é o mais desejável. De facto, o tempo de vida útil de uma vinha comercial é da ordem dos 30 anos e, por ser uma espécie arbustiva, a área ocupada pelo ensaio tem significado em termos de encargos de gestão cultural. Frequentemente instalam-se ensaios em 2-4 locais e procura-se compensar esse reduzido número com avaliações em mais anos em cada local. A forma mais simples de tratar o problema passa pelo ajustamento de modelos mistos que admitam que os erros aleatórios associados a observações feitas em anos diferentes na mesma unidade experimental (ou seja, sobre as mesmas plantas) são variáveis aleatórias independentes e identicamente distribuídas. No entanto, o que se tem na prática no mesmo local são avaliações sucessivas nos mesmos

indivíduos de cada genótipo ao longo de vários anos. Ou seja, os erros aleatórios associados a duas quaisquer observações na mesma unidade experimental não são independentes. Contudo, para algumas características poderá pensar-se que tal terá pouco significado prático. Por exemplo, uma das características mais importantes alvo do estudo da interacção $G \times E$, o rendimento, é uma característica que é avaliada de ano a ano. Os factores que a influenciam são de tal ordem numerosos que, comparativamente a outras causas de variação, a correlação que existe entre observações feitas na mesma unidade experimental tende a ser negligenciada. No entanto, é necessário compreender se tal se verifica e em que medida afecta a avaliação da interacção $G \times E$. Este trabalho tem precisamente como objectivo desenvolver uma abordagem que responda a este problema.

2 Modelos

Neste trabalho a metodologia proposta, também aplicável a outras espécies lenhosas de propagação vegetativa, representa uma nova abordagem para o estudo da interacção $G \times E$ em clones de videira. Trata-se de uma análise que permite obter os valores médios dos genótipos para uma determinada característica a nível global dos ambientes, juntamente com os desvios da interacção $G \times E$ para cada ambiente e que toma em conta que os erros aleatórios associados a observações feitas em anos diferentes na mesma unidade experimental não são independentes.

Como já referido, nesta espécie em um mesmo local uma determinada característica é avaliada durante vários anos. A noção de ambiente está, portanto, hierarquizada: por um lado, o local distinto (que abrange as condições edafo-climáticas, porta-enxerto, etc.) e, dentro do local, os diferentes anos (não necessariamente os mesmos anos em cada local). Poderia ser construído um modelo que incluísse os efeitos do local e do ano subordinado ao local. No entanto, para simplificação e com o objectivo de centrar o estudo no principal efeito

global do ambiente, nesta abordagem será considerado no modelo o efeito do ambiente visto como a combinação local/ano.

Admitamos dados de rendimento (kg/planta) provenientes de ensaios com delineamento experimental em blocos completos casualizados. Matricialmente, o modelo linear misto aplicável a este tipo de estudo pode ser genericamente descrito como

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (1)$$

em que:

\mathbf{Y} é o vector $n \times 1$ das observações (valores fenotípicos, ou observados, do rendimento), ordenado por local, ambiente (combinação local/ano) e unidade experimental dentro de cada ambiente;

\mathbf{X} é a matriz de delineamento $n \times p$ dos efeitos fixos (matriz cujas colunas são variáveis indicatrizes que identificam as observações de cada nível de cada factor de efeitos fixos);

$\boldsymbol{\beta}$ é o vector $p \times 1$ de efeitos fixos (média populacional, efeitos dos ambientes);

\mathbf{Z} é a matriz de delineamento $n \times q$ dos efeitos aleatórios (matriz cujas colunas são variáveis indicatrizes que identificam as observações de cada nível de cada factor de efeitos aleatórios);

\mathbf{u} é o vector $q \times 1$ de efeitos aleatórios que contém os efeitos genotípicos, os efeitos dos blocos por ambiente e os efeitos da interacção genótipo \times ambiente ($q = \sum_{i=1}^r q_i$, sendo q_i o número de níveis do factor de efeitos aleatórios i e r o número de factores de efeitos aleatórios em estudo);

\mathbf{e} é o vector $n \times 1$ de erros aleatórios.

