

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Delineamentos ótimos para experimentos multi-ambientais de
melhoramento genético de plantas**

Denize Palmito dos Santos

Tese apresentada para obtenção do título de Doutora
em Ciências. Área de concentração: Estatística e Ex-
perimentação Agronômica

**Piracicaba
2023**

Denize Palmito dos Santos
Licenciada em Matemática

**Delineamentos ótimos para experimentos multi-ambientais de
melhoramento genético de plantas**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientadora:

Profa. Dra. **RENATA ALCARDDE SERMARINI**

Tese apresentada para obtenção do título de Doutora
em Ciências. Área de concentração: Estatística e Ex-
perimentação Agronômica

Piracicaba
2023

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Santos, Denize Palmito dos

Delineamentos ótimos para experimentos multi-ambientais de melhora-
mento genético de plantas / Denize Palmito dos Santos. -- versão revisada
de acordo com a resolução CoPGr 6018 de 2011. -- Piracicaba, 2023 .

74 p.

Tese (Doutorado) -- USP / Escola Superior de Agricultura "Luiz de
Queiroz".

1. Análise conjunta 2. Delineamento parcialmente replicado 3. Deline-
amento *grid-plot* 4. Ganho genético 5. Qualidade de seleção 6. Matriz de
relacionamento . I. Título.

DEDICATÓRIA

À minha família.

AGRADECIMENTOS

Agradeço a Deus e a todos que me deram suporte e apoio para que eu conseguisse tornar esse sonho possível, em especial minha mãe, Maria Ednalva Palmito dos Santos, meu irmão Denisar Palmito dos Santos, meu noivo Julio Cezar Souza Vasconcelos, meu amigo Lucas Castro e amigos que fiz ao longo desse caminho.

Agradeço também à minha orientadora, a professora Dra. Renata Alcarde Sermarini, pela imensa ajuda, motivação e entusiasmo durante o meu doutorado, e à equipe do Departamento de Ciências Exatas da ESALQ/USP.

Meu trabalho foi apoiado financeiramente pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

SUMÁRIO

Resumo	6
Abstract	8
1 Introdução	9
Referências	11
2 Delineamentos ótimos em experimentos de melhoramento de plantas: um estudo de simulação comparando delineamentos <i>grid-plot</i> e <i>p-rep</i>	13
Resumo	13
2.1 Introdução	13
2.2 Metodologia	15
2.2.1 Delineamentos gerados	15
2.2.2 Estudo de simulação e medidas de comparação	17
2.2.2.1 Simulação dos dados	17
2.2.2.2 Modelagem estatística e medidas de comparação dos delineamentos	18
2.3 Resultados e discussão	18
2.3.1 Layout dos delineamentos	18
2.3.2 Estudo de simulação	20
2.4 Conclusão	22
Referências	23
3 Delineamentos ótimos incluindo matriz de parentesco: um estudo de simulação em ensaios multi-ambientais de melhoramento genético de trigo	25
Resumo	25
3.1 Introdução	26
3.2 Metodologia	28
3.2.1 Matriz de relacionamento	28
3.2.2 Delineamentos gerados	29
3.2.3 Estudo de simulação e medidas das comparações	31
3.2.3.1 Simulação dos dados	31
3.2.3.2 Modelagem estatística e medidas de comparação dos delineamentos	32
3.2.3.3 Estimação dos parâmetros de variância usando o <i>asreml-R</i>	32
3.3 Resultados e discussão	34
3.4 Conclusão	52
Referências	53
4 Considerações finais	57
Apêndices	59

RESUMO

Delineamentos ótimos para experimentos multi-ambientais de melhoramento genético de plantas

Os programas de melhoramento genético de plantas têm como objetivo selecionar diferentes linhagens com características desejáveis. Ensaio multi-ambientais podem ser utilizados nesses estudos, pois permitem avaliar a adequação de diferentes linhagens em diversas condições ambientais. Dessa forma, métodos estatísticos apropriados são fundamentais para a tomada de decisões com relação à seleção de linhagens sob teste (*test lines*). Os delineamentos ótimos são amplamente considerados em estudos dessa natureza, pois permitem selecionar delineamentos fundamentados em condições não usuais, como por exemplo, dados desbalanceados e a não repetição de tratamentos. Além disso, é possível obter delineamentos do tipo parcialmente replicados (*p*-rep). Diversos estudos mostram que esses delineamentos são mais eficientes do que delineamentos *grid-plot*. Desse modo, o presente trabalho tem como objetivo avaliar o ganho genético e a qualidade da seleção do material genético, por meio de estudos de simulação, para comparar delineamentos *grid-plot* e *p*-rep especialmente otimizados em ensaios multi-ambientais. Considerando algumas medidas para as comparações dos delineamentos, sendo elas: o Ganho Genético Relativo Percebido - RRGG (*Relative Realized Genetic Gain*), a precisão genética, e a probabilidade de seleção. Os delineamentos considerados neste estudo são para uma área experimental de 24 linhas por 10 colunas e 180 *test lines*, em três locais, sendo cada um dividido em dois blocos, compostos por 12 colunas cada, assumindo-se o critério *A* de otimização. Para o delineamento *p*-rep otimizado variou-se as *p*% porcentagens de linhagens repetidas em 11%, 22% e 33% e diferentes número de variedades padrões-*checks* (0, 5, 10, 15, 20), que são variedades já consolidadas. Além disso, também foram considerados diferentes valores para os parâmetros associados às variâncias genéticas e residuais dentro de cada local, e valores distintos para a correlação genética entre as mesmas linhagens aplicadas em diferentes locais, caracterizando diferentes cenários. Modelos lineares mistos incorporando a variação espacial nos erros de parcelas, foram utilizados para gerar os delineamentos, simular e analisar os dados. As análises foram conduzidas de forma conjunta e individual com relação aos locais. No Capítulo 2, investigou-se com relação aos efeitos de *test lines*, se fixos ou aleatórios, no modelo para a obtenção do delineamento. Os resultados indicaram que não foram identificadas diferenças entre as medidas de comparação ao considerar efeitos de *test lines* fixos ou aleatórios no modelo do delineamento. No entanto, houve uma distribuição mais uniforme das *test lines* repetidas e/ou dos tratamentos na área experimental ao assumir efeitos fixos para as *test lines*. Estendendo o estudo, o Capítulo 3 teve como objetivo comparar delineamentos gerados considerando a suposição de dependência ou independência entre os efeitos de linhagens, verificando o ganho genético e a qualidade de seleção do material genético. A matriz de parentesco neste capítulo é proveniente de um programa simulado de melhoramento genético de trigo e as linhagens consideradas foram referentes a fase do teste preliminar de rendimento (*Preliminary Yield Trial* - PYT). Além disso, considerou-se também outros valores para as variâncias genéticas e residual dentro dos locais, esses valores foram assumidos de tal modo a garantir que as razões entre essas duas variâncias fossem as mesmas em todos os locais. Modelos lineares mistos incorporando a variação espacial nos erros de parcelas com matriz de parentesco foram utilizados para realizar as análises. Os resultados indicaram que quando considerou-se valores mais altos para os parâmetros de variâncias genéticas e residual por local, foi possível notar uma melhor aleatorização das variedades em teste na área experimental e melhores resultados para as medidas utilizadas. Em ambos os capítulos, observou-se que os delineamentos *p*-rep especialmente otimizados apresentaram superioridade em relação ao *grid-plot*, e entre as diferentes caracterizações para os delineamentos *p*-rep, os que não apresentaram a presença de variedade padrão, bem como, o *p*-rep com $p = 22\%$ das linhagens duplicadas e menor quantidade de diferentes variedades padrões apresentaram os melhores resultados sendo estes recomendados para realizar estudos dessa natureza. De modo geral, os resultados apresentados indicaram que, a razão

entre a variância genética e residual e a correlação entre as mesmas linhagens em locais diferentes sobre a qualidade da seleção; quanto maiores forem esses valores, melhores serão os resultados.

Palavras-chave: Análise conjunta, Delineamento parcialmente replicado, Delineamento *grid-plot*, Ganho genético, Matriz de relacionamento, Qualidade de seleção.

ABSTRACT

Optimal designs for multi-environmental plant breeding experiments

Plant genetic improvement programs aim to select different lineages with desirable characteristics. Multi-environment tests can be used in these studies, as they allow evaluating the suitability of different strains in different environmental conditions. Thus, appropriate statistical methods are fundamental for making decisions regarding the selection of lines under test (test lines). Optimal designs are widely considered in studies of this nature, as they allow selecting designs based on unusual conditions, such as unbalanced data and non-repetition of treatments. Furthermore, it is possible to obtain partially replicated designs (p -rep). Several studies show that these designs are more efficient than grid-plot designs. Thus, the present work aims to evaluate the genetic gain and the quality of genetic material selection, through simulation studies, to compare spatially optimized grid-plot and p -rep designs in multi-environmental. Considering some measures for the design comparisons, namely: the Perceived Relative Genetic Gain-RRGG (Relative Realized Genetic Gain), the genetic accuracy, and the selection probability. The designs considered in this study are for an experimental area of 24 lines by 10 columns and 180 test lines, two blocks and three locations, assuming the A optimization criterion. For the optimized p -rep design, the $p\%$ percentages of repeated lines were varied by 11%, 22% and 33% and different number of standard-checks varieties (0, 5, 10, 15, 20), which are already consolidated varieties. In addition, different values were also considered for the parameters associated with genetic and residual variances within each location, and different values for the genetic correlation between the same strains applied in different locations, characterizing different scenarios. Mixed linear models incorporating spatial variation in plot errors were used to generate the designs, simulate and analyze the data. The analyzes were conducted jointly and individually with respect to the locations. In Chapter 2, we investigated the effects of test lines, whether fixed or random, on the model to obtain the design. The results indicated that no differences were identified between the comparison measures when considering fixed or random effects of test lines in the design model. However, there was a more uniform distribution of repeated test lines and/or treatments in the experimental area when assuming random effects for the test lines. Extending the study, Chapter 3 aimed to compare the designs generated considering the assumption of dependence or independence between the effects of lineages, verifying the genetic gain and the quality of selection of the genetic material. The parentage matrix in this chapter comes from a simulated wheat genetic improvement program and the lines considered were related to the preliminary yield test phase (Preliminary Yield Trial - PYT). In addition, other values were also considered for the genetic and residual variances within the locations, these values were assumed in such a way as to guarantee that the ratios between these two variances were the same in all locations. Mixed linear models incorporating spatial variation in plot errors with parentage matrix was used to perform the analyses. The results indicated that when higher values were considered for the parameters of genetic and residual variances per location, it was possible to notice a better randomization of the test varieties in the experimental area and better results for the measures used. In both chapters, it was observed that the specially optimized p -rep designs showed superiority in relation to the grid-plot, and among the different characterizations for the p -rep designs, those that did not present the presence of a standard variety, as well as the p -rep with $p = 22\%$ of the duplicated lines and a smaller amount of different standard varieties showed the best results, which are recommended for studies of this nature. In general, the results presented indicated that the ratio between the genetic and residual variance and the correlation between the same strains in different locations on the quality of selection; the higher these values are, the better the results.

Keywords: Joint analysis, Partially replicated design, Grid-plot design; Genetic gain, relationship matrix, Selection quality.

1 INTRODUÇÃO

Os programas de melhoramento genético de plantas têm como objetivo selecionar características desejáveis que serão perpetuadas a uma espécie vegetal, obtendo assim, variedades superiores. Essas apresentam um bom desempenho, melhorando fatores relacionados ao aumento da produtividade, resistência às pragas e doenças, resistência às condições adversas do clima e do solo, melhores arquiteturas das plantas, dentre outros. No entanto, a expressão observável dessas novas linhagens ou genótipos, ou seja, os fenótipos, são influenciados por diversos fatores, entre eles destaca-se o ambiente em que essas novas variedades estão sendo cultivadas (Amabile et al., 2018). Nessa perspectiva, os ensaios multi-ambientais são frequentemente utilizados em estudos dessa natureza, pois permitem avaliar diferentes linhagens em diversas condições ambientais.

