

Análise com e sem uso de funções polinomiais no ajuste espacial de dados nas etapas iniciais de programas de seleção – comparação via simulação

Márcio Pupin de Mello¹; Luiz Alexandre Peterlini²; Mauro Sergio de Oliveira Leite³

Introdução

Um dos grandes problemas encontrados nas etapas iniciais de programas de melhoramento genético é a escassez de materiais perante o grande número necessário para a realização do experimento (DUARTE, 2000; MATSUOKA et al., 1999). Nesse contexto, é importante a realização de estudos que visem otimizar os recursos sem comprometer as avaliações genéticas (SOUZA et al., 2006; PETERNELLI e BARBOSA, 2004).

Uma alternativa viável seria o uso de apenas uma repetição de cada novo genótipo em experimentos segundo o Delineamento Aumentado (FEDERER, 1956). Recentemente, Federer (1998) e Federer et al. (2001) propõem dispor as unidades experimentais em “*grids*” que contemplem variações bidimensionais. A análise inicia-se com uma função polinomial envolvendo esse *grid*, que absorva da melhor maneira possível, as variações sistemáticas dos dados da amostra, segundo o modelo de delineamento usado. A função então corrigirá os valores associados a cada uma dessas unidades experimentais para prosseguir a análise.

O objetivo deste trabalho foi comparar três modelos de análise, onde ocorre ou não o uso de funções polinomiais para ajuste da variabilidade espacial em experimentos de campo simulados, usando o software livre R.

Material e Métodos

Os dados foram simulados como num experimento em campo com: 20 linhas; 10 colunas; 10 blocos (cada duas linhas formam um bloco); 4 genótipos testemunha (901, a 904) que se repetem 10 vezes cada. O restante das unidades experimentais foi preenchida com os novos genótipos (1 a 160). No total, temos 200 unidades experimentais (20 linhas x 10 colunas) onde a distribuição das testemunhas nos blocos se deu de duas maneiras: aleatória (ALE) e sistemática (SIS) (Figura 1).

¹ Mestrando em Sensoriamento Remoto – Instituto Nacional de Pesquisas Espaciais (INPE) – pupin@dsr.inpe.br

² Professor Adjunto IV – DPI / Setor de Estatística – Universidade Federal de Viçosa (UFV) – peterlini@dpi.ufv.br

³ Doutorando em Genética e Melhoramento – Universidade Federal de Viçosa (UFV) – mauro@vicosa.ufv.br



Figura 1 - Exemplo de desenho experimental com distribuições (a) ALEatória e (b) SIStemática das testemunhas

Os dados foram simulados seguindo o modelo: $y_{ijk} = \mu + g_i + \psi_{ijk}$, onde: y_{ijk} : valor associado ao genótipo i , na unidade experimental jk ; μ : constante do modelo; g_i : efeito do i -ésimo genótipo; ψ_{ijk} : erro aleatório associado ao valor do i -ésimo genótipo na j -ésima linha e k -ésima coluna, que pode ser decomposto em: $\psi_{ijk} = l_j + c_k + \varepsilon_{ijk}$, onde: l_j : efeito da j -ésima linha; c_k : efeito da k -ésima coluna; ε_{ijk} : erro puro associado ao i -ésimo genótipo na j -ésima linha e k -ésima coluna.

Foram realizadas 1000 simulações de cada um de 72 cenários, onde esses se variavam entre ALE e SIS, Coeficiente de Variação (cv=10;30%), correlação no eixo x (cx=0;30;60%), correlação no eixo y (cy=0;30;60%) e herdabilidade (H=30;70%), conforme Tabela 1.

Tabela 1: Definição dos cenários em função dos valores dos fatores incluídos nas simulações

c	10										30									
v	10										30									
c	0			30			60			0			30			60				
x	0			30			60			0			30			60				
y	0	30	60	0	30	60	0	30	60	0	30	60	0	30	60	0	30	60		
H	3	7	3	7	3	7	3	7	3	7	3	7	3	7	3	7	3	7		
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
ALEATÓRIO - ALE																				
n	1	2	3	4	5	6	7	8	9	1	1	1	1	1	1	1	1	1		
c										0	1	2	3	4	5	6	7	8	9	
SISTEMÁTICA - SIS																				
n	3	3	3	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	
c	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	

nc: número do cenário (1 a 72); H: herdabilidade (30 e 70%); cy: correlação espacial na direção das linhas (0, 30 e 60%); cx: correlação espacial na direção das colunas (0,30 e 60%); cv: coeficiente de variação experimental (10 e 30%).

