

Avaliando delineamentos parcialmente replicados: Um estudo de simulação

Renata Alcarde Sermarini ¹, Chris Brien ², Clarice Garcia Borges Demétrio ³,
Alessandra dos Santos ³

Introdução

Em estudos de melhoramento genético o principal objetivo consiste em maximizar o ganho genético, o que envolve a seleção dos melhores genótipos avaliados, sendo a experimentação uma parte essencial para a tomada de decisões. Desse modo a obtenção de delineamentos eficientes tem grande relevância em tais estudos e os delineamentos ótimos consistem um modo de se obter um delineamento tão eficiente quanto possível.

Tratando-se do melhoramento genético de plantas, em fases iniciais do programa tem-se um grande número de genótipos a serem avaliados e uma limitação no material genético disponível para estudo, bem como na área experimental, não sendo possível a realização de repetições de todos os tratamentos de interesse. Consequentemente, delineamentos grid-plot são amplamente utilizados. Estes são delineamentos em linhas e colunas que não apresentam repetições dos novos genótipos e sim um grid de parcelas contendo repetições para variedades padrão (KEMPTON, 1984).

Como alternativa a tais delineamentos, Cullis, Smith e Coombes (2006) propuseram os delineamentos parcialmente replicados, chamados *p*-rep *designs*, em que *p* representa a proporção de novos genótipos que apresentam repetição. Em tais delineamentos, as parcelas ocupadas pelas variedades padrão são substituídas por parcelas com repetições dos novos genótipos. Cullis, Smith e Coombes (2006) afirmam que os delineamentos *p*-rep apresentam uma melhor seleção de novas variedades quando comparados aos delineamentos grid-plot. Tal superioridade foi confirmada por Clarke e Stefanova (2011) e Santos (2017).

Os *p*-rep *designs* foram inicialmente propostos considerando-se delineamentos em linhas e colunas espacialmente otimizados, usualmente apresentando dois blocos e, desse modo, duas repetições para os genótipos replicados, caracterizando assim delineamentos resolvíveis. Entretanto, Clarke e Stefanova (2011) não empregaram delineamentos resolvíveis, enquanto que Santos (2017) sim.

De acordo com Piepho, Williams e Michel (2016) delineamentos resolvíveis em linhas e colunas são utilizados em experimentos com um grande número de tratamentos, já os não resolvíveis são frequentemente utilizados quando o número de tratamentos, bem como suas repetições são pequenos.

Embora não seja bem conhecida a importância de delineamentos resolvíveis para a seleção de novos genótipos, existe um esforço em afastar as repetições do mesmo tratamento tanto quanto possível, mesmo que o delineamento seja não resolvível (PIEPHO; WILLIAMS, 2016; WILLIAMS; PIEPHO, 2017).

¹LCE/ESALQ/USP. e-mail: ralcarde@usp.br

²Australian Plant Phenomics Facility, The Plant Accelerator®, School of Agriculture Food & Wine, University of Adelaide, Urrbrae, SA 5064, Australia. e-mail: chris.brien@unisa.edu.au

³LCE/ESALQ/USP. e-mail: clarice.demetrio@usp.br

³LCE/ESALQ/USP. e-mail: ale.estatistica@gmail.com

Conforme mencionado anteriormente, delineamento eficientes podem ser obtidos por meio de delineamentos ótimos, entretanto estes delineamentos são ótimos segundo algum critério previamente estabelecido. O critério mais adotado em estudos de melhoramento genético de plantas é aquele que consistem em minimizar a média das variâncias das diferenças entre os pares de efeitos de tratamentos (AVPD - *average variance pairwise differences of effects*), sendo este um critério que depende da matriz de informação relacionada aos tratamentos de interesse.

Tal matriz de informação será influenciada por suposições referentes ao modelo para o delineamento e conseqüentemente, se efeitos de tratamentos são assumidos fixos ou aleatórios. Smith, Cullis e Thompson (2005) afirmam que se o interesse da pesquisa consiste em determinar diferenças entre pares de tratamentos específicos, então os efeitos de tratamentos devem ser fixos, e neste caso o critério AVPD é indicado. Por outro lado, Cullis, Smith e Coombes (2006) e Bueno Filho e S. G. Gilmour (2007) verificaram que minimizar a variância das diferenças entre os pares dos erros de predição é equivalente a maximizar o ganho genético esperado, no qual os efeitos de tratamentos são assumidos aleatórios.