Em muitos estudos de melhoramento de plantas, é violado um dos princípios básicos apresentados por Fisher et al. (1937), que é a repetição, o qual possibilita medir a variabilidade não controlada (Federer, 1956a). Dessa forma, nas fases iniciais do melhoramento, há uma escassez relacionada à disponibilidade de material genético, devido ao grande número de linhagens que precisam ser testadas após a hibridação. Essas limitações podem estar relacionadas aos recursos financeiros ou áreas experimentais disponíveis para realizar os testes. Um outro fator que também deve ser observado, além da heterogeneidade do ambiente, são os efeitos de interação entre variedades e o ambiente. Para contornar essas e outras dificuldades alguns delineamentos surgiram durante o século XX. Entre eles podem-se destacar os delineamentos aumentados.

Um delineamento aumentado consiste de duas classes de tratamentos, os comuns (ou *checks*), que correspondem às variedades padrões e as linhagens sob teste (*test lines*). Assim, um delineamento base é utilizado para os tratamentos comuns, podendo ser em blocos completos ou incompletos, delineamento em linhas e colunas, entre outros, e tal delineamento é aumentado com as *test lines* (Federer, 1961). Em outras palavras, seguindo o delineamento proposto, as *checks* são repetidas várias vezes, e as *test lines* são incluídas no experimento apenas uma vez. Entre os principais estudos que foram responsáveis pela introdução desse tipo de delineamento encontram-se Federer (1956a), Federer (1961), Searle et al. (1965), Federer et al. (1970), Federer e Raghavarao (1975), Lin e Poushinsky (1983) e Lin et al. (1983). Federer e Crossa (2012b) mencionam que os delineamentos aumentados são eficientes no controle da variabilidade e na avaliação dos efeitos das linhagens em teste de maneira prática e eficaz.

Um outro tipo de delineamento frequentemente utilizado são os delineamentos não replicados (Kempton, 1984) que foram posteriormente denominados por Cullis et al. (2006) de *grid-plot*. Assim como os delineamentos aumentados, o *grid-plot* também foi utilizado para controlar a variabilidade ambiental em ensaios de campo, pois as parcelas com variedades padrões são distribuídas entre parcelas com as *test lines* não repetidas, de modo que, o número de parcelas do experimento é mantido, e a homogeneidade é alcançada, subtraindo-se um “índice de fertilidade” com base nos rendimentos das variedades padrões.

Para Santos et al. (2002), os delineamentos aumentados alocam um número grande de parcelas para as *checks*, que geralmente, são variedades bem estabelecidas e que, normalmente, não são de interesse do pesquisador para a seleção. Cullis et al. (2006) propuseram substituir algumas parcelas ocupadas por variedades padrões por *test lines* em um *grid-plot*, eles designaram esse delineamento como parcialmente replicado (*p*-rep).

O delineamento *p*-rep surgiu como uma alternativa ao delineamento *grid-plot*. Nesse tipo de delineamento *p*% é a porcentagem de *test lines* repetidas no experimento. Por meio de estudo de simulação, Cullis et al. (2006) mostraram que para um tamanho de ensaio fixo, os delineamentos *p*-rep resultaram em ganhos genéticos maiores do que os delineamentos de *grid-plot*. Stefanova et al. (2009), Moehring et al. (2014), Santos (2017) e Goes (2020) também mostraram a superioridade do delineamento *p*-rep em comparação ao delineamento *grid-plot* em ensaios individuais, esses autores consideraram diferentes

modelos lineares mistos, diferentes cultivares, diferentes critérios de otimização e outros parâmetros de variâncias que não foram considerados por Cullis et al. (2006). Em contrapartida, outros estudos, também, apresentaram vantagens do delineamento p -rep em ensaios multi-ambientais (Beeck et al., 2010; Moehring et al., 2014). Sermarini et al. (2020) avaliaram diferentes tamanhos para os delineamentos p -rep, nas condições consideradas pelos autores, os delineamentos foram assumidos como ensaios grandes quando compostos por 400 *test lines* em 480 parcelas e em experimentos pequenos quando são avaliadas 100 *test lines* em 120 parcelas. Em ambos os casos foi assumido que $p = 20\%$ de *test lines* repetidas, em diferentes condições experimentais, e concluíram que em experimentos grandes o ganho genético não é impactado ao considerar diferentes cenários, mas, em experimentos pequenos o RGG (*realized genetic gain*) é afetado.

Os delineamentos utilizados nesses estudos podem ser classificados em duas categorias. Na primeira categoria, destacam-se os delineamentos clássicos sendo eles: os delineamentos em blocos completos ou incompletos, delineamentos de linhas e colunas e os α -*designs* (Cullis et al., 2020). Na segunda categoria, destacam-se os delineamentos ótimos que são uma classe de delineamentos experimentais, que são ótimos no que diz respeito ao critério de busca e o modelo utilizado para o delineamento adotado (Shah e Sinha, 2012; Cullis et al., 2020). Um critério de otimização é geralmente definido como uma função da matriz de informações de *Fisher*, que está relacionada com o modelo do delineamento. Desta forma, no contexto de modelos lineares, maximizar a informação de *Fisher* corresponde a minimizar a variância de um estimador (Dawid e Sebastiani, 1999).

Existem vários critérios de otimização, e a escolha por um deles dependerá do interesse do pesquisador. Para exemplificar, o critério A, minimiza a soma das variâncias das estimativas dos tratamentos. O critério C, por sua vez, minimiza a variância de um melhor estimador linear não tendencioso de uma combinação linear predeterminada dos parâmetros do modelo. O critério D, maximiza o determinante da matriz de informação *Fisher*. As medidas utilizadas por um critério ótimo para busca do delineamento podem ser consideradas para compará-los, com a utilização do critério A, por exemplo, o delineamento selecionado apresenta uma variância mínima de erro de predição dos efeitos das *test lines*. Desta forma, o ganho genético esperado é maximizado quando a média das variâncias do erro de previsão (*average variance of pairwise differences* - AVPD) de todos os contrastes elementares entre as *test lines* é minimizada.

Por meio dos delineamentos ótimos ou quase ótimos é possível obter um delineamento p -rep, visto que estes não são balanceados e é necessário considerar valores para os parâmetros de interesse. A obtenção dos delineamentos ótimos depende do critério de busca e de um modelo pré-especificado (Shah e Sinha, 2012). Os modelos lineares mistos são preferíveis na busca desses delineamentos pois são flexíveis e permitem a incorporação de termos em sua estrutura, como por exemplo, a inclusão da matriz de parentesco que possibilita uma aproximação mais realista frequentemente encontrada nos experimentos de campo (Cullis et al., 2020). Esses modelos, também, são preferíveis para a análise dos dados. Van der Werf e De Boer (1990) mencionam que a inclusão da matriz de parentesco na seleção de linhagens pode diminuir o viés evitando estimativas tendenciosas. Piepho et al. (2008) mostraram que os BLUPs (*Best Linear Unbiased Predictions*) obtidos ao incluir matriz de parentesco têm uma boa precisão preditiva quando comparados com os BLUPs sem a informação de parentesco, melhorando a eficiência na seleção do material genético. Dessa forma, considera-se importante incluir essa informação tanto no modelo para obtenção dos delineamentos quanto na análise de seus dados.

Com o propósito de usar os conceitos mencionados anteriormente, o objetivo desta tese é avaliar delineamentos ótimos por meio de estudo de simulação para ensaios multi-ambientais de melhoramento genético de plantas. Neste estudo, são gerados delineamentos do tipo *grid-plot* e p -rep otimizados, em que ambos apresentam uma área experimental composta de 24 linhas por 10 colunas, 180 *test lines* e dois blocos avaliados em três ambientes (locais). No Capítulo 2, os delineamentos p -rep e os delineamentos *grid-plot* são comparados considerando o ganho e a qualidade da seleção do material genético. Esses

delineamentos são gerados considerando efeitos de tratamentos fixos ou aleatórios, diferentes valores de parâmetros associados às variâncias genéticas e residuais dentro de cada ambiente, diferentes valores para a correlação genética para os efeitos das mesmas *test lines* em locais distintos e porcentagens diferentes ($p\%$) de *test lines* repetidas e números de *checks*. As análises conjuntas e individuais são conduzidas usando modelos lineares mistos que incorporam a variação espacial nos erros de parcela. No Capítulo 3, os delineamentos são gerados considerando duas suposições para os tratamentos, isto é, considerando a relação de dependência ou independência dos tratamentos e dois cenários para as variâncias genética e residual. Assim como no Capítulo 2, também, são considerados diferentes valores para a correlação genética para os efeitos das mesmas *test lines* em locais distintos e porcentagens diferentes ($p\%$) de *test lines* repetidas e números de *checks*. Modelos lineares mistos com e sem matriz de parentesco são utilizados para gerar os delineamentos. As análises conjuntas e individuais são conduzidas usando modelos lineares mistos, assumindo-se o relacionamento entre as *test lines*, ou seja, a matriz de parentesco, e variação espacial nos erros de parcelas. A conclusão é apresentada no Capítulo 4.

Referências

- Amabile, R., VILELA, M., e Peixoto, J. (2018). Melhoramento de plantas: variabilidade genética, ferramentas e mercado. *Embrapa Cerrados-Livro técnico (INFOTECA-E)*.
- Beeck, C., Cowling, W., Smith, A., e Cullis, B. R. (2010). Analysis of yield and oil from a series of canola breeding trials. part i. fitting factor analytic mixed models with pedigree information. *Genome*, 53(11):992–1001.
- Cullis, B. R., Smith, A. B., Cocks, N. A., e Butler, D. G. (2020). The design of early-stage plant breeding trials using genetic relatedness. *Journal of Agricultural, Biological and Environmental Statistics*, 25(4):553–578.
- Cullis, B. R., Smith, A. B., e Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *Journal of agricultural, biological, and environmental statistics*, 11(4):381.
- Dawid, A. P. e Sebastiani, P. (1999). Coherent dispersion criteria for optimal experimental design. *Annals of Statistics*, pages 65–81.
- Federer, W. (1956a). Augmented designs, “hawaiian planters”. *Record*, 55:191–208.
- Federer, W. T. (1961). Augmented designs with one-way elimination of heterogeneity. *Biometrics*, 17(3):447–473.
- Federer, W. T. e Crossa, J. (2012b). I. 4 screening experimental designs for quantitative trait loci, association mapping, genotype-by environment interaction, and other investigations. *Frontiers in Physiology*, 3.
- Federer, W. T. et al. (1970). Construction of classes of experimental designs using transversals in latin squares and hedayat’s sum composition method.
- Federer, W. T. e Raghavarao, D. (1975). On augmented designs. *Biometrics*, pages 29–35.
- Fisher, R. A. et al. (1937). The design of experiments. *The design of experiments.*, (2nd Ed).
- Goes, A. d. L. (2020). *Delineamentos ótimos para experimentos com cana-de-açúcar*. PhD thesis, Universidade de São Paulo.