Os vectores \mathbf{u} e \mathbf{e} admitem-se independentes, com distribuição normal multivariada de vector de valores médios nulo e matrizes de covariâncias \mathbf{G} e \mathbf{R} , respectivamente, isto é,

$$\text{Cov}[\mathbf{u}, \mathbf{e}] = \mathbf{0}, \quad \mathbf{u} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{G}), \quad \mathbf{e} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}).$$

A distribuição de \mathbf{Y} admite-se assim normal multivariada, com vector de valores médios $\mathbf{X}\boldsymbol{\beta}$ e matriz de covariâncias $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$,

$$\mathbf{Y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}).$$

No contexto biológico em análise, estudam-se genótipos sem relações de parentesco e o delineamento experimental é específico de cada ensaio. Assim, relativamente ao vector \mathbf{u} , sendo \mathbf{u}_i o vector dos efeitos aleatórios do factor i , admite-se:

$$\text{var}[\mathbf{u}_i] = \mathbf{G}_i = \sigma_{u_i}^2 \mathbf{I}_{q_i}, \text{ para } i = 1, \dots, r, \text{ e } \text{Cov}[\mathbf{u}_i, \mathbf{u}_{i'}] = \mathbf{0}, \text{ para } \forall i \neq i'.$$

Consequentemente, a matriz de covariâncias do vector \mathbf{u} é definida como $\mathbf{G} = \oplus_{i=1}^r \mathbf{G}_i$, em que \oplus representa a soma directa de matrizes. Especificando de acordo com efeitos aleatórios descritos no modelo (1), \mathbf{G} é a soma directa das matrizes $\mathbf{G}_g = \sigma_g^2 \mathbf{I}_{q_1}$, $\mathbf{G}_{b(A)} = \sigma_{b(A)}^2 \mathbf{I}_{q_2}$, $\mathbf{G}_{ge} = \sigma_{ge}^2 \mathbf{I}_{q_3}$, sendo σ_g^2 a variância genotípica, $\sigma_{b(A)}^2$ a variância dos blocos subordinados ao ambiente e σ_{ge}^2 a variância da interacção $G \times E$.

Relativamente ao vector \mathbf{e} , a forma mais simples de abordar o problema é admitir que os elementos deste vector são variáveis aleatórias independentes e identicamente distribuídas, isto é, $\mathbf{R} = \sigma_e^2 \mathbf{I}_n$. Admite-se, portanto, homogeneidade de variâncias e que os erros aleatórios associados a observações feitas em anos diferentes na mesma unidade experimental (em ambientes diferentes no mesmo local) são variáveis aleatórias independentes e identicamente distribuídas (adiante designado modelo *IND*). No entanto, o que se tem na prática no mesmo local são avaliações sucessivas na mesma unidade experimental ao longo de vários anos.

Seja, então, o vector \mathbf{e} definido como

$$\mathbf{e} = \left[\mathbf{e}_{1(n_1 \times 1)}^T \quad \mathbf{e}_{2(n_2 \times 1)}^T \quad \cdots \quad \mathbf{e}_{l(n_l \times 1)}^T \right]^T,$$

representando cada subvector, \mathbf{e}_j , com $j = 1, \dots, l$, o vector dos erros aleatórios no local j , e n_j o número de observações no respectivo local. Seja $\text{Var}[\mathbf{e}_j] = \mathbf{R}_j$, representando, portanto, \mathbf{R}_j a respectiva matriz de covariâncias. Admite-se $\text{Cov}[\mathbf{e}_j, \mathbf{e}_{j'}] = \mathbf{0}$, para $\forall j \neq j'$,

consequentemente, a matriz de covariâncias do vector \mathbf{e} é definida como

$$\mathbf{R} = \bigoplus_{j=1}^r \mathbf{R}_j.$$

Trata-se agora de definir a estrutura da matriz de covariâncias do vector dos erros aleatórios no local j (matriz \mathbf{R}_j). Admite-se que os erros aleatórios associados a observações de unidades experimentais diferentes são independentes e que os associados a observações na mesma unidade experimental ao longo dos anos não são independentes. No local j com p unidades experimentais, a matriz \mathbf{R}_j é definida como

$$\mathbf{R}_j = \mathbf{I}_p \otimes \tilde{\Sigma}_{e_j},$$

sendo \mathbf{I}_p a matriz identidade de ordem p e \otimes o produto de *Kronecker* de matrizes.