O presente estudo tem como foco a avaliação dos *p*-rep *designs* espacialmente otimizados por meio de estudos de simulação, verificando como o ganho genético é afetado por características do modelo assumido para a obtenção do delineamento ótimo. Neste, as características estudadas foram (i) o efeito de genótipos, se fixo ou aleatório e (ii) a condição do delineamento ser resolvível ou não.

Estudo de Simulação

Modelo para o delineamento

Foram avaliados 24 *p*-rep *designs* por meio de estudo de simulação, considerando-se uma área experimental fixa correspondendo a 20 linhas por 24 colunas, em que as parcelas são retangulares e estreitas na direção das linhas. Também foi fixado o número de genótipos, sendo 400, dos quais 320 não apresentaram repetição e 80 foram duplicados, caracterizando $p = 20\%$ como parâmetro de repetição. Tais valores são comuns em estudo de melhoramento genético de cana-de-açúcar (SANTOS, 2017).

Os modelos para a obtenção dos delineamentos são definidos conforme A. R. Gilmour, Cullis e Verbyla (1997),

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon} + \boldsymbol{\eta}, \quad (1)$$

em que \mathbf{y} é o vetor de observações na ordem do campo (colunas \times linhas), $\boldsymbol{\beta}$ é o vetor de efeitos fixos com matriz de delineamento \mathbf{X} , \mathbf{u} é o vetor de efeitos aleatórios, com matriz de delineamento \mathbf{Z} , $\boldsymbol{\epsilon}$ é o vetor de erros espacialmente dependentes e $\boldsymbol{\eta}$ é o vetor de erros independentes (ruído branco). Assumiu-se $(\mathbf{u}, \boldsymbol{\epsilon}, \boldsymbol{\eta})$ com média zero e independentes, tais que $\text{Var}(\mathbf{u}) = \sigma^2 \mathbf{G}(\boldsymbol{\gamma})$, $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{R}(\boldsymbol{\phi}) = \sigma^2 \boldsymbol{\Sigma}_c(\boldsymbol{\phi}_c) \otimes \boldsymbol{\Sigma}_r(\boldsymbol{\phi}_r)$, dada por processos separáveis auto-regressivos de primeira ordem e $\text{Var}(\boldsymbol{\eta}) = \sigma^2 \gamma_n \mathbf{I}$, em que \mathbf{I} é uma matriz identidade.

Os delineamentos foram obtidos de acordo com seis conjuntos de parâmetros não genéticos, dados de acordo com os valores obtidos por Santos (2017), denotados por (a) $\gamma_c = \gamma_r = 0,30$ e $\gamma_n = 0,10$; (b) $\gamma_c = \gamma_r = 0,30$ e $\gamma_n = 0,50$; (c) $\gamma_c = \gamma_r = 0,75$ e $\gamma_n = 0,10$; (d) $\gamma_c = \gamma_r = 0,75$ e $\gamma_n = 0,50$; (e) $\gamma_c = \gamma_r = 2,50$ e $\gamma_n = 0,10$ e (f) $\gamma_c = \gamma_r = 2,50$ e $\gamma_n = 0,50$, fixando-se $\sigma^2 = 1$, $\phi_c = 0,5$ e $\phi_r = 0,4$, em que γ_c e

γ_r são os termos de variância correspondentes aos efeitos aleatórios de colunas e linhas, respectivamente, tais que $\gamma_i = \sigma_i^2/\sigma^2$.

Delineamentos em que a AVPD é minimizada foram obtidos por meio da biblioteca *od* (BUTLER, 2014) para o software estatístico R (R CORE TEAM, 2017), considerando-se a estratégia de busca aleatória e 20.000 iterações, de acordo com as combinações das condições para o efeito de genótipos (fixo ou aleatório) e se o delineamento era resolvível ou não. Quando os efeitos de genótipos foram considerados aleatórios seu correspondente componente de variância foi dado por $\sigma_g^2 = 1$. Quando o delineamento foi resolvível, este apresentou dois blocos, correspondendo a grupos de 12 colunas consecutivas, representando repetições completas dos genótipos repetidos.