- Kempton, R. (1984). The design and analysis of unreplicated field trials. *Vortraege fuer Pflanzenzuechtung (Germany)*.
- Lin, C., Poushinsky, G., e Jui, P. (1983). Simulation study of three adjustment methods for the modified augmented design and comparison with the balanced lattice square design. *The Journal of Agricultural Science*, 100(3):527–534.
- Lin, C.-S. e Poushinsky, G. (1983). A modified augmented design for an early stage of plant selection involving a large number of test lines without replication. *Biometrics*, pages 553–561.
- Moehring, J., Williams, E. R., e Piepho, H.-P. (2014). Efficiency of augmented p-rep designs in multi-environmental trials. *Theoretical and applied genetics*, 127(5):1049–1060.
- Piepho, H., Möhring, J., Melchinger, A., e Büchse, A. (2008). Blup for phenotypic selection in plant breeding and variety testing. *Euphytica*, 161(1-2):209–228.
- Santos, A. d. (2017). *Design and analysis of sugarcane breeding experiments: a case study*. PhD thesis, Universidade de São Paulo.
- Santos, A. H., Bearzoti, E., Ferreira, D. F., e Silva Filho, J. L. d. (2002). Simulation of mixed models in augmented block design. *Scientia Agricola*, 59(3):483–489.
- Searle, S. et al. (1965). Computing formulae for analyzing augmented randomized complete block designs.
- Sermarini, R. A., Brien, C., Demétrio, C. G. B., e dos Santos, A. (2020). Impact on genetic gain from using misspecified statistical models in generating p-rep designs for early generation plant-breeding experiments. *Crop Science*, 60(6):3083–3095.
- Shah, K. R. e Sinha, B. (2012). *Theory of optimal designs*, volume 54. Springer Science & Business Media.
- Stefanova, K. T., Smith, A. B., e Cullis, B. R. (2009). Enhanced diagnostics for the spatial analysis of field trials. *Journal of agricultural, biological, and environmental statistics*, 14(4):392.
- Van der Werf, J. e De Boer, I. (1990). Estimation of additive genetic variance when base populations are selected. *Journal of Animal Science*, 68(10):3124–3132.

2 DELINEAMENTOS ÓTIMOS EM EXPERIMENTOS DE MELHORAMENTO DE PLANTAS: UM ESTUDO DE SIMULAÇÃO COMPARANDO DELINEAMENTOS *GRID-PLOT* E *P-REP*

Resumo

No estudo do melhoramento genético de plantas, é comum a utilização de ensaios multi-ambientais que visam avaliar e selecionar diferentes linhagens em diferentes condições ambientais. Outro objetivo é medir as principais características quantitativas/qualitativas, como rendimento de grãos, para selecionar as melhores linhagens sob teste (*test lines*) ou a arquitetura das plantas. Para atingir esses e outros objetivos, um bom delineamento experimental e uma boa análise estatística são essenciais. Nesta perspectiva, o presente trabalho usa dados simulados de melhoramento genético de cana-de-açúcar para comparar delineamentos especialmente otimizados do tipo *p-rep* e *grid-plot* em ensaios multi-ambientais considerando o ganho genético e a qualidade da seleção do material. Para o *p-rep*, variaram-se as porcentagens de repetições das *test lines* e o número de variedades padrões (*checks*). As análises conjuntas e individuais foram conduzidas usando modelos lineares mistos que incorporam a variação espacial nos erros de parcela. Os resultados indicaram que os delineamentos *p-rep* apresentaram superioridade em relação ao delineamento *grid-plot* nas medidas de Ganho Genético Relativo Percebido (RRGG), precisão genética e probabilidade de seleção. Ao comparar os diferentes delineamentos *p-rep*, observou-se que aqueles sem a presença de *checks* apresentaram os melhores resultados em relação ao ganho genético. Além disso, observa-se que para cada conjunto de parâmetros, em que se variou apenas os valores relacionados a correlação genética, que mede a correlação entre as mesmas *test lines* aplicadas em diferentes locais quanto maior a correlação genética, melhor o ganho e a qualidade da seleção. Em todos os locais considerados, observou-se maior precisão nas estimativas dos componentes de variância quando obtidos por meio de análise conjunta, conforme esperado.

Palavras-chave: Análise individual e conjunta; Diferentes porcentagens de *test lines*; Ganho genético; Qualidade da seleção; Quantidades diferentes de *checks*.

2.1 Introdução

Nos programas de melhoramento de plantas, em específico o melhoramento de cana-de-açúcar, um dos objetivos é selecionar novas linhagens que apresentem desempenho superior às já conhecidas e, em muitos casos, possam ser utilizadas como futuros pais nos cruzamentos ou lançadas comercialmente. O Brasil é o maior produtor mundial de cana-de-açúcar, seguido por Índia, China, Tailândia, Paquistão e México (Cursi et al., 2022). Responsável por mais de 70% do total de açúcar produzido no mundo, também é utilizada como cultura energética, na alimentação animal, na produção de bioetanol, bebidas e outros. Dessa forma, variedades melhoradas de cana-de-açúcar têm desempenhado um papel fundamental no desenvolvimento das indústrias açucareiras em todo o mundo (Yadav et al., 2020).

Durante as fases de um programa de melhoramento de cana-de-açúcar é possível selecionar as principais características qualitativas/quantitativas, como: maior resistência às pragas e doenças, maior tolerância ao estresse hídrico, melhor adaptação à colheita mecanizada, desenvolver variedades mais produtivas e, então, estimar o potencial genético durante as etapas do processo (Cullis et al., 2006). Um aspecto muito importante destes estudos é a avaliação da interação linhagem-ambiente por meio de ensaios multi-ambientes que permitem identificar aqueles com adaptações estáveis ou específicas (Jarquin et al., 2020), tendo em conta a variação edafoclimática. Para obter uma boa qualidade e precisão na seleção das linhagens a serem testadas, um planejamento experimental adequado e uma análise estatística

correta são essenciais.

Modelos lineares mistos têm sido utilizados para a obtenção de delineamentos ótimos que são amplamente usados em ensaios de campos (Gilmour et al., 1997; Oakey et al., 2006, 2007; Moehring et al., 2014; Hunt et al., 2020; Cullis et al., 2020; Sermarini et al., 2020), uma vez que incorporam efeitos fixos e aleatórios, permitindo modelar a dependência espacial, a inclusão da matriz de parentesco e outros efeitos de controle local. Os modelos lineares mistos são flexíveis e, frequentemente, usados para a análise de dados de experimentos de melhoramento de plantas (Cullis et al., 2020; Sermarini et al., 2020; Piepho et al., 2020).

Uma vez que os experimentos em fase inicial dos programas de melhoramento de plantas apresentam, em geral, limitações como grande número de linhagens sob teste (*test lines*), restrições relacionadas ao material genético disponível e à área experimental, nem todas as *test lines* podem ser repetidas em um experimento, afetando as estimativas dos parâmetros e suas precisões. Para tentar atenuar este tipo de problemas, foram propostos os delineamentos *grid-plot* e aumentados (Federer, 1956b; Kempton, 1984; Federer e Crossa, 2012a; Moehring et al., 2014). Os delineamentos aumentados são um tipo de delineamento experimental formado a partir de duas classes de tratamentos, os comuns, que são as variedades padrões ou *checks* e as novas linhagens que serão testadas, que são as *test lines*. Desta forma, um delineamento base é utilizado para acomodar as *checks* e esse delineamento é então aumentado com parcelas contendo *test lines*.

Os *grid-plot* são delineamentos em linha-coluna nos quais as *test lines* não são repetidas, isto é, as *test lines* aparecem apenas uma vez em cada experimento, e as variedades conhecidas (*checks*) são repetidas, permitindo a estimativa da variabilidade ambiental ou não genética. As variedades conhecidas podem ocupar as unidades experimentais de acordo com uma alocação sistemática ou aleatória. Uma alternativa ao delineamento *grid-plot* é o delineamento parcialmente replicado (*p-rep*), também, em linha-coluna, proposto por Cullis et al. (2006), em que $p\%$ das *test lines* são repetidas substituindo parte ou todas as parcelas com as *checks* no experimento, resultando em um delineamento desbalanceado. Os delineamentos *p-rep* foram, inicialmente, propostos como delineamentos resolvíveis e espacialmente otimizados.

Delineamentos que apresentam melhores resultados em termos de ganho genético podem ser obtidos usando a teoria de delineamento ótimo, que requer um modelo apropriado para a sua obtenção. Segundo Butler e Cullis (2019) o modelo do delineamento deve ser o mais próximo possível do utilizado para realizar a análise, o que justifica a preferência pelo uso de modelos lineares mistos. Além disso, para encontrar um delineamento ótimo é necessário um critério de busca (Shah e Sinha, 2012; Stefanova et al., 2009; Cullis et al., 2020), que pode ser relacionado à matriz de informações do modelo. Por exemplo, o critério de otimização A , seleciona um delineamento com variância mínima de erro de predição dos efeitos das *test lines*. Desta forma, o ganho genético esperado é maximizado quando a média das variâncias do erro de previsão (*average variance of pairwise differences* - AVPD) de todos os contrastes elementares entre as *test lines* é minimizada.

Para ensaios simulados para um único local, diferentes modelos lineares mistos, critérios de busca e tamanho experimental, foi demonstrado que os delineamentos *p-rep* são superiores ao *grid-plot* em termos de ganho genético (Cullis et al., 2006; Santos, 2017; Goes, 2020; Sermarini et al., 2020). Considerando ensaios de uniformidade, Clarke e Stefanova (2011), também, concluíram pela superioridade dos delineamentos *p-rep*, quando comparados aos *grid-plot*, usando como medida de comparação o *SE ratio*, que corresponde à razão entre o erro-padrão da comparação entre *test lines* e *checks*, calculada a partir do referido delineamento e também a partir de um delineamento inteiramente casualizado.

No caso de ensaios multi-ambientais, os delineamentos *p-rep* aumentados foram propostos por Williams et al. (2011), como uma extensão de Cullis et al. (2006). Os autores adaptaram o α -*design* e produziram delineamentos *p-rep* eficientes, para os quais as *test lines* que são repetidas em um ambiente

não são repetidas em outro. Para comparar delineamentos p -rep aumentados, delineamentos blocos aumentados, *grid-plot* e delineamentos totalmente replicados, Moehring et al. (2014) realizaram estudos de simulação considerando a fase inicial de programas de melhoramento para triticales e milho. Os autores mostraram que os delineamentos *grid-plot* apresentaram maior eficiência ao avaliar ensaios multi-ambientais. No entanto, se a repetição for necessária, os delineamentos p -rep aumentados superam os delineamentos aumentados e o p -rep, e podem apresentar mais vantagens, se houver interesse na análise de dados individuais (*single-site*) ou uma análise por etapa considerando os ambientes de forma individual. Ao comparar delineamentos totalmente replicados e delineamentos p -rep aumentados, Paget et al. (2017) concluíram com base em testes empíricos e simulações, que delineamentos p -rep aumentam a taxa de ganho genético para produção de tubérculos em um programa de melhoramento de batata. Eles destacaram as maiores vantagens de usar ensaios multi-ambientais.

Embora existam alguns trabalhos comparando análises individuais com análises conjuntas para ensaios multi-ambientais (Santos, 2017; Tanaka, 2020), as vantagens de lidar com essas análises de dados de delineamentos espacialmente otimizados em ensaios multi-ambientais ainda precisam ser investigadas. Com essa perspectiva, usando um estudo de simulação, o objetivo deste trabalho é comparar delineamentos espacialmente otimizados utilizando o critério A com relação aos (i) efeitos de *test lines*, sejam eles fixos ou aleatórios, (ii) diferentes valores de parâmetros associados às variâncias genética e residual dentro de cada ambiente, (iii) valores diferentes para a correlação genética para os efeitos das mesmas *test lines* em locais diferentes e (iv) porcentagens diferentes ($p\%$) de *test lines* repetidas e números de *checks*. Os dados foram analisados considerando as análises individuais e conjuntas. A Seção 2.2 apresenta as especificações do modelo para obtenção dos delineamentos, para a simulação e análise dos dados e as medidas utilizadas para comparar os delineamentos. Os principais resultados são apresentados na Seção 2.3, enquanto as conclusões estão na Seção 2.4.

2.2 Metodologia

2.2.1 Delineamentos gerados

Os delineamentos *grid-plot* e p -rep foram gerados considerando três locais (n_s), com uma área experimental de 24 linhas (n_r) por 10 colunas por local (n_c), sendo a área dividida em dois blocos (cada bloco tem 5 colunas) e 180 *test lines* (m). Os delineamentos ótimos foram gerados assumindo quatro valores para $p\%$ (0%, 11%, 22% e 33%) das *test lines* a serem repetidas em cada local e diferentes números de *checks* (0, 1, 2, 4), conforme apresentado na Tabela 2.1. Assim como assumido por Williams et al. (2011), as *test lines* que foram repetidas em um local não foram repetidas em outro. Além disso, cada bloco recebeu apenas uma repetição das *test lines* repetidas, conforme considerado por Cullis et al. (2006) e Sermarini et al. (2020), o que caracteriza um delineamento resolvível. Dessa forma, o número máximo de observações para cada *test line* foi quatro, considerando os três locais. Os tratamentos foram compostos pelas *test lines*, rotuladas de 1 a 180, e as *checks*, rotuladas de A a D, que foram distribuídas em quantidades iguais (quando possível) dentro de cada bloco.