A estrutura mais complexa para $\tilde{\Sigma}_{e_j}$ seria uma matriz não estruturada, que admitiria variâncias distintas para cada ano no local j e diferentes covariâncias para todos os pares de anos avaliados nesse local. Por local, para a anos de avaliação, estimar-se-iam $a + a(a-1)/2$ parâmetros, o que resultaria num modelo excessivamente parametrizado e um exagero em termos do que realmente se quer ter em conta [12]. Não será, portanto, a opção a seguir. As estruturas estudadas serão as que fazem sentido neste contexto biológico: uma estrutura que traduza uma contribuição comum a todas as observações feitas na mesma unidade experimental, ou que considere que a correlação entre observações da mesma unidade experimental diminui à medida que a distância de separação entre os anos aumenta. Neste sentido, para a matriz $\tilde{\Sigma}_{e_j}$ foram definidos os dois tipos de estrutura seguidamente descritos.

(1) Uma matriz de simetria composta (adiante designado modelo CS, do Inglês *Compound Symmetry*), que tem como elementos diagonais $\sigma_{e_j}^2$ (variância dos erros aleatórios do local j) e elementos não diagonais $\sigma_{e_j}^2 \rho$ (sendo ρ a correlação entre pares de observações na mesma unidade experimental ao longo dos anos). Isto é, admite-se homogeneidade de variâncias no local j e que todos os pares de observações na mesma unidade experimental têm a mesma correlação.

(2) Uma matriz autorregressiva de primeira ordem (adiante designado modelo AR1), que tem como elementos diagonais $\sigma_{e_j}^2$ e elementos não diagonais $\sigma_{e_j}^2 \rho^{|k-k'|}$, sendo $|k - k'|$ o intervalo de separação entre os anos k e k' . Isto é, admite-se homogeneidade de variâncias no local j e que a correlação entre observações da mesma unidade experimental diminui à medida que a distância entre anos aumenta. Faz sentido quando se avaliam anos consecutivos.

O método de estimação dos parâmetros incluídos nas matrizes \mathbf{G} e \mathbf{R} foi o método de máxima verosimilhança restrita, REML [8], actualmente o mais recomendado e utilizado no contexto dos modelos lineares mistos para estimar parâmetros covariância em grandes conjuntos de dados com estrutura complexa [7]. A comparação e selecção de modelos com estruturas de covariância distintas (modelos IND, CS e AR1) foi baseada no critério de informação de Akaike (AIC) [10], definido como [12]:

$$AIC_i = -2lr_i + 2npar_i,$$

em que lr_i é a log-verosimilhança restrita do modelo i e $npar_i$ o respectivo número de parâmetros covariância do modelo. Modelos encaixados foram também formalmente comparados com base em testes de razão de verosimilhanças.

Por último, é útil reforçar que esta metodologia é aplicada na última fase da metodologia de selecção da videira [5], quando todos os genótipos (clones) presentes já foram seleccionados quanto a várias características de interesse, mas ainda não quanto à sua estabilidade ambiental. Por isso, o objectivo principal nesta fase é avaliar a variabilidade da interacção $G \times E$ e, posteriormente, seleccionar com base na menor sensibilidade a essa interacção. Neste sentido, o interesse deste tipo de estudo está essencialmente nesta componente de variância (σ_{ge}^2). A inferência relativa a este parâmetro foi baseada num teste de razão de verosimilhanças restritas [11] ($H_0 : \sigma_{ge}^2 = 0$ vs $H_0 : \sigma_{ge}^2 > 0$) e foi efectuada para todos os modelos em estudo,

de modo a compreender em que medida a estrutura de covariância dos erros aleatórios afecta a avaliação da interacção $G \times E$.

3 Uma aplicação

A aplicação utiliza dados de rendimento (kg de uva/planta) de vários ensaios de selecção de variedades antigas de videira, instalados em 2 e 3 locais por variedade, segundo um delineamento experimental em blocos completos casualizados (RCB, do Inglês *Randomized Complete Block*), com 3 a 9 repetições, e avaliados durante vários anos (Tabela 1). O delineamento experimental RCB é um dos mais simples para controlar variabilidade indesejável presente num ensaio (tipo de solo, exposição, declive, etc.). Neste sentido, nos modelos de análise de dados de ensaios com este tipo de delineamento, os efeitos dos blocos por ambiente são sempre incluídos, de modo a respeitar o processo de casualização associado a este tipo de delineamento.