Três tipos de análise foram usadas: “DBA” (Delineamento em Blocos Aumentados), “DBA + Espacial” e “Espacial”. Para os casos onde funções Espaciais

são envolvidas seguem o modelo utilizado apenas para as linhas e para as colunas e linhas, respectivamente: $y_{ijk} = \mu + g_i + b_j + c_k^1 + c_k^2 + c_k^3 + \dots + c_k^d + \varepsilon_{ijk}$, e

$y_{ijk} = \mu + g_i + l_j^1 + c_k^1 + l_j^2 + l_j^1.c_k^1 + c_k^2 + l_j^3 + l_j^2.c_k^1 + l_j^1.c_k^2 + c_k^3 + \dots + c_k^d + \varepsilon_{ijk}$, onde: y_{ijk} : valor associado ao genótipo i, unidade experimental jk; μ : constante; g_i : efeito do i-ésimo genótipo; b_j : efeito do j-ésimo bloco; l_j^n : efeito da j-ésima linha elevada a n-ésima potência; c_k^d : efeito da k-ésima coluna elevada a d-ésima potência; $l_j^n.c_k^m$: efeito da interação entre a j-ésima linha elevada a n-ésima potência e a k-ésima coluna elevada a m-ésima potência; ε_{ijk} : erro puro.

Resultados e Discussões

Foram avaliados: a porcentagem de coincidência entre os 10% melhores genótipos (%c) e a correlação de Spearman entre os resultados recuperados pela análise e os dados simulados, além das diferenças entre herdabilidades (H) recuperadas e simuladas.

Em praticamente todos os casos (Figura 2) o valor da %c entre as análises espaciais (DBA + espacial, Espacial) foi superior ao DBA. O DBA foi superior ao Espacial apenas nos cenários 2, 22, 40, 57 e 59, que não apresentam nenhuma particularidade em comum, levando-nos a concluir que isso se deu apenas devido ao acaso. Não ocorreram casos onde o DBA ou o Espacial tivesse porcentagem de coincidência superior ao DBA + espacial.

Em média o DBA + espacial teve porcentagem de coincidência 2,8% maior que o DBA e 2,2% maior que o Espacial que por sua vez foi apenas 0,6% maior que o DBA. Logo, o DBA + espacial foi o método mais eficiente em ranquear os genótipos.

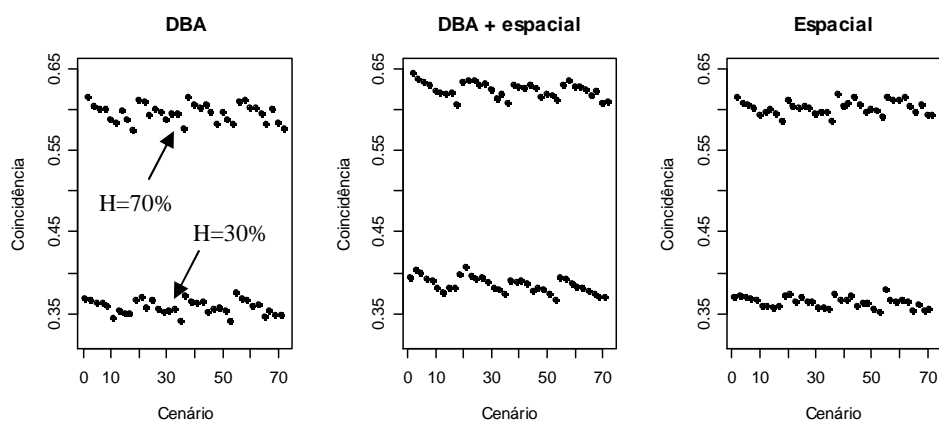


Figura 2 - Diagrama das Porcentagens de coincidência no ranqueamento para as três análises

Na Figura 2 observa-se que cenários consecutivos possuem alternância entre baixa e alta %c de acordo com a H. Na Tabela 2 se percebe uma nítida diferença nas %c médias para os dois valores de herdabilidade em cada modelo de análise. Além disso, observa-se que existem 4 grupos distintos (em cada valor de H) onde a porcentagem de coincidência começa com certo valor e, a medida que o número do cenário aumenta, ela diminui. Os 4 grupos são explicados pelas variações do valor do CV e as diminuições graduais da porcentagem de coincidência são explicadas pelo aumento da correlação espacial dentro de cada grupo de CV. Isso causa uma diminuição gradual na %c entre o ranqueamento. Contudo, mesmo nos casos onde não há correlação espacial (cx e /ou cy iguais a zero) o DBA + espacial mostrou-se mais eficiente que os outros métodos.

Tabela 2 – Valores médios das Porcentagens de Coincidência das 3 análises dentro de cada grupo de valores de herdabilidade

H (herdabilidade)	DBA	DBA + espacial	Espacial	Média Geral
30%	35,7%	38,5%	36,3%	36,8%
70%	59,6%	62,4%	60,3%	60,8%

Observe que, em média, as porcentagens de coincidência dos cenários onde a herdabilidade é de 70% foi 24% maior que nos casos onde a herdabilidade é de 30%.