Tabela 2.1. Caracterização dos delineamentos de acordo com o número de *test lines* repetidas e *checks*.

	$p\%$	Número de <i>test lines</i>		Número de <i>checks</i>	
		Não repetidas	Repetidas	Quantidade	Número de repetição
Delineamento 1	0	180	0	4	15
Delineamento 2	11	160	20	4	10
Delineamento 3	22	140	40	4	5
Delineamento 4	22	140	40	2	10
Delineamento 5	22	140	40	1	20
Delineamento 6	33	120	60	0	0

O modelo do delineamento considerou conjuntamente os experimentos nos três locais, assumindo os efeitos de *test lines* aleatórios e incorporando termos para avaliar variações locais, globais e externas (Gilmour et al., 1997), é definido por:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_o\mathbf{u}_o + \mathbf{Z}_g\mathbf{u}_g + \boldsymbol{\epsilon} \quad (2.1)$$

em que $\mathbf{y}_{(n \times 1)}$ é o vetor de observações, $n = n_s \times n_r \times n_c$; $\boldsymbol{\beta}_{(q \times 1)}$ é o vetor de q efeitos fixos, incluindo efeitos de ambiente (local), efeitos de blocos e *checks* com matriz de delineamento $\mathbf{X}_{(n \times q)}$; $\mathbf{u}_o_{(b \times 1)}$ é o vetor de b efeitos aleatórios não genéticos (linhas e colunas dentro de cada local) com matriz de delineamento $\mathbf{Z}_o_{(n \times b)}$, $b = n_s \times (n_r + n_c)$; $\mathbf{u}_g_{(mn_s \times 1)}$ é o vetor de efeitos genéticos aleatórios dentro de cada local com matriz de delineamento $\mathbf{Z}_g_{(n \times mn_s)}$; $\boldsymbol{\epsilon}_{(n \times 1)}$ é o vetor de erros aleatórios. Neste estudo, considerou-se que o vetor dos efeitos genéticos \mathbf{u}_g é composto apenas pelos efeitos genéticos aditivos, isto é, não se considerou os efeitos de dominância e os residuais não aditivos. Assumiu-se uma distribuição Gaussiana multivariada para $(\mathbf{u}_o, \mathbf{u}_g, \boldsymbol{\epsilon})$ com vetor de média zero e matriz de variâncias-covariâncias

$$\begin{bmatrix} \mathbf{G}_o(\boldsymbol{\sigma}_o) & 0 & 0 \\ 0 & \mathbf{G}_g(\boldsymbol{\sigma}_{g_s}) & 0 \\ 0 & 0 & \mathbf{R}(\boldsymbol{\phi}) \end{bmatrix},$$

sendo $\mathbf{G}_o(\boldsymbol{\sigma}_o) = \mathbf{I}_{n_s} \otimes (\sigma_r^2 \mathbf{I}_{n_r} \oplus \sigma_c^2 \mathbf{I}_{n_c})$ em que $\boldsymbol{\sigma}_o = (\sigma_r^2, \sigma_c^2)'$ e σ_r^2 e σ_c^2 são as variâncias dos efeitos de linhas e colunas, consideradas iguais em todos os ambientes; assim, assumiu-se que $\sigma_r^2 = 0, 1$, $\sigma_c^2 = 0, 3$, \mathbf{I}_{n_s} , \mathbf{I}_{n_r} e \mathbf{I}_{n_c} são matrizes identidades de ordens n_s , n_r e n_c , respectivamente. Além disso, assumiu-se que

$$\mathbf{G}_g(\boldsymbol{\sigma}_{g_s}) = \begin{bmatrix} \sigma_{g_{s_1}}^2 & \phi_g \sigma_{g_{s_1}} \sigma_{g_{s_2}} & \phi_g \sigma_{g_{s_1}} \sigma_{g_{s_3}} \\ \phi_g \sigma_{g_{s_1}} \sigma_{g_{s_2}} & \sigma_{g_{s_2}}^2 & \phi_g \sigma_{g_{s_2}} \sigma_{g_{s_3}} \\ \phi_g \sigma_{g_{s_1}} \sigma_{g_{s_3}} & \phi_g \sigma_{g_{s_2}} \sigma_{g_{s_3}} & \sigma_{g_{s_3}}^2 \end{bmatrix} \otimes \mathbf{I}_m,$$

em que $\boldsymbol{\sigma}_{g_s} = (\sigma_{g_{s_1}}^2, \sigma_{g_{s_2}}^2, \sigma_{g_{s_3}}^2, \phi_g)'$, σ_{g_s} corresponde à variância genética dentro de cada ambiente e ϕ_g é a correlação genética entre as mesmas linhagens aplicadas em diferentes locais. No modelo para gerar os delineamentos assumiu-se: $\sigma_{g_{s_1}}^2 = 0, 8$, $\sigma_{g_{s_2}}^2 = 1$, $\sigma_{g_{s_3}}^2 = 2$. Três valores diferentes, sendo: $\phi_g = 0, 7$, $\phi_g = 0, 8$ e $\phi_g = 0, 9$, caracterizando a independência entre as linhagens aplicadas no mesmo local, porém com dependência entre os efeitos das mesmas *test lines* aplicadas em diferentes locais. Além disso, os diferentes valores de ϕ_g representam três conjunto de parâmetros (cenários) avaliados de acordo com a influência da correlação genética. \mathbf{I}_m é uma matriz identidade de ordem m . Por fim, $\mathbf{R}(\boldsymbol{\phi}) = \bigoplus_{i=1}^3 \sigma_{s_i}^2 \text{AR1}(\phi_r) \otimes \text{AR1}(\phi_c)$, ou seja, uma soma direta das variâncias residuais dentro de cada local ($\sigma_{s_i}^2$) e dos processos separáveis autorregressivos de primeira ordem, como sugerido por Gilmour et al. (1997). Neste caso, $\boldsymbol{\phi} = (\phi_r, \phi_c)'$, sendo, $\phi_r = \phi_c = 0, 5$ são os parâmetros de correlações residuais entre linhas e entre colunas dentro de cada local. As variâncias residuais dentro de cada local são $\sigma_{s_1}^2 = 1, 0$ e $\sigma_{s_2}^2 = \sigma_{s_3}^2 = 1, 25$. Assim, $\sigma_{g_{s_1}}^2 / \sigma_{s_1}^2$ é 0,8, $\sigma_{g_{s_2}}^2 / \sigma_{s_2}^2$ é 0,8 e $\sigma_{g_{s_3}}^2 / \sigma_{s_3}^2$ é 1,6. Com exceção da correlação genética, todos os valores considerados para os parâmetros de variâncias foram considerados iguais em todos os cenários. Esses valores foram definidos com base no trabalho de Santos (2017); Goes (2020) e Sermarini et al. (2020) que abordaram estudos de simulação com dados de cana-de-açúcar.

Alternativamente, considerou-se *test lines* de efeitos fixos no modelo para gerar o delineamento, para verificar se há mudança nos resultados quando este efeito é inserido de forma fixa no modelo do delineamento. Desta forma, o modelo é definido como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{X}_g\boldsymbol{\beta}_g + \mathbf{Z}_o\mathbf{u}_o + \boldsymbol{\epsilon} \quad (2.2)$$

sendo \mathbf{y} , \mathbf{X} , $\boldsymbol{\beta}$, \mathbf{Z}_o , \mathbf{u}_o e $\boldsymbol{\epsilon}$ são definidos como na Equação (2.1), $\boldsymbol{\beta}_{\mathbf{g}(mn_s \times 1)}$ é o vetor de efeitos genéticos fixos dentro do local com matriz de delineamento $\mathbf{X}_{\mathbf{g}(n \times mn_s)}$. Assumiu-se uma distribuição Gaussiana multivariada para $(\mathbf{u}_o, \boldsymbol{\epsilon})$ com média zero e matriz de variâncias-covariâncias

$$\begin{bmatrix} \mathbf{G}_o(\boldsymbol{\sigma}_o) & 0 \\ 0 & \mathbf{R}(\phi) \end{bmatrix},$$

sendo que $\mathbf{G}_o(\boldsymbol{\sigma}_o)$ e $\mathbf{R}(\phi)$ são conforme definido anteriormente.

Os delineamentos gerados corresponderam a uma combinação de (i) seis caracterizações de delineamento (Tabela 2.1) (ii) suposições sobre os efeitos das *test lines* (aleatórios ou fixos) e (iii) três valores para ϕ_g quando os efeitos das *test lines* foram considerados aleatórios, resultando em um total de 24 delineamentos.

O critério de otimização adotado foi o A , que minimiza a média das variâncias das diferenças pareadas (AVPD) dos efeitos das *test lines*, dado por,

$$\text{AVPD} = \frac{2}{m-1} \left(\text{tr}(\mathbf{C}^{-1}) - \frac{1}{m} \mathbf{1}' \mathbf{C}^{-1} \mathbf{1} \right),$$

em que m é o número de *test lines*, $\text{tr}()$ é o traço da matriz, \mathbf{C} é a matriz de informação das *test lines* e $\mathbf{1}$ é um vetor de uns. De acordo com Hooks et al. (2009), ao considerar os efeitos aleatórios das *test lines*, a matriz \mathbf{C} é definida da seguinte forma:

$$\begin{aligned} \mathbf{C} = & \mathbf{Z}'_{\mathbf{g}}(\mathbf{Z}_o \mathbf{G}_o(\boldsymbol{\sigma}_o) \mathbf{Z}'_o + \mathbf{R}(\phi))^{-1} \mathbf{Z}_{\mathbf{g}} + \mathbf{G}_{\mathbf{g}}^{-1}(\boldsymbol{\sigma}_{g_s}) - \mathbf{Z}'_{\mathbf{g}}(\mathbf{Z}_o \mathbf{G}_o(\boldsymbol{\sigma}_o) \mathbf{Z}'_o + \mathbf{R}(\phi))^{-1} \\ & \mathbf{X}(\mathbf{X}'(\mathbf{Z}_o \mathbf{G}_o(\boldsymbol{\sigma}_o) \mathbf{Z}'_o + \mathbf{R}(\phi))^{-1} \mathbf{X})^{-1} \mathbf{X}'(\mathbf{Z}_o \mathbf{G}_o(\boldsymbol{\sigma}_o) \mathbf{Z}'_o + \mathbf{R}(\phi))^{-1} \mathbf{Z}_{\mathbf{g}}. \end{aligned} \quad (2.3)$$

Ao considerar os efeitos fixos das *test lines*, a matriz \mathbf{C} é semelhante à Equação 2.3, porém $\mathbf{G}'_{\mathbf{g}}(\boldsymbol{\sigma}_{g_s}) = 0$. Foram considerados 20.000 movimentos para a busca do delineamento. Mais detalhes podem ser obtidos consultando Hooks et al. (2009), Atkinson e Donev (1992) e Cullis et al. (2006). Ao considerar quatro variações para a porcentagem $p\%$ de *test lines* repetidas, três valores para os parâmetros de correlação genética entre as mesmas linhagens aplicadas em diferentes locais e duas suposições para os efeitos de *test lines* (fixos/aleatórios), foram obtidas 24 caracterizações para os delineamentos. Os delineamentos foram gerados usando o pacote *owd* (Butler e Cullis, 2019) para o *software* estatístico R.