Procedeu-se de seguida ao ajustamento dos modelos lineares mistos propostos na secção anterior, considerando-se: o factor ambiente como um factor de efeitos fixos; os efeitos genotípicos, os efeitos dos blocos por ambiente e os efeitos da interacção genótipo \times ambiente como aleatórios; as três estruturas de covariância (IND, CS and AR1). Para tal, recorreu-se ao *Software R, package ASReml – R* [1] (método de estimação REML, usando o algoritmo de informação média).

Os resultados obtidos com o ajustamento dos vários modelos aos dados de rendimento das 7 variedades antigas constam da Tabela 2. Em todos os casos estudados verificou-se que os modelos com estruturas de covariância CS e AR1 revelaram melhor ajustamento face ao modelo que admitiu erros aleatórios independentes entre observações da mesma unidade experimental (IND), tendo este último modelo revelado sempre maior valor de *AIC*. Também se chega a esta conclusão comparando formalmente os modelos IND e CS e os modelos IND e AR1 através de um teste de razão de verosimilhanças restritas. Em qualquer dos casos, pelos valores da log-verosimilhança

| |
|---|
| <p>Alvarinho (AI), 40 clones avaliados em 20 ambientes: <i>l1</i>, Monção (RCB, 3 blocos)/1990 a 1992; <i>l2</i>, Monção-Pias (RCB, 9 blocos)/1995 a 2000; <i>l3</i>, Monção-Ceivães (RCB, 9 blocos)/1994 a 2004.</p> |
| <p>Antão Vaz (AN), 40 clones avaliados em 14 ambientes: <i>l1</i>, Évora (RCB, 5 blocos)/1988;1989; 1990; <i>l2</i>, Palmela (RCB, 8 blocos)/1993 a 1998; <i>l3</i>, Vidigueira (RCB, 8 blocos)/1998 a 2002.</p> |
| <p>Aragonez (RZ), 40 clones avaliados em 13 ambientes: <i>l1</i>, Estremoz (RCB, 8 blocos)/1992 a 1999; <i>l2</i>, Tabuaço (RCB, 8 blocos)/1994 a 1998.</p> |
| <p>Fernão Pires (FP), 40 clones avaliados em 11 ambientes: <i>l1</i>, Alpiarça (RCB, 8 blocos)/1993 a 1997; <i>l2</i>, Anadia (RCB, 8 blocos)/1999,2000; <i>l3</i>, Caldas da Rainha (RCB, 8 blocos)/1995 a 1998.</p> |
| <p>Negra Mole (NM), 40 clones avaliados em 7 ambientes: <i>l1</i>, Lagoa (RCB, 5 blocos)/1989, 1990; <i>l2</i>, Loulé (RCB, 8 blocos)/1994 a 1998.</p> |
| <p>Rabo de Ovelha (OV), 33 clones avaliados em 8 ambientes: <i>l1</i>, Redondo (RCB, 8 blocos)/1996 a 2000; <i>l2</i>, Reguengos de Monsaraz (RCB, 5 blocos)/1990 a 1992.</p> |
| <p>Síria (CR), 40 clones avaliados em 10 ambientes: <i>l1</i>, Estremoz (RCB, 8 blocos)/1992 a 1999; <i>l2</i>, Pinhel (RCB, 5 blocos)/1988 e 1989.</p> |

Tabela 1: Descrição dos dados usados na aplicação: variedade antiga, número de clones (genótipos) e número de ambientes por variedade, com indicação do local, delineamento experimental e anos de avaliação em cada local.