Da mesma forma a H influenciou na correlação entre os efeitos dos novos genótipos estimados com os simulados (Figura 3). Os melhores resultados foram encontrados nos cenários onde a herdabilidade era alta (H=70%). A análise DBA + espacial foi a que apresentou melhores resultados.

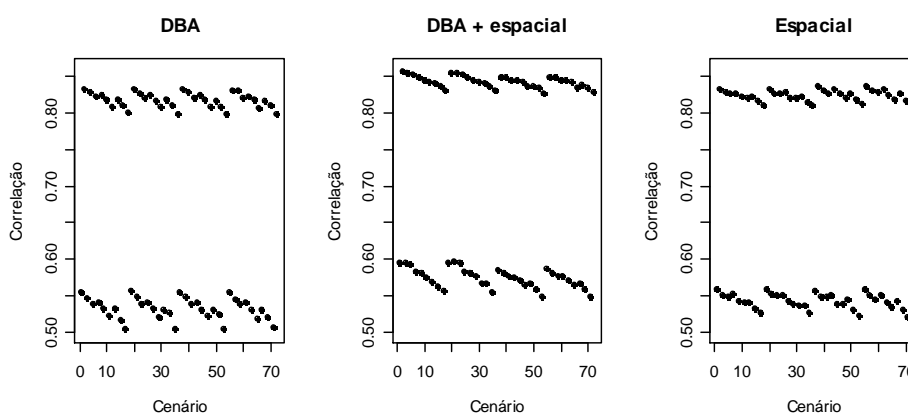


Figura 3 - Diagrama das correlações de Spearman obtidas para as três análises

MELLO, M. P.; PETERNELLI, L. A.; LEITE, M. S. O. Análise com e sem uso de funções polinomiais no ajuste espacial de dados nas etapas iniciais de programas de seleção – comparação via simulação. In: Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria (RBRAS) / Simpósio de Estatística Aplicada a Experimentação Agronômica, 52. / 12., 2007, Santa Maria, RS. **Resumos...** São Paulo, SP: RBRAS, 2007. p. 1-5. CD-ROM. Disponível em: <<http://mtc-m17.sid.inpe.br:80/rep-/sid.inpe.br/mtc-m18@80/2009/04.23.04.44>>. Acesso em: 29 abr. 2008.

A forma de distribuição das testemunhas nas unidades experimentais influenciou apenas na estimativa da herdabilidade. Na SIS os valores de H tenderam a ser menos superestimados que na ALE, especialmente nas análises DBA + espacial e Espacial (dados não apresentados).

Conclusões

a) Quanto maior a herdabilidade (H), maior a porcentagem de coincidência entre os ranqueamentos, assim como é maior a correlação entre os efeitos dos novos genótipos estimados com os simulados; b) A distribuição das testemunhas influenciou apenas na estimativa da herdabilidade, sendo que os valores de H tenderam a ser menos superestimados na SIS; c) O DBA mostrou-se o mais eficiente dos métodos de análises avaliados, quanto à estimativa da herdabilidade; d) O uso de métodos de ajuste espacial dentro dos blocos (DBA + espacial) melhora a seleção, no sentido de melhorar as porcentagens de coincidência dos genótipos selecionados com os verdadeiros melhores genótipos simulados.

Bibliografia

DUARTE, J.B. *Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal*. Piracicaba, 2000. 293p. Tese (Doutorado) – ESALQ/USP.

FEDERER, W.T. Recovery of interblock, intergradient, and intervariety information in incomplete block and lattice rectangle designed experiments. *Biometrics*, v. 54, p. 471-481. 1998.

FEDERER, W.T.; REYNOLDS, M.; CROSSA, J. Combining results from augmented designs over sites. *Agronomy J.*, v. 93, p. 389-395. 2001.

FEDERER, W.T. Augmented (hoonuiaku) designs. *Hawaiian Planters' Record, Aica*, v. 55, p.191-208. 1956.

MATSUOKA, S.; GARCIA, A.A.F; ARIZONO, H. In: *BORÉM, A. Melhoramento de espécies cultivadas*. Viçosa: Editora UFV, 1999. cap. 7, p. 205-251.

PETERNELLI, L.A.; BARBOSA, M.H.P. Desafios na condução e planejamento de experimentos de avaliação de clones de cana-de-açúcar. In: *Anais do VII Simpósio de Atualização em Genética e Melhoramento de Plantas: desafios e perspectivas da interação da interação biometria e o melhoramento de plantas*. Editora UFLA, Lavras. 2004. p 17-27.

SOUZA, E.F.M.; PETERNELLI, L.A.; BARBOSA, M.H.P. Designs and model effects definitions in the initial stage of a plant breeding program. *Pesquisa Agropecuária Brasileira, Brasília*, v. 41, n. 3, p. 369-375. 2006.