2.2.2 Estudo de simulação e medidas de comparação

2.2.2.1 Simulação dos dados

Após a obtenção dos delineamentos espacialmente otimizados, conforme apresentado na seção 2.2.1, realizou-se um estudo de simulação para obter o conjunto dos dados que foram utilizados para avaliar a robustez dos delineamentos gerados. Os dados simulados seguiram o modelo linear misto apresentado na Equação 2.1, considerando efeitos fixos para as *checks*, os blocos e os ambientes. No estudo de simulação e posteriormente na análise dos dados os efeitos das *test lines* foram sempre aleatórios. Os mesmos conjuntos de parâmetros usados para gerar os delineamentos foram atribuídos no modelo de simulação de dados. Considerando a distribuição normal padrão, simularam-se (i) 1.000 vetores de tamanho 540 (180×3) para os efeitos genéticos dentro de cada local (\mathbf{u}_g), (ii) 1.000 vetores de tamanho 102 ($(24 + 10) \times 3$) para os efeitos de linhas e colunas (não genéticos) (\mathbf{u}_o) e (iii) 1.000 vetores de tamanho 720 ($24 \times 10 \times 3$) para os efeitos residuais. Para obter os dados simulados para cada cenário de dados, cada um dos 1.000 vetores foi transformado usando a matriz de variâncias para esses respectivos efeitos. Os três grupos de vetores foram, então, combinados para cada um dos 24 caracterizações de delineamentos do tipo *grid-plot* e *p-rep* usando suas matrizes de delineamento.

2.2.2.2 Modelagem estatística e medidas de comparação dos delineamentos

Realizaram-se as análises individuais e conjuntas para os três locais seguindo o modelo utilizado para gerar os dados com o método de estimação da máxima verossimilhança restrita. Para cada modelo ajustado registraram-se as estimativas dos componentes de variância e seus respectivos erros padrão, e os *Empirical Best Linear Unbiased Predictions* (EBLUPs), a fim de calcular o Ganho Genético Relativo Percebido (*Realized Relative Genetic Gain* - RRG) que é obtido a partir do *realized genetic gain* (RGG), como a razão entre a média dos maiores $s\%$ EBLUPs e a média dos maiores $s\%$ verdadeiros efeitos genéticos, em que $s\%$ é a porcentagem de seleção, aqui assumido 10%, um valor comumente adotado pelos programas brasileiros de melhoramento de plantas (Sermarini et al., 2020; Goes, 2020).

Seguindo Sermarini et al. (2020), também foram calculadas medidas secundárias (i) precisão genética que é a correlação entre os verdadeiros efeitos genéticos e seus EBLUPs, (ii) a probabilidade de seleção, em termos das verdadeiramente $s\%$ superiores *test lines*, referido como um sucesso de seleção, (iii) AVPD, calculado a partir das estimativas dos componentes de variância, que permite a comparação entre *test lines* repetidas, entre as não repetidas e entre as repetidas e não repetidas e (iv) número de graus residuais (nedf).

Estudos de simulação e análises de dados foram realizados no software R (R Core Team, 2020) utilizando os pacotes *asreml-R* (Butler, 2021), para ajustar os modelos; e *dae* (Brien, 2021b) para calcular o AVPD.

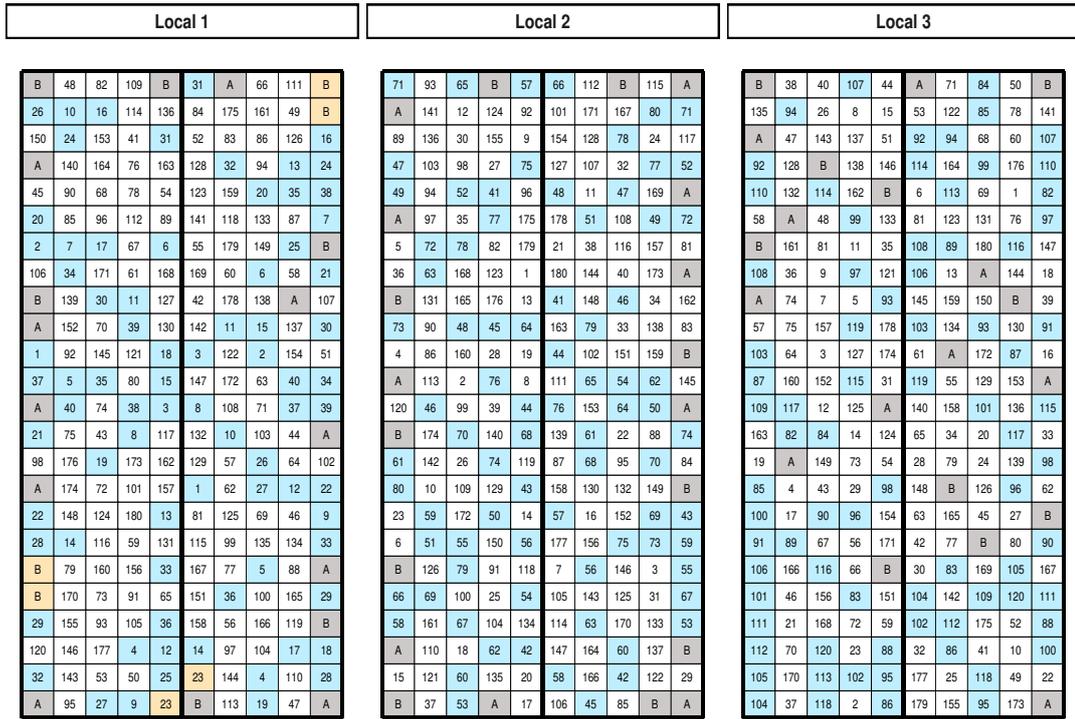
2.3 Resultados e discussão

2.3.1 Layout dos delineamentos

Na Figura 2.1, os layouts para o Delineamento 4 (Tabela 2.1), que são delineamentos p -rep otimizados com $\phi_g = 0,9$, são apresentados como exemplo, na qual, (a) os efeitos das *test lines* são assumidos aleatórios enquanto em (b) são efeitos fixos. Isso mostra que as *test lines* e as *checks* são, em geral, distribuídas aleatoriamente ao longo da área experimental e não foi observado nenhum arranjo sistemático das *checks*. Embora tenham sido adotados delineamentos resolvíveis, há ocorrências de repetições da mesma *test lines* ocupando parcelas vizinhas (para o limite definido pelos blocos), ocorrendo para não mais do que 3% das *test lines* repetidas por local.

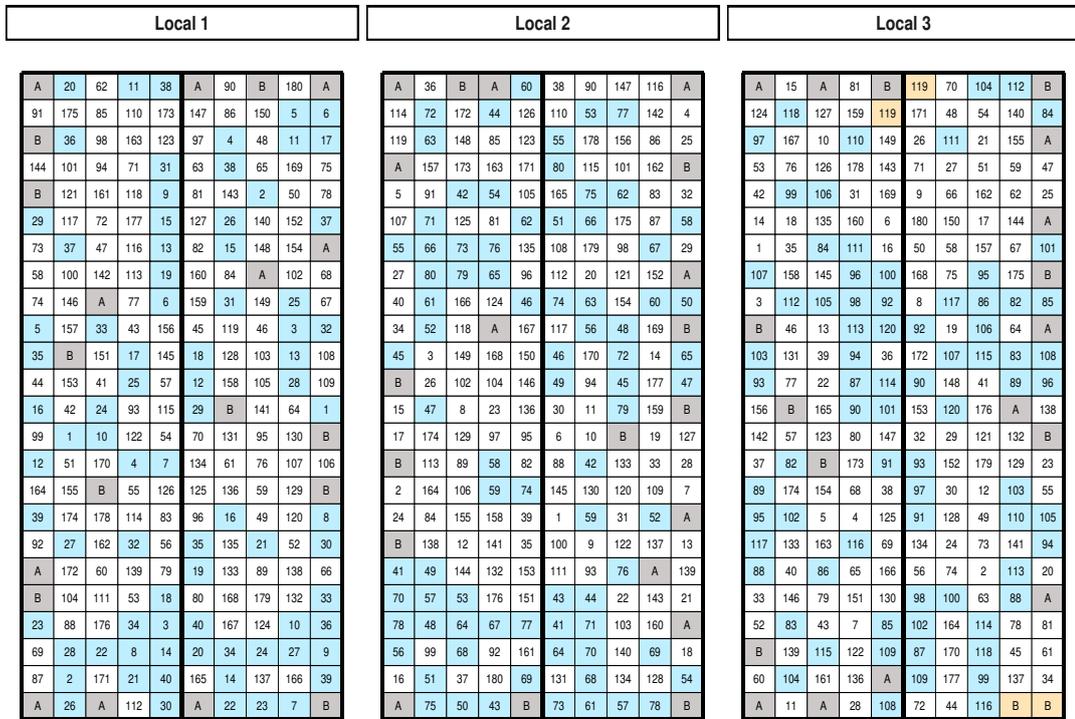
No geral, todos os delineamentos tiveram um comportamento semelhante ao Delineamento 4. No entanto, destacou-se: (i) ao assumir os efeitos fixos das *test lines*, elas apresentaram maior incidência de *test lines* repetidas próximas às bordas da área experimental; (ii) os maiores percentuais de *checks* ocupando parcelas vizinhas foram observados para os Delineamentos 1 e 5, com 8,89% e 6,66%, respectivamente. (iii) quanto maior o valor de $p\%$ maior é o número de *test lines* repetidas nas parcelas da borda da área experimental, (iv) ao assumir efeitos fixos de *test lines*, a quantidade de *test lines* repetidas em parcelas vizinhas tende a aumentar conforme ϕ_g aumenta.

(a)



Checks não vizinhas
 Test lines e/ou checks vizinhas
 Test lines repetidas não vizinhas
 Test lines não repetidas

(b)



Checks não vizinhas
 Test lines e/ou checks vizinhas
 Test lines repetidas não vizinhas
 Test lines não repetidas

Figura 2.1. Layouts para o Delineamento 4 (Tabela 2.1), assumindo $p\%$ como 22%, 140 *test lines* não repetidas, 40 *test lines* repetidas, duas *checks* repetidas 10 vezes e ϕ_g como 0,9, ao assumir (a) efeitos aleatórios de *test lines* e (b) efeitos fixos de *test lines* no modelo para gerar o delineamento.

2.3.2 Estudo de simulação

A Figura 2.2 resume as medidas de comparação para os delineamentos, que são: o RRG, a precisão genética e o sucesso de seleção, considerando, *test lines* de efeitos aleatórios e fixos no modelo para o delineamento, para as análises individuais por local e a análise conjunta. Existem poucas diferenças entre os delineamentos, sendo as mais perceptíveis aquelas relacionadas à precisão e ao sucesso da seleção. Essas medidas de comparação não são afetadas pela suposição acerca dos efeitos de *test lines* serem considerados fixos ou aleatórios no modelo para o delineamento. No entanto, quanto maior a porcentagem de *test lines* repetidas, maiores são as medidas, como mostram os resultados para o Delineamento 6. Além disso, um número menor de *checks* diferentes tende a melhorar os resultados ao definir $p\%$ como 22% e assumir efeitos de *test lines* aleatórios nos modelos de delineamento, principalmente, para as análises conjuntas.

Adicionalmente, observa-se que quanto maior a razão entre variâncias genéticas e residuais em cada local, melhores são os resultados para as medidas de comparação, o que corrobora com Cullis et al. (2006), Santos (2017), Goes (2020) e Sermarini et al. (2020).

De maneira geral, as análises conjuntas apresentaram um melhor desempenho para as três medidas, em todos os cenários e locais, concordando com os resultados obtidos por Santos (2017) e Tanaka (2020). Observa-se, também, que à medida que aumenta a correlação genética, ϕ_g , entre as mesmas linhagens em diferentes locais, ocorre um aumento nos valores obtidos para as medidas de comparação. A influência da razão entre as variâncias genéticas e residuais nas medidas de comparação diminui à medida que ϕ_g tende a um.