| Variedade antiga | Modelo | lr | $npar$ | AIC |
|---------------------|------------|---------|--------|---------|
| Alvarinho (AI) | <i>IND</i> | -5509.5 | 4 | 11027.0 |
| | <i>CS</i> | -5202.5 | 9 | 10423.0 |
| | <i>AR1</i> | -5224.9 | 9 | 10467.8 |
| Antao Vaz (AN) | <i>IND</i> | -3506.1 | 4 | 7020.2 |
| | <i>CS</i> | -3155.9 | 9 | 6329.9 |
| | <i>AR1</i> | -3167.7 | 9 | 6353.3 |
| Aragonez (RZ) | <i>IND</i> | -1820.8 | 4 | 3649.6 |
| | <i>CS</i> | -1573.4 | 7 | 3160.9 |
| | <i>AR1</i> | -1711.0 | 7 | 3435.9 |
| Fernaõ Pires (FP) | <i>IND</i> | -1548.6 | 4 | 3105.3 |
| | <i>CS</i> | -1480.9 | 9 | 2979.9 |
| | <i>AR1</i> | -1493.3 | 9 | 3004.5 |
| Negra Mole (NM) | <i>IND</i> | -738.6 | 4 | 1485.2 |
| | <i>CS</i> | -535.1 | 7 | 1084.3 |
| | <i>AR1</i> | -531.2 | 7 | 1076.4 |
| Rabo de Ovelha (OV) | <i>IND</i> | -1173.6 | 4 | 2355.2 |
| | <i>CS</i> | -1156.5 | 7 | 2326.9 |
| | <i>AR1</i> | -1164.4 | 7 | 2342.8 |
| Síria (CR) | <i>IND</i> | -1303.9 | 4 | 2615.9 |
| | <i>CS</i> | -1250.1 | 7 | 2514.2 |
| | <i>AR1</i> | -1285.7 | 7 | 2585.5 |

Tabela 2: Log-verossimilhança restrita (lr), número de parâmetros covariância ($npar$) e critério de informação de Akaike (AIC) obtidos com o ajustamento dos modelos com matrizes de covariâncias do vector dos erros aleatórios diagonal (IND), de simetria composta (CS) e autorregressiva de primeira ordem (AR1).

restrita constantes da Tabela 2, facilmente se percebe que o valor calculado da estatística do teste de razão de verosimilhanças restritas conduzirá à rejeição do modelo IND para qualquer nível de significância usual. Entre os modelos que admitem dependência entre observações realizadas na mesma unidade experimental, o modelo CS revelou sempre um melhor ajustamento, com excepção de um único caso de estudo, a casta Negra Mole (Tabela 2). Os resultados obtidos com o modelo AR1, que não evidenciam uma vantagem geral deste modelo face ao CS, podem dever-se ao facto do número de anos avaliados em cada local ser baixo ou moderado.

As estimativas dos parâmetros covariância obtidas com o ajustamento dos vários modelos encontram-se na Tabela 3. Torna-se clara a heterogeneidade de variâncias do erro entre locais. Relativamente às correlações entre observações realizadas na mesma unidade experimental, estas revelaram-se fracas a moderadas, variando entre castas e, relativamente a algumas castas, também entre locais (diferenças mais marcantes nas castas Alvarinho, Antão Vaz e Negra Mole). A variação dessa correlação entre locais da mesma casta pode dever-se às condições edafo-climáticas e culturais específicas de cada local. Relativamente às estimativas da variância genotípica e da variância dos blocos subordinados ao ambiente, observa-se que com o ajustamento dos modelos CS e AR1 os valores dessas estimativas são, em geral, menores. Pelo contrário, com o ajustamento dos modelos CS e AR1 a estimativa da variância da interacção $G \times E$ tende, em geral, a ser maior.

Quando se testa a componente de variância da interacção $G \times E$ através de um teste de razão de verosimilhanças restritas (Tabela 4), conclui-se que essa variabilidade é significativa, para qualquer nível de significância usual, em todos os casos estudados, excepto no caso Negra Mole com o modelo IND. Verifica-se também que a diferença entre log-verosimilhanças restritas dos modelos com e sem efeitos da interacção é mais marcante nos modelos CS e AR1, resultando num maior valor calculado da estatística do teste de razão de verosimilhanças restritas (REMLRT). No caso da casta Negra Mole a variabilidade da interacção $G \times E$ apenas se revelou signifi-