Processo de Simulação

Como o foco do estudo é a seleção genética em experimentos de fases iniciais do melhoramento de plantas, foram examinados apenas os casos em que os efeitos de genótipos são aleatórios. O modelo para gerar os dados foi dado por:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_o\mathbf{u}_o + \boldsymbol{\epsilon} + \boldsymbol{\eta},$$

em que \mathbf{u}_g é o vetor de efeitos aleatórios de genótipos com matriz de delineamento \mathbf{Z}_g e \mathbf{u}_o é o vetor de efeitos aleatórios de linhas e colunas com matriz de delineamento \mathbf{Z}_o . Os demais termos são descritos conforme Equação 1. Ainda, assumiu-se $(\mathbf{u}_g, \mathbf{u}_o, \boldsymbol{\epsilon}, \boldsymbol{\eta})$ com média zero e independentes, tal que $\text{Var}(\mathbf{u}_g) = \sigma^2 \gamma_g \mathbf{I}$, $\text{Var}(\mathbf{u}_o) = \sigma^2 \mathbf{G}(\boldsymbol{\gamma}_o)$, em que $\boldsymbol{\gamma}_o = (\gamma_c, \gamma_r)'$ e $\gamma_i = \sigma_i^2/\sigma^2$.

Foram gerados dados para cada delineamento a partir da combinação de dois valores para γ_g (0,6 ou 2,0) com os seis conjuntos de parâmetros não genéticos definidos como no modelo do delineamento, com e sem efeito de blocos. Na presença do efeito de blocos somou-se e subtraiu-se aos dados uma constante, $\sqrt{5}$, grande o suficiente para corresponder a um efeito significativo, de acordo com o bloco. Desse modo, foram avaliados 24 cenários para os dados.

Uma única coleção de 1000 conjuntos de efeitos aleatórios de genótipos foi simulada e utilizada em todos os casos de acordo com γ_g . Procedeu-se de modo análogo para a simulação dos demais efeitos aleatórios, compondo 1000 conjuntos de dados para cada cenário avaliado.

Os ajustes dos modelos se deram conforme o modelo utilizado para a obtenção do delineamento ótimo, com a exceção dos efeitos de genótipos serem sempre aleatórios. Quando o delineamento foi resolvível, os efeitos de blocos foram considerados fixos no modelo.

Para cada modelo ajustado, foram registrados os valores preditos dos genótipos, com o objetivo de calcular o ganho genético (CULLIS; SMITH; COOMBES, 2006) para uma seleção de 10% dos genótipos com melhor desempenho.

As simulações, bem como os ajustes dos modelos foram realizados no software estatístico R (R CORE TEAM, 2017) e com o uso do pacote *asreml-R* (BUTLER, 2009).