O ganho genético pode ser afetado pelo número de graus de liberdade residuais (*nedf*) (Clarke e Stefanova, 2011). Neste estudo, os delineamentos com o mesmo número de *checks* e diferentes percentuais de *test lines* repetidas, Delineamentos 1 (*grid-plot*), 2 e 3 (*p-rep*), tiveram os menores valores para *nedf*, que são 234 para a análise individual e 702 para a análise conjunta. Apesar das diferenças nas suposições dos delineamentos, o ganho genético é ligeiramente afetado pela porcentagem de *test lines* repetidas. No entanto, o delineamento com *test lines* não repetidas, o Delineamento 1, apresentou os menores resultados em todos os cenários, locais e análises considerados. Os *nedf* para os Delineamentos 4, 5 e 6 (*p-rep*) foram 236, 237 e 238 (para cada análise individual) e 708, 711 e 716 (para análise conjunta), respectivamente. O Delineamento 6, com o maior número de *nedf*, apresentou maiores valores de RRG em todos os cenários e locais, como também notado por Clarke e Stefanova (2011). Ao definir $p\%$ como 22%, Delineamentos 3, 4 e 5, um RRG maior é resultado da suposição de um número menor de *checks*. Os resultados da comparação entre os Delineamentos 3 e 4 variaram conforme o local e o tipo de análise. Esses resultados são consistentes com Paget et al. (2017), Santos (2017) e Goes (2020). Os valores de sucesso de seleção (Figuras 2.2 (e) e (f)) aumentam à medida que ϕ_g aumenta, com os melhores resultados para as análises conjuntas, em que o sucesso de seleção ficou acima de 59% em todos os cenários. Os menores resultados foram observados para o Delineamento 1 no local 1, sendo que o sucesso de seleção foi de 59%, 62% e 66%, respectivamente, para $\phi_g = 0,7$, $\phi_g = 0,8$ e $\phi_g = 0,9$, representando 10, 11 e 12 das 18 verdadeiramente melhores *test lines* selecionadas. Os maiores resultados foram observados para o Delineamento 6 no local 3, em que os sucessos de seleção foram 67%, 68% e 72%, para $\phi_g = 0,7$, $\phi_g = 0,8$ e $\phi_g = 0,9$, representando 12, 12 e 13 das 18 verdadeiramente melhores *test lines* selecionadas. Dessa forma, há um aumento de pelo menos uma *test line* superior selecionada ao comparar o Delineamento 6 com o Delineamento 1, o que é uma pequena melhoria, mas contribui para o processo de seleção genética.

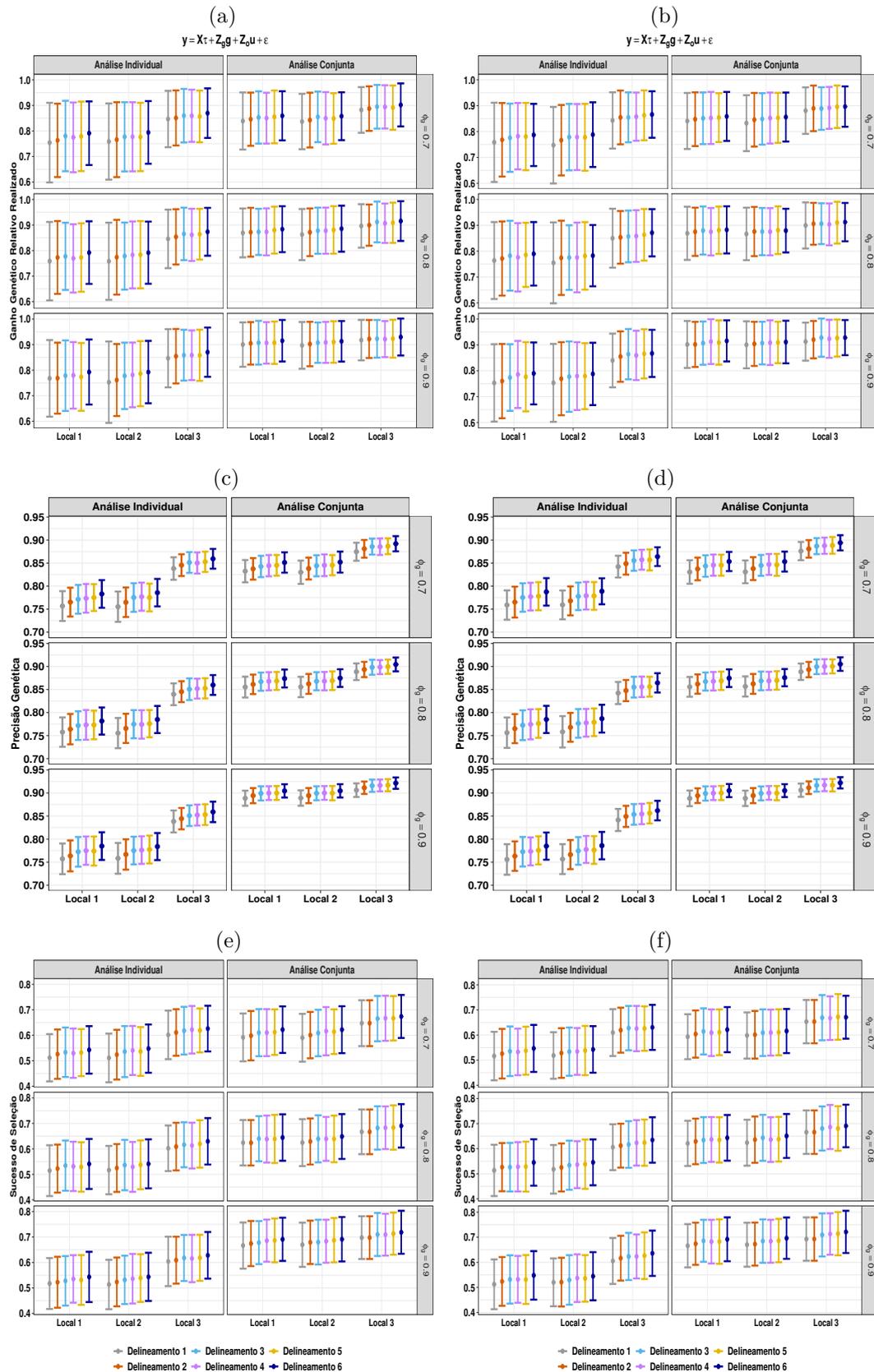


Figura 2.2. Médias para o Ganho Genético Relativo Percebido ((a) e (b)), a precisão genética ((c) e (d)) e o sucesso da seleção ((e) e (f)), ao assumir ((a), (c) e (e)) efeitos aleatórios de *test lines* e ((b), (d) e (f)) efeitos fixos de *test lines* no modelo para gerar o delineamento, para as análises individuais e análises conjuntas dos três locais. As médias são representadas por pontos e as barras de erro representam a média \pm e o desvio padrão.

As *test lines* repetidas comparadas com as não repetidas tiveram os menores valores médios para AVPD (Tabelas A.1 e A.2, localizadas no apêndice) o que é esperado uma vez que o número de *test lines* no primeiro grupo é menor do que no segundo. Além disso, a variação de $p\%$ e a suposição dos efeitos de *test lines*, sejam eles fixos ou aleatórios, pouco afetam os valores dessa medida. No entanto, a correlação genética entre os mesmos genótipos em locais diferentes influencia o AVPD, quanto maior ϕ_g menor foi o seu valor. Também, observou-se um aumento do ganho genético à medida que o AVPD diminui, conforme enfatizado por Cullis et al. (2006).

As suposições de efeitos de *test lines* no modelo de delineamento não afetaram as estimativas dos componentes de variância (Figuras A.6, A.7 e A.8), ainda que em todos os locais e cenários haja uma ligeira subestimação da componente de variância genética e da correlação genética entre as mesmas linhagens em locais diferentes, e uma tendência de subestimar a variância global da coluna e as correlações residuais para linhas e colunas. Quanto maior a razão entre as variâncias genéticas e residuais, maior é o erro padrão. No entanto, essa variabilidade diminui à medida que ϕ_g aumenta. Conforme ressaltado, não há grandes diferenças entre os valores relacionados ao RRG, a precisão genética, ao sucesso da seleção e às estimativas dos componentes de variância quando comparados os delineamentos espacialmente otimizados assumindo efeitos de *test lines* fixas ou aleatórias. Assim como em Sermarini et al. (2020), observou-se que as *test lines* repetidas têm mais aglomeração quando os delineamentos são gerados assumindo efeitos fixos de *test lines*. No entanto, notou-se uma frequência maior de *test lines* ocupando parcelas próximas à borda da área experimental ao gerar delineamentos p -rep assumindo efeitos aleatórios para as *test lines*, para dois tamanhos de área experimental diferentes e $p\%$ como 20%. O delineamento p -rep teve o melhor desempenho nas medidas de comparação, o que foi observado ao avaliar ensaios uni e multi-ambientais, corroborando com Cullis et al. (2006), Clarke e Stefanova (2011), Santos (2017) e Goes (2020). No entanto, se o objetivo é comparar *test lines* com *checks*, é recomendável usar um delineamento p -rep otimizado com repetição de algumas variedades de interesse, em vez do delineamento *grid-plot*. O delineamento que considera apenas uma *check* (Delineamento 5) apresentou os melhores resultados quando comparado a outros delineamentos p -rep que incluíram diferentes números de *checks*.

2.4 Conclusão

Delineamentos *grid-plot* espacialmente otimizados e p -rep espacialmente otimizados foram avaliados para 180 *test lines* em uma área experimental de 24 linhas por 10 colunas e dois blocos para ensaios multi-ambientais, sob a hipótese de efeitos fixos e aleatórios para as *test lines*. Para as análises individuais e conjuntas, os delineamentos p -rep foram ligeiramente superiores aos delineamentos *grid-plot* na seleção do material genético para todos os cenários explorados. Destacou-se a influência da razão entre a variância genética e residual e a correlação entre os efeitos genéticos para as mesmas linhagens em locais diferentes sobre a qualidade da seleção; quanto maiores forem esses valores, melhores serão os resultados. Resultados semelhantes foram encontrados por Cullis et al. (2006), Santos (2017), Goes (2020) e Sermarini et al. (2020). Apesar das pequenas diferenças entre os delineamentos em cada cenário e local, o delineamento p -rep sem *checks* resultou no maior ganho genético, conforme apontado por Santos (2017) e Goes (2020). Assim, é possível concluir que o número de *checks*, bem como seu número de repetições, pode afetar o valor do ganho genético. Portanto, recomenda-se o uso de delineamentos parcialmente replicados sem *checks* para realizar estudos para programas de melhoramento genético de plantas em fases iniciais. Porém, caso seja necessário o uso de *checks*, recomenda-se o delineamento p -rep com pequeno número de repetições para as *checks*, ao invés dos delineamentos *grid-plot*. Finalmente, não foram identificadas diferenças entre as medidas de comparação ao considerar efeitos fixos ou aleatórios no modelo de delineamento. No entanto, houve uma distribuição uniforme das *test lines* repetidas e/ou dos tratamentos na área experimental ao assumir efeitos fixos para as *test lines*. Aqui, foram consideradas

apenas simulações adotando conjuntos de parâmetros cujos valores foram estimados em estudos desenvolvidos por programas brasileiros de melhoramento de cana-de-açúcar. Além disso, uma forte suposição adotada é de independência entre as *test lines*. Os resultados podem diferir quando se consideram áreas experimentais de diferentes tamanhos, ensaios de uniformidade e/ou correlação entre as *test lines*.