| Estimativas | AI | AN | RZ | FP | NM | OV | CR |
|--------------------------------|------------------|------------------|------------------|------------------|------------------|-------------------|------------------|
| Modelo IND | | | | | | | |
| $\hat{\sigma}_g^2 (SE)$ | 0.936 (0.246) | 0.041 (0.015) | 0.054 (0.015) | 0.091 (0.027) | 0.104 (0.027) | 0.381 (0.115) | 0.056 (0.017) |
| $\hat{\sigma}_{b(A)}^2 (SE)$ | 0.233 (0.046) | 0.506 (0.083) | 0.177 (0.029) | 0.205 (0.035) | 0.092 (0.024) | 0.737 (0.165) | 0.145 (0.029) |
| $\hat{\sigma}_{ge}^2 (SE)$ | 0.883 (0.090) | 0.072 (0.021) | 0.046 (0.010) | 0.130 (0.017) | 0.010 (0.010) | 0.321 (0.050) | 0.084 (0.014) |
| $\hat{\sigma}_e^2 (SE)$ | 3.684 (0.087) | 1.830 (0.044) | 0.775 (0.018) | 0.758 (0.020) | 0.689 (0.024) | 1.184 (0.046) | 0.757 (0.021) |
| Modelo CS | | | | | | | |
| $\hat{\sigma}_g^2 (SE)$ | 0.797 (0.215) | 0.013 (0.012) | 0.037 (0.014) | 0.088 (0.026) | 0.067 (0.024) | 0.367 (0.114) | 0.040 (0.016) |
| $\hat{\sigma}_{b(A)}^2 (SE)$ | 0.161 (0.037) | 0.468 (0.078) | 0.164 (0.027) | 0.202 (0.035) | 0.064 (0.018) | 0.729 (0.164) | 0.138 (0.028) |
| $\hat{\sigma}_{ge}^2 (SE)$ | 0.808 (0.077) | 0.100 (0.018) | 0.065 (0.009) | 0.123 (0.016) | 0.041 (0.009) | 0.312 (0.048) | 0.093 (0.014) |
| $\hat{\sigma}_{e_{l1}}^2 (SE)$ | 1.251 (0.139) | 1.842 (0.112) | 0.819 (0.030) | 0.985 (0.041) | 0.868 (0.068) | 1.243 (0.056) | 0.741 (0.024) |
| $\hat{\rho}_{l1} (SE)$ | 0.423 (0.072) | 0.089 (0.046) | 0.273 (0.024) | 0.070 (0.024) | 0.319 (0.067) | 0.100 (0.028) | 0.142 (0.020) |
| $\hat{\sigma}_{e_{l2}}^2 (SE)$ | 5.862 (0.223) | 2.496 (0.108) | 0.702 (0.030) | 0.688 (0.044) | 0.654 (0.034) | 1.012 (0.089) | 0.875 (0.067) |
| $\hat{\rho}_{l2} (SE)$ | 0.087 (0.023) | 0.357 (0.027) | 0.284 (0.029) | 0.059 (0.064) | 0.457 (0.030) | 0.211 (0.066) | 0.110 (0.075) |
| $\hat{\sigma}_{e_{l3}}^2 (SE)$ | 2.571 (0.092) | 1.026 (0.042) | | 0.572 (0.024) | | | |
| $\hat{\rho}_{l3} (SE)$ | 0.189 (0.023) | 0.190 (0.027) | | 0.130 (0.027) | | | |
| Modelo AR1 | | | | | | | |
| $\hat{\sigma}_g^2 (SE)$ | 0.842 (0.224) | 0.019 (0.012) | 0.048 (0.014) | 0.089 (0.026) | 0.080 (0.025) | 0.394 (0.118) | 0.054 (0.017) |
| $\hat{\sigma}_{b(A)}^2 (SE)$ | 0.168 (0.038) | 0.464 (0.077) | 0.170 (0.028) | 0.203 (0.035) | 0.070 (0.019) | 0.734 (0.165) | 0.144 (0.029) |
| $\hat{\sigma}_{ge}^2 (SE)$ | 0.805 (0.078) | 0.091 (0.017) | 0.058 (0.009) | 0.122 (0.016) | 0.037 (0.009) | 0.322 (0.050) | 0.090 (0.014) |
| $\hat{\sigma}_{e_{l1}}^2 (SE)$ | 1.247 (0.130) | 1.850 (0.113) | 0.813 (0.025) | 0.984 (0.040) | 0.873 (0.068) | 1.234 (0.054) | 0.735 (0.022) |
| $\hat{\rho}_{l1} (SE)$ | 0.394 (0.066) | 0.106 (0.055) | 0.232 (0.019) | 0.075 (0.030) | 0.322 (0.067) | -0.109 (0.033) | 0.120 (0.021) |
| $\hat{\sigma}_{e_{l2}}^2 (SE)$ | 5.840 (0.219) | 2.479 (0.096) | 0.697 (0.027) | 0.688 (0.044) | 0.606 (0.027) | 1.009 (0.086) | 0.882 (0.067) |
| $\hat{\rho}_{l2} (SE)$ | 0.056 (0.028) | 0.438 (0.020) | 0.271 (0.024) | 0.059 (0.064) | 0.485 (0.022) | 0.159 (0.065) | 0.116 (0.075) |
| $\hat{\sigma}_{e_{l3}}^2 (SE)$ | 2.607 (0.089) | 1.029 (0.041) | | 0.570 (0.023) | | | |
| $\hat{\rho}_{l3} (SE)$ | 0.305 (0.021) | 0.258 (0.029) | | 0.110 (0.030) | | | |