Resultados e Discussão

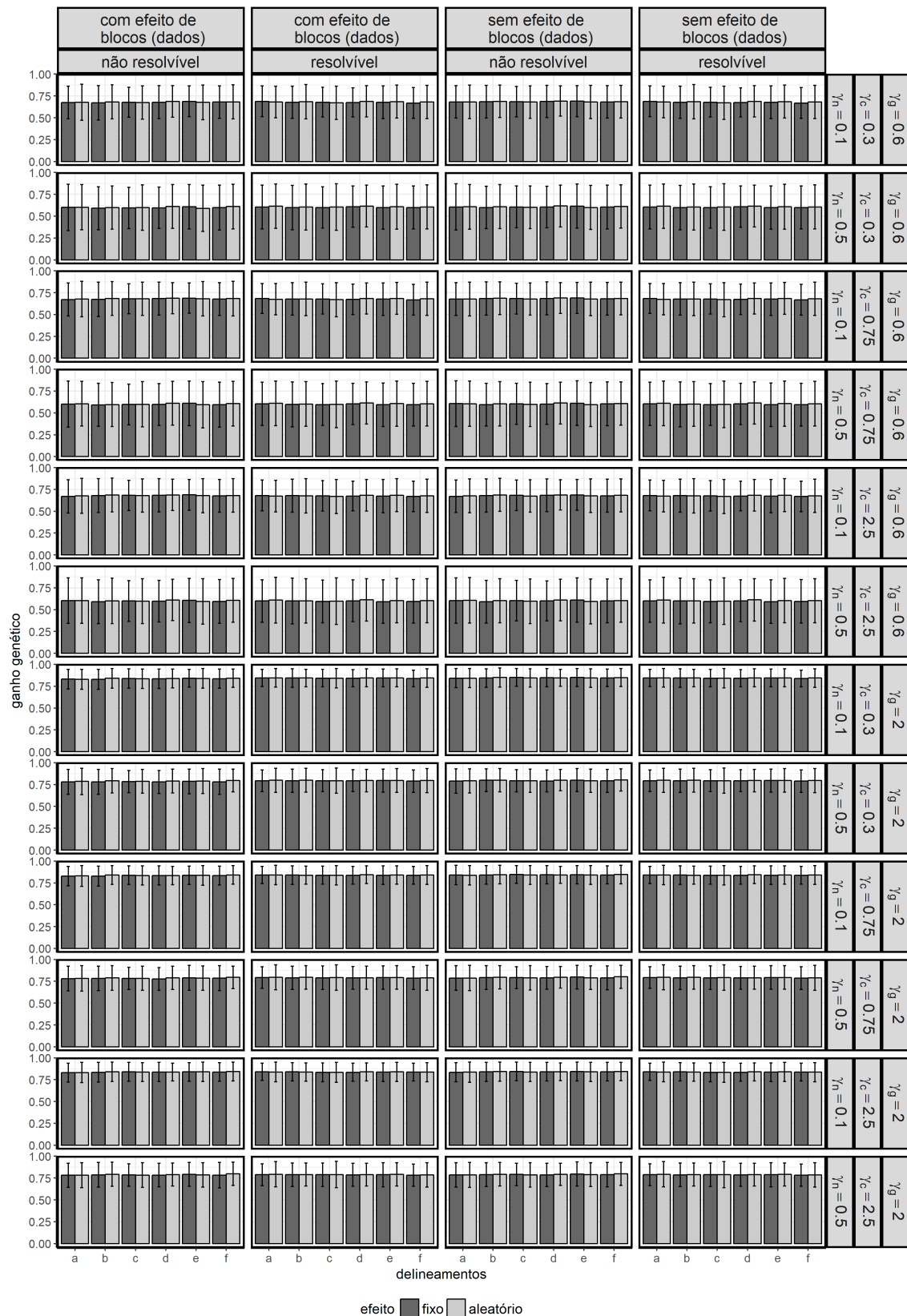


Figura 1: Ganho genético (mediana \pm amplitude interquartílica) para cada delineamento p -rep espacialmente otimizado de acordo com o critério que minimiza a AVPD, considerando-se os diferentes cenários para os conjuntos de dados (a) $\gamma_c = \gamma_r = 0,30$ e $\gamma_n = 0,10$; (b) $\gamma_c = \gamma_r = 0,30$ e $\gamma_n = 0,50$; (c) $\gamma_c = \gamma_r = 0,75$ e $\gamma_n = 0,10$; (d) $\gamma_c = \gamma_r = 0,75$ e $\gamma_n = 0,50$; (e) $\gamma_c = \gamma_r = 2,50$ e $\gamma_n = 0,10$ e (f) $\gamma_c = \gamma_r = 2,50$ e $\gamma_n = 0,50$

Uma característica dos delineamentos para os quais os efeitos de genótipos foram aleatórios é a alta frequência de genótipos repetidos em uma mesma coluna, apresentando uma alocação pobre para as repetições dos tratamentos. Entretanto, não foi observado confundimento total do efeito de colunas com o efeito de genótipos. Comportamento similar foi relatado por Cullis, Smith e Coombes (2006) avaliando *p*-rep designs com 120 genótipos, dos quais 30 foram duplicados, em uma área experimental de 30 linhas por 5 colunas quando $\gamma_g = 0.1$. Neste caso, os genótipos que apresentam repetições se encontraram próximos as bordas do experimento.

A convergência foi obtida em no mínimo 994 dos 1000 conjuntos de dados para cada cenário avaliado.

De modo geral, não existem grandes diferenças entre os delineamentos. É evidente que as características dos dados são as responsáveis por alterações no comportamento do ganho genético, conforme pode ser observado na Figura 1. Uma melhor seleção é possível quando os dados apresentam uma maior variabilidade genética e menor variância associada ao ruído branco, como esperado.

Embora diferentes valores para os parâmetros não genéticos, supostos no modelo para a obtenção dos delineamentos, possam afetar a variâncias das diferenças entre os pares dos efeitos de tratamentos, como apontado por Williams e Piepho (2013) em estudo no qual todos os tratamentos foram repetidos, nota-se, observando-se a Figura 1, que o ganho genético não é influenciado por tais valores, assim como não é afetado pela característica do delineamento, se resolvível ou não.

Adicionalmente, não foram observadas grandes diferenças entre delineamentos que consideraram efeitos de genótipos fixos e delineamentos que consideraram efeitos de genótipos aleatórios em seu modelo, com relação ao ganho genético. Tal resultado corrobora os de Cullis, Smith e Coombes (2006) para uma seleção de 20% dos melhores genótipos. Quando os efeitos de genótipos foram assumidos fixos, o ganho genético apresentou uma melhor precisão.