Referências

- Atkinson, A. C. e Donev, A. N. (1992). *Optimum experimental designs*.
- Brien, C. (2021b). *dae*: Functions useful in the design and anova of experiments. R package version 3.1-37.
- Butler, D. (2021). *asreml*: Fits the linear mixed model. R package version 4.1.0.149.
- Butler, D. e Cullis, B. (2019). *od*: Generate optimal experimental designs. r package version 2.0. 0.
- Clarke, G. P. Y. e Stefanova, K. T. (2011). Optimal design for early-generation plant-breeding trials with unreplicated or partially replicated test lines. *Australian & New Zealand Journal of Statistics*, 53(4):461–480.
- Cullis, B. R., Smith, A. B., Cocks, N. A., e Butler, D. G. (2020). The design of early-stage plant breeding trials using genetic relatedness. *Journal of Agricultural, Biological and Environmental Statistics*, 25(4):553–578.
- Cullis, B. R., Smith, A. B., e Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *Journal of agricultural, biological, and environmental statistics*, 11(4):381.
- Cursi, D. E., Hoffmann, H. P., Barbosa, G., Bressiani, J. A., Gazaffi, R., Chapola, R. G., Fernandes Junior, A., Balsalobre, T. W. A., Diniz, C. A., Santos, J. M., et al. (2022). History and current status of sugarcane breeding, germplasm development and molecular genetics in brazil. *Sugar Tech*, 24(1):112–133.
- Federer, W. (1956b). Augmented (or hoonuiaku) designs. biometrics unit. Technical report, Cornell Univ. Mimeo. BU-74-M, February.
- Federer, W. T. e Crossa, J. (2012a). I. 4 screening experimental designs for quantitative trait loci, association mapping, genotype-by environment interaction, and other investigations. *Frontiers in Physiology*, 3:156.
- Gilmour, A. R., Cullis, B. R., e Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics*, pages 269–293.
- Goes, A. d. L. (2020). *Delineamentos ótimos para experimentos com cana-de-açúcar*. PhD thesis, Universidade de São Paulo.
- Hooks, T., Marx, D., Kachman, S., e Pedersen, J. (2009). Optimality criteria for models with random effects. *Revista Colombiana de Estadística*, 32(1):17–31.
- Hunt, C. H., Hayes, B. J., Van Eeuwijk, F. A., Mace, E. S., e Jordan, D. R. (2020). Multi-environment analysis of sorghum breeding trials using additive and dominance genomic relationships. *Theoretical and Applied Genetics*, 133(3):1009–1018.

- Jarquín, D., Howard, R., Crossa, J., Beyene, Y., Gowda, M., Martini, J. W., Pazaran, G. C., Burgueño, J., Pacheco, A., Grondona, M., et al. (2020). Genomic prediction enhanced sparse testing for multi-environment trials. *G3: Genes, Genomes, Genetics*, 10(8):2725–2739.
- Kempton, R. (1984). The design and analysis of unreplicated field trials. *Vortraege fuer Pflanzenzuechtung (Germany)*.
- Moehring, J., Williams, E. R., e Piepho, H.-P. (2014). Efficiency of augmented p-rep designs in multi-environmental trials. *Theoretical and applied genetics*, 127(5):1049–1060.
- Oakey, H., Verbyla, A., Pitchford, W., Cullis, B., e Kuchel, H. (2006). Joint modeling of additive and non-additive genetic line effects in single field trials. *Theoretical and Applied Genetics*, 113(5):809–819.
- Oakey, H., Verbyla, A. P., Cullis, B. R., Wei, X., e Pitchford, W. S. (2007). Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. *Theoretical and Applied Genetics*, 114(8):1319–1332.
- Paget, M., Alspach, P., Anderson, J., Genet, R., Braam, W., e Apiolaza, L. (2017). Replicate allocation to improve selection efficiency in the early stages of a potato breeding scheme. *Euphytica*, 213(9):1–15.
- Piepho, H.-P., Vo-Thanh, N., e Tobias, R. (2020). Generating experimental designs for estimation of genetically related treatment effects using sas. *Agronomy Journal*, 112(5):3929–3940.
- Santos, A. d. (2017). *Design and analysis of sugarcane breeding experiments: a case study*. PhD thesis, Universidade de São Paulo.
- Sermarini, R. A., Brien, C., Demétrio, C. G. B., e dos Santos, A. (2020). Impact on genetic gain from using misspecified statistical models in generating p-rep designs for early generation plant-breeding experiments. *Crop Science*, 60(6):3083–3095.
- Shah, K. R. e Sinha, B. (2012). *Theory of optimal designs*, volume 54. Springer Science & Business Media.
- Stefanova, K. T., Smith, A. B., e Cullis, B. R. (2009). Enhanced diagnostics for the spatial analysis of field trials. *Journal of agricultural, biological, and environmental statistics*, 14(4):392.
- Tanaka, E. (2020). Simple outlier detection for a multi-environmental field trial. *Biometrics*, 76(4):1374–1382.
- Williams, E., Piepho, H.-P., e Whitaker, D. (2011). Augmented p-rep designs. *Biometrical Journal*, 53(1):19–27.
- Yadav, S., Jackson, P., Wei, X., Ross, E. M., Aitken, K., Deomano, E., Atkin, F., Hayes, B. J., e Voss-Fels, K. P. (2020). Accelerating genetic gain in sugarcane breeding using genomic selection. *Agronomy*, 10(4):585.

3 DELINEAMENTOS ÓTIMOS INCLUINDO MATRIZ DE PARENTESCO: UM ESTUDO DE SIMULAÇÃO EM ENSAIOS MULTI-AMBIENTAIS DE MELHORAMENTO GENÉTICO DE TRIGO

Resumo

Um programa de melhoramento genético de plantas visa produzir novas cultivares com características de rendimento e qualidade superiores às variedades existentes, como por exemplo, aumento da produtividade, maior resistência às doenças ou variações edafoclimáticas. No melhoramento genético de trigo, por exemplo, a avaliação das novas linhagens pode ocorrer em poucos ambientes, sendo estas não replicadas no ambiente, nos estágios iniciais do ciclo de melhoramento e nos estágios avançados, as réplicas e ambientes tendem a aumentar. Diante disso, ensaios multi-ambientais são amplamente utilizados nesses programas para verificar, entre outros fatores, o efeito da interação entre o genótipo e o ambiente. Os delineamentos ótimos têm sido amplamente utilizados nos ensaios multi-ambientais e são gerados considerando um modelo estatístico e um critério de otimização e com base nessa abordagem é possível obter delineamentos alternativos, como por exemplo, os delineamentos *grid-plot* e *p-rep* otimizados, ambos os delineamentos são desbalanceados e em linhas e colunas, porém no *grid-plot* não há repetição das novas linhagens na área experimental, e no *p-rep* há uma porcentagem $p\%$ de novas linhagens que são duplicadas. Nessa perspectiva, o objetivo do trabalho é avaliar o uso de dados simulados provenientes de delineamentos espacialmente otimizados em ensaios multi-ambientais para comparar delineamentos *grid-plot* e *p-rep* otimizados. Para o delineamento *p-rep* otimizado variou-se as porcentagens $p\%$ de linhagens duplicadas (11%, 22% e 33%) e diferentes número de variedades padrões, denominadas, *checks* (0, 5, 10, 15, 20), que são variedades consolidadas. Em ambos os delineamentos consideraram-se a suposição de dependência ou independência entre os efeitos de linhagens, verificando o ganho genético e a qualidade de seleção do material genético. Os dados e a matriz de relacionamento considerados neste estudo foram provenientes de um estudo de simulação. Foram ajustados modelos lineares mistos considerando-se as análises conjunta e individuais, que incorporam a variação espacial nos erros das parcelas e a informação de *pedigree*. O ganho genético foi avaliado considerando as medidas de *Relative Realized Genetic Gain* (RRGG), da precisão genética e do sucesso de seleção. Os resultados indicaram que os delineamentos *p-rep* apresentaram superioridade em relação ao delineamento *grid-plot* em todos os cenários considerados. Os resultados relevantes indicaram que, nas duas situações consideradas na obtenção do delineamento, quanto maior a correlação entre os mesmos genótipos aplicados em diferentes locais, maior é o ganho genético e melhor é o sucesso de seleção. Além disso, quando são fixados valores iguais para os parâmetros de variâncias dos efeitos não genéticos de linhas e colunas e residuais dentro de cada local e varia-se apenas o parâmetro associado à variância genética, quanto maior é o valor deste parâmetro, melhores são os resultados apresentados. Curiosamente, observou-se que as suposições de dependência/independência para a relação entre as linhagens na obtenção do delineamento pouco afetam nos resultados obtidos. Por fim, notou-se que o delineamento *p-rep* sem a presença de variedade padrão, bem como, o *p-rep* com $p = 22\%$ das linhagens duplicadas e maior quantidade de diferentes variedades padrões apresentaram os melhores resultados sendo estes recomendados para realizar estudos dessa natureza.

Palavras-chave: Análise conjunta e individual; Diferentes valores para os parâmetros de variâncias; Estágio preliminar de rendimento; Ganho genético; Qualidade de seleção; Melhoramento Genético de Trigo.

3.1 Introdução

O trigo é um dos cereais mais cultivados do mundo e se destaca pela sua relevância na economia global. Em um programa de melhoramento de grãos, em específico no de trigo, as linhagens são produzidas nas fases iniciais, quando os melhoristas cruzam diferentes variedades de trigo para criar uma população segregante. Essa população é então cultivada em campo e avaliada para identificar indivíduos que apresentam características desejáveis em termos de produtividade, qualidade, resistência a estresses abióticos e bióticos, entre outras aplicações agrícolas adaptadas a diferentes condições ambientais. Esses indivíduos selecionados são então cultivados para produzir linhagens puras, que são avaliadas em diferentes condições ambientais para identificar aquelas com características superiores.

Assim, ensaios multi-ambientais são amplamente utilizados em experimentos dessa natureza pois as sementes são replicadas em diferentes ambientes (combinação particular de local/ano) para avaliar o comportamento das linhagens experimentais frente às variações edafoclimáticas, por exemplo, no melhoramento de trigo as linhagens são plantadas em alguns ambientes e a seleção das melhores ocorre através de estágios dos testes avançados.

Um dos estágios intrínsecos no programa de melhoramento de trigo é o teste preliminar de rendimento (*Preliminary Yield Trial* - PYT). Esse estágio constitui o primeiro ano de teste de rendimento das linhagens e tem como função selecionar linhagens superiores, na ausência de marcadores moleculares, que serão avaliadas nos próximos anos e/ou nos próximos estágios, como por exemplo, no teste de rendimento avançado (*Advanced Yield Trial* - AYT) (Endelman et al., 2014; Verges e Van Sanford, 2020; Lara et al., 2022). Além disso, também podem ser possíveis candidatas utilizadas como pais em futuros cruzamentos.

Endelman et al. (2014) mencionam que a precisão genética é a correlação entre os verdadeiros efeitos genéticos e seus EBLUPs (*Empirical Best Linear Unbiased Predictors*) e sugerem que uma maior precisão genética é alcançada no estágio PYT quando as linhagens são testadas em vários locais e/ou várias repetições das linhagens dentro do local (Endelman et al., 2014; Verges e Van Sanford, 2020). As linhagens do estágio PYT têm uma quantidade limitada de sementes, assim elas são avaliadas em uma quantidade menor de locais do que no estágio AYT. Cullis et al. (2020) mencionam que a seleção das linhagens nos estágios preliminares e avançados ocorre de forma sequencial (S_1 , S_2 , S_3 e S_4) em que a principal característica é o rendimento dos grãos. S_1 e S_2 são referidos como os ensaios em estágios iniciais em que os números de locais geralmente são menores do que quatro e não há repetição dentro de cada local. Além das novas linhagens também são consideradas parcelas contendo variedades de verificação, denominadas de variedades padrões (*checks*). Nas etapas S_3 e S_4 , que são os estágios mais avançados, as novas linhagens são avaliadas em até dez locais, com duas ou três repetições por local (Cullis et al., 2020).

Os delineamentos clássicos e os delineamentos baseados em modelos (delineamentos ótimos) são amplamente utilizados em ensaios multi-ambientais de melhoramento genético de plantas. Os delineamentos ótimos são uma classe de delineamentos experimentais que dependem de um modelo estatístico e de um critério de otimização. O modelo estatístico deve ser, quando possível, o mais próximo do modelo esperado para realizar a análise.