Tabela 3: Estimativas dos parâmetros covariância (e respectivos erros padrão, SE) resultantes do ajustamento dos modelos com matrizes de covariâncias do vector dos erros aleatórios diagonal (IND), de simetria composta (CS) e autorregressiva de primeira ordem (AR1): $\hat{\sigma}_g^2$ - estimativa da variância genotípica; $\hat{\sigma}_{b(A)}^2$ - estimativa da variância dos blocos subordinados ao ambiente; $\hat{\sigma}_{ge}^2$ - estimativa da variância da interacção $G \times E$, $\hat{\sigma}_{e_l}^2$ - estimativas da variância dos erros aleatórios, $\hat{\rho}_l$ - estimativas das correlações entre observações feitas na mesma unidade experimental (no modelo AR1, correlações entre observações feitas na mesma unidade experimental em dois anos consecutivos).

| Variedade antiga | Modelo | REMLRT | Valor-p |
|---------------------|----------------|--------|---------|
| Alvarinho (AI) | <i>IND</i> | 238.5 | <0.0001 |
| | <i>CS</i> | 253.0 | <0.0001 |
| | <i>AR1</i> | 257.2 | <0.0001 |
| Antao Vaz (AN) | <i>IND</i> | 15.9 | <0.0001 |
| | <i>CS</i> | 57.1 | <0.0001 |
| | <i>AR1</i> | 51.4 | <0.0001 |
| Aragonez (RZ) | <i>R = IND</i> | 35.0 | <0.0001 |
| | <i>CS</i> | 97.7 | <0.0001 |
| | <i>AR1</i> | 71.0 | <0.0001 |
| Fernaio Pires (FP) | <i>IND</i> | 153.5 | <0.0001 |
| | <i>CS</i> | 157.4 | <0.0001 |
| | <i>AR1</i> | 147.3 | <0.0001 |
| Negra Mole (NM) | <i>IND</i> | 1.0 | 0.3142 |
| | <i>CS</i> | 37.7 | <0.0001 |
| | <i>AR1</i> | 36.1 | <0.0001 |
| Rabo de Ovelha (OV) | <i>IND</i> | 125.2 | <0.0001 |
| | <i>CS</i> | 126.5 | <0.0001 |
| | <i>RAR1</i> | 123.0 | <0.0001 |
| Sírria (CR) | <i>IND</i> | 64.6 | <0.0001 |
| | <i>CS</i> | 95.9 | <0.0001 |
| | <i>AR1</i> | 81.2 | <0.0001 |

Tabela 4: Teste de razão de verossimilhanças restritas à componente de variância da interacção genótipo \times ambiente ($H_0 : \sigma_{ge}^2 = 0$ vs $H_0 : \sigma_{ge}^2 > 0$) para cada um dos modelos com matrizes de covariâncias do vector dos erros aleatórios diagonal (*IND*), de simetria composta (*CS*) e autorregressiva de primeira ordem (*AR1*) e respectivo valor-p. REMLRT, valor calculado da estatística do teste de razão de verossimilhanças restritas.

cativa com os ajustamento destes últimos modelos. Este resultado reforça a vantagem da utilização da abordagem proposta no estudo da interacção $G \times E$. Finalmente, deve-se acrescentar que, inerente a este resultado, está a consequente vantagem na posterior utilização dos melhores preditores empíricos lineares não enviesados dos efeitos da interacção $G \times E$, obtidos com o ajustamento dos modelos CS e AR1, para fins de selecção de clones com menor sensibilidade a essa interacção.