Conclusão

Este estudo limitou-se a avaliar os *p*-rep designs fixando-se o tamanho do experimento, bem como o número de tratamentos avaliados, considerando-se 20% dos genótipos repetidos. Neste caso, não foram observadas diferenças entre os diferentes delineamentos espacialmente otimizados, assumindo-se o critério que minimiza AVPD, com relação ao ganho genético (seleção de 10% dos melhores genótipos), uma das principais características de interesse em estudo de melhoramento genético de plantas. Sendo assim, conclui-se que os *p*-rep designs são robustos sob tais condições.

Agradecimentos

Processo nº 2016/26003-4, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Referências

- BUENO FILHO, J. S. De S.; GILMOUR, S. G. Block designs for random treatment effects. v. 137, p. 1446–1451, 2007. DOI: [10.1016/j.jspi.2006.02.002](https://doi.org/10.1016/j.jspi.2006.02.002).
- BUTLER, D. **asreml: asreml() fits the linear mixed model**. [S.l.], 2009. R package version 3.0. Disponível em: www.vsni.co.uk.
- _____. **od: generate optimal designs**. [S.l.], 2014. R package version 0.75.
- CLARKE, G. P. Y.; STEFANOVA, K. T. Optimal design for early-generation plant-breeding trials with unreplicated or partially replicated test lines. **Australian and New Zealand Journal of Statistics**, v. 53, n. 4, p. 461–480, 2011. ISSN 13691473. DOI: [10.1111/j.1467-842X.2011.00642.x](https://doi.org/10.1111/j.1467-842X.2011.00642.x).
- CULLIS, B. R.; SMITH, A. B.; COOMBES, N. E. On the design of early generation variety trials with correlated data. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 11, n. 4, p. 381–393, 2006. ISSN 1085-7117. DOI: [10.1198/108571106X154443](https://doi.org/10.1198/108571106X154443).
- GILMOUR, A. R.; CULLIS, B. R.; VERBYLA, A. P. Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 2, n. 3, p. 269–293, 1997. Disponível em: <http://www.jstor.org/stable/1400446>.
- KEMPTON, R. A. The design and analysis of unreplicated field trials. **Vortrafe für Pflanzenzuchtung**, v. 7, p. 219–242, 1984.
- PIEPHO, H.-P.; WILLIAMS, E. R. Augmented row–column designs for a small number of checks. **Agronomy Journal**, v. 108, n. 6, p. 2256–2262, 2016. ISSN 14350645. DOI: [10.2134/agronj2016.06.0325](https://doi.org/10.2134/agronj2016.06.0325).
- PIEPHO, H.-P.; WILLIAMS, E. R.; MICHEL, V. Nonresolvable row–column designs with an even distribution of treatment replications. **Journal of Agricultural, Biological, and Environmental Statistics**, Springer US, v. 21, n. 2, p. 227–242, 2016. ISSN 15372693. DOI: [10.1007/s13253-015-0241-2](https://doi.org/10.1007/s13253-015-0241-2).
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2017. Disponível em: <https://www.R-project.org/>.
- SANTOS, A. dos. **Design and analysis of sugarcane breeding experiments : a case study**. 2017. f. 173. Tese (Doutorado) – Universidade de São Paulo.
- SMITH, A. B.; CULLIS, B. R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **The Journal of Agricultural Science**, v. 143, n. 6, p. 449–462, 2005. ISSN 0021-8596. DOI: [10.1017/S0021859605005587](https://doi.org/10.1017/S0021859605005587). Disponível em: http://www.journals.cambridge.org/abstract%7B%5C_%7DS0021859605005587.
- WILLIAMS, E. R.; PIEPHO, H.-P. A comparison of spatial designs for field variety trials. **Australian and New Zealand Journal of Statistics**, v. 55, n. 3, p. 253–258, 2013. ISSN 13691473. DOI: [10.1111/anzs.12034](https://doi.org/10.1111/anzs.12034).
- _____. An Evaluation of Error Variance Bias in Spatial Designs. **Journal of Agricultural, Biological, and Environmental Statistics**, Springer US, v. 23, n. 1, p. 1–9, 2017. ISSN 15372693. DOI: [10.1007/s13253-017-0309-2](https://doi.org/10.1007/s13253-017-0309-2).