4 Conclusões

Analisando dados de rendimento, os modelos que admitiram que observações na mesma unidade experimental não são independentes (modelos CS e AR1), revelaram melhor ajustamento face ao modelo que admitiu erros independentes (modelo IND), afectando, consequentemente, a avaliação da interacção $G \times E$. De entre as estruturas de covariância estudadas, o modelo CS revelou-se quase sempre como o mais adequado. Isto revela que em ensaios com videira a estrutura que mais se adequa é a que traduz a existência de um efeito comum, ainda que pequeno a moderado, a todas as observações feitas na mesma unidade experimental (o solo que partilha, a estrutura radicular, etc.).

Resumindo, este trabalho dá resposta ao modelo que deverá ser usado no estudo da interacção $G \times E$ na fase final da metodologia de selecção da videira. Este conhecimento é fundamental, pois é o primeiro passo para alcançar o objectivo final: selecção dos clones com menor sensibilidade à interacção $G \times E$. Essa selecção deverá basear-se nos melhores preditores empíricos lineares não enviesados dos efeitos da interacção $G \times E$ obtidos com o ajustamento desse modelo (quanto mais próximos de zero, menor a sensibilidade à interacção).

Agradecimentos

Aos colegas da “Rede Nacional de Selecção da Videira” e da Associação Portuguesa para a Diversidade da Videira (PORVID) pela sua contribuição em todo o processo de selecção das castas. À Fundação para a Ciência e Tecnologia (UID/AGR/04129/2013) pelo apoio financeiro.

Referências

- [1] Butler, D., Cullis, B.R., Gilmour, A.R., Gogel, B.J. (2007). ASReml-R reference manual. AsReml-R estimates variance components under a general linear mixed model by residual maximum likelihood (REML). NSW Department of Primary Industries, Queensland Government. Queensland.
- [2] De Faveri, J., Verbyla, A.P., Pitchford, W.S., Venkatanagappa, S., Cullis, B.R.. Statistical methods for analysis of multi-harvest data from perennial pasture variety selection trials. *Crop and Pasture Science*, 66, 947–962, 2015.
- [3] Gonçalves, E., Carrasquinho, I., Almeida, R., Pedroso, V., Martins, A.. Genetic correlations in grapevine and their effects on selection. *Australian Journal of Grape and Wine Research*, 22, 52–63, 2016.
- [4] Gonçalves, E., Martins, A.. Metodologias estatísticas para estudo da interacção genótipo×ambiente em clones de videira. *Atas XXI Congresso da Sociedade Portuguesa de Estatística*, 89–103, 2014.
- [5] Martins, A., Gonçalves, E.. Grapevine breeding programmes in Portugal. In *Grapevine Breeding Programs for the Wine Industry*. A. G. Reynolds ed., Woodhead Publishing, Elsevier, UK, 159–182, 2015.
- [6] Mathews, K.L., Chapman, S.C., Trethowan, R., Pfeiffer, W., Ginkel, M., Crossa, J., Payne, T., DeLacy, I., Fox, P.N., Cooper, M.. Global adaptation patterns of Australian and CIMMYT spring bread wheat. *Theor Appl Genet*, 115, 819–855, 2007.
- [7] McCulloch, C.E., Searle, S.R., Neuhaus, J.M. (2008). *Generalized, linear and mixed models*. John Wiley & Sons, New York.

- [8] Patterson, H.D., Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554.
- [9] Piepho, H.P., Eckl, T.. Analysis of series of variety trials with perennial crops. *Grass and Forage Science*, 69, 431–440, 2014.
- [10] Sakamoto, Y., Ishiguro, M., Kitagawa, G. (1986). Akaike information criterion statistics. Dordrecht: D. Reidel.
- [11] Stram, D.O., J.W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* 50, 1171–1177.
- [12] Stroup, W.W. (2013). Generalized Linear Mixed Models: Modern Concepts, Methods and Applications. CRC Press, Boca Raton.