O critério estatístico pode estar relacionado à matriz de variâncias e covariâncias dos estimadores ou, analogamente, à matriz de informação (Shah e Sinha, 2012; Cullis et al., 2020). Os delineamentos ótimos possibilitam a obtenção de delineamentos alternativos que podem apresentar bons resultados para o ganho genético (Sermarini et al., 2020; Piepho et al., 2020; Paget et al., 2017; Santos, 2017). Existem vários critérios de otimização, que são utilizados a depender do interesse do pesquisador. O critério de otimização A , minimiza a soma das variâncias das estimativas dos tratamentos. com a utilização desse critério o delineamento selecionado apresenta uma variância mínima de erro de predição dos tratamentos.

Desta forma, o ganho genético esperado é maximizado quando a média das variâncias do erro de previsão (*average variance of pairwise differences* - AVPD) de todos os contrastes elementares entre os tratamentos é minimizada.

Com base nessas abordagens, é possível obter os delineamentos parcialmente replicados (*p*-rep) otimizados, propostos por Cullis et al. (2006). Os delineamentos *p*-rep otimizados são delineamentos em linhas e colunas, desbalanceados, em que uma porcentagem *p*% de novas linhagens são repetidas no experimento. Essa porcentagem *p*% é definida de acordo com o material genético disponível e do interesse prático do pesquisador. Cullis et al. (2020) mencionam que o delineamento *p*-rep foi introduzido para solucionar os problemas inerentes nos estágios iniciais dos programas de melhoramento genético de plantas, por exemplo, a disponibilidade de material genético nos estágios iniciais é limitada. Cullis et al. (2006) mostraram que o delineamento *p*-rep melhora a precisão da seleção das linhagens quando comparado ao delineamento *grid-plot* (Kempton, 1984), que, também, é um delineamento de linhas e colunas, mas parcelas com *checks*, que são variedades de verificação, são distribuídas entre parcelas com as novas linhagens que não são replicadas. Os modelos lineares mistos são preferíveis para gerar delineamentos ótimos, pois são flexíveis e possibilitam variações nas estruturas das matrizes de variâncias e covariâncias (Piepho et al., 2020; Sermarini et al., 2020; Cullis et al., 2020).

Além disso, os modelos lineares mistos são amplamente utilizados nas análises de dados de melhoramento genético de plantas, pois é possível investigar, por exemplo, o efeito da interação entre variedades e o ambiente. Gilmour et al. (1997) utilizaram esses modelos e identificaram três componentes relacionados à variação espacial dos dados em experimentos de campo, que são: variação local, global e estranha. Considerando a abordagem espacial dada por Gilmour et al. (1997), Smith et al. (2001) analisaram dados de cevada e apresentaram o modelo misto de análise fatorial para ensaios multi-ambientais, permitindo incorporar a variância genética de cada ambiente separadamente. Uma outra característica dos modelos lineares mistos é a possibilidade de incluir a matriz de relacionamento, que explora a correlação genética entre linhagens relacionadas, melhorando as estimativas de efeitos genéticos e reduzindo o viés. Existe uma vasta literatura sobre a inclusão da matriz de relacionamento em modelos estatísticos (Oakey et al., 2006, 2007; Piepho et al., 2008; Cullis et al., 2020; Piepho et al., 2020). Piepho et al. (2008) mostraram que os BLUPs (Best Linear Unbiased Predictions) com informações de parentesco têm boa precisão preditiva quando comparado aos BLUPs sem informações de parentesco.

Oakey et al. (2006) analisaram dados de melhoramento genético de trigo em um único local modelando a variação ambiental frequentemente presente em ensaios de campo e adicionaram matriz de parentesco. O modelo proposto abordou a decomposição dos efeitos genéticos das linhagens em efeitos genéticos aditivos e não aditivos. Posteriormente, Oakey et al. (2007) estenderam a decomposição da variância genética para os casos de ensaios multi-ambientais, com base no modelo misto de análise fatorial apresentado por Smith et al. (2001), reforçando que o uso da matriz de relacionamento permite que estruturas populacionais mais gerais sejam consideradas. Os autores ainda consideraram que os efeitos não aditivos podem ser divididos em efeitos de dominância e epistasia. Piepho et al. (2020) apresentaram um tutorial demonstrando os procedimentos necessários para gerar delineamentos ótimos considerando a dependência entre os efeitos da linhagem com base no OPTEx do Sistema SAS. Cullis et al. (2020) propuseram o uso de modelos lineares mistos para gerar delineamentos ótimos em um único local, incluindo a matriz de relacionamentos baseando-se em dados históricos dos programas de melhoramento. Os autores utilizaram um modelo semelhante ao proposto por Oakey et al. (2006).

Cullis et al. (2020) mencionam que o *p*-rep foi desenvolvido para incluir a matriz de relacionamento; entretanto, há poucos relatos na literatura considerando a inclusão de informações de *pedigree* em modelos para geração de delineamentos otimizados em ensaios multi-ambientais. O principal objetivo deste trabalho é comparar, por meio de estudo de simulação, delineamentos espacialmente otimizados para uma área experimental de 24 linhas por 10 colunas, 180 *test lines*, quatro variedades de verificação

e dois blocos em três locais, assumindo-se o critério A de otimização, com relação (i) à suposição de independência ou dependência entre os efeitos de *test lines*; (ii) a diferentes valores de parâmetros associados às variâncias genética e residual dentro de cada local, (iii) a diferentes valores para o parâmetro associado à correlação entre os efeitos de uma mesma *test lines* em locais distintos; (iv) e diferentes tipos de delineamentos (*grid-plot* e *p-rep*). Para o delineamento *p-rep* foram consideradas diferentes porcentagens ($p\%$) de linhagens repetidas e diferentes números de *checks*. Como medidas de comparação entre os delineamentos foram utilizados: o ganho genético, representado por medida *Relative Realized Genetic Gain* (RRGG), a precisão genética e o sucesso de seleção das verdadeiramente melhores linhagens, considerando as análises individuais por local e conjunta, com base no modelo linear misto que incorpora a modelagem da variação ambiental e a matriz de relacionamento. A matriz de parentesco utilizada foi proveniente de um programa de melhoramento genético de trigo que foi simulado de acordo com Lara et al. (2022) e as linhagens consideradas foram referentes ao estágio PYT. Na Seção 3.2, são apresentadas as especificações do modelo para a obtenção dos delineamentos e, também, para o estudo de simulação dos dados. Além disso, são apresentadas as medidas utilizadas para a comparação dos delineamentos. Na Seção 3.3, são apresentados os principais resultados. A conclusão é apresentada na Seção 3.4.

3.2 Metodologia

3.2.1 Matriz de relacionamento

Os programas de melhoramento de trigo concentram-se no desenvolvimento de linhagens endogâmicas ou consanguíneas, isto significa que, o cruzamento ocorre pela união de indivíduos geneticamente semelhantes (aparentados). Tradicionalmente, após o cruzamento e a melhoria da população, as linhagens endogâmicas são desenvolvidas através da autopolinização ou de haploides duplicados. As linhagens endogâmicas são, então, fenotipadas em *headrows* (viveiros) e ensaios de campo antes de serem selecionadas como futuros pais. No trigo, o ciclo de reprodução pode levar de quatro a seis anos, a depender da estrutura do programa de melhoramento e da preferência do criador (Merrick et al., 2022; Lara et al., 2022; Gaynor et al., 2017).

Os *headrows* são o primeiro estágio da fenotipagem, e, no quarto ano, as melhores linhagens observadas no *headrows* são selecionadas e avaliadas em um ensaio de rendimento preliminar, que ocorre no quinto ano. Só então, após seis anos, são iniciados os testes de campo replicados nos ensaios avançados. As linhagens endogâmicas permanecem em ensaios de rendimento replicado por mais cinco anos, quando as variedades são então liberadas (Merrick et al., 2022). Seguindo essa estrutura, a matriz de parentesco utilizada neste estudo foi obtida após a simulação de um programa completo de melhoramento de trigo por 40 anos, e foi baseado na pesquisa de Lara et al. (2022), em que os autores seguiram um programa descrito por Gaynor et al. (2017) com 21 anos de seleção fenotípica convencional para rendimento.

Inicialmente, realizou-se uma simulação coalescente de um genoma completo sequenciado. Para isto, fixou-se a quantidade de haplótipos fundadores que foram usados para formar os primeiros pais no programa de melhoramento. Neste estudo, foram simulados os haplótipos fundadores para 70 indivíduos consanguíneos com 21 pares de cromossomos cada um com 600 polimorfismos de nucleotídeo único (*single nucleotide polymorphism* - SNP) como marcadores por cromossomo, e atribuiu-se, de forma aleatória, 100 SNP como QTL (locos de caracteres quantitativos) por cromossomo. O SNP é um tipo específico de marcador genético que é usado para identificar variações individuais na sequência de DNA em uma população. Desta forma, refere-se a uma variação na sequência de um único nucleotídeo (polimorfismo) que é comum em uma população. O QTL, por sua vez, é um marcador genético que é associado a uma característica quantitativa, ou seja, uma característica que é controlada por múltiplos genes e pelo ambiente. Desta forma, os SNPs foram utilizados como marcadores genéticos para mapear as regiões no genoma que estão associadas a traços quantitativos, obtidos pelos QTLs.

Em um segundo momento, definiram-se os parâmetros globais, assumindo-se que o QTL 2.100 teve efeito aditivo no rendimento e seus efeitos foram amostrados a partir de uma distribuição normal.

Os genótipos foram codificados como 0 para homozigoto de referência (ancestral), 1 para heterozigoto e 2 para homozigoto alternativo (derivado). A partir das sequências genômicas simuladas, foram criadas 70 linhagens consanguíneas. A variância genética aditiva entre essas linhagens foi fixada em 0,1. Posteriormente, cruzaram-se as linhagens consanguíneas para gerar 100 populações biparentais. Cada população teve 100 progênie (F_1) em que o seu genoma foi dobrado e avaliadas nos *headrows*, ao todo 10.000 linhagens foram observadas. As linhagens experimentais foram selecionadas de acordo com seu desempenho e as 500 melhores foram escolhidas com base em seleção fenotípica em um ensaio PYT, das quais as 50 melhores foram consideradas em um ensaio AYT e, posteriormente, as 10 melhores em um ensaio de rendimento de elite (EYT). No ensaio de rendimento de elite as linhagens são avaliadas por dois anos consecutivos em um grande ensaio replicado em vários locais e, então, uma variedade é selecionada. Este programa de melhoramento de trigo foi simulado utilizando o pacote **AlphaSimR** (Gaynor et al., 2021). Para obter a matriz de parentesco foram utilizados os 40 anos de seleção fenotípica convencional para rendimento considerando o *pedrigree* das linhagens no estágio PYT do 39º ano e conhecendo seis gerações anteriores. A inversa da matriz de relacionamento foi obtida utilizando o pacote **Pedicure** (Butler, 2019) atribuindo 0 para os coeficientes de endogamia e 10 para linhagens.

3.2.2 Delineamentos gerados

As configurações dos delineamentos ótimos utilizados neste estudo foram análogos aos considerados por Goes (2020), e são apresentadas na Tabela 3.1. Em todos os casos, a área experimental foi constituída por três locais (n_s), sendo a área dividida em dois blocos (cada bloco tem 5 colunas), 24 linhas (n_r) por 10 colunas ($n_c = 10$) e $m = 184$ tratamentos (novas linhagens e *checks*), provenientes dos estágio PYT do programa de melhoramento genético de trigo. Para os delineamentos *p*-rep otimizados, consideraram-se diferentes valores para as porcentagens *p*% de linhagens duplicadas e diferentes números de *checks* e suas repetições.

Tabela 3.1. Caracterização dos delineamentos de acordo com o número de linhagens duplicadas e *checks*.

Caracterização	<i>p</i> %	Número de linhagens		Número de <i>checks</i>	
		Não duplicadas	Duplicadas	Quantidade	Número de repetição
grid-plot	0	180	0	4	15
<i>p</i> -rep 1	11	160	20	4	10
<i>p</i> -rep 2	22	140	40	4	5
<i>p</i> -rep 3	22	140	40	2	10
<i>p</i> -rep 4	22	140	40	1	20
<i>p</i> -rep 5	33	120	60	0	0

Os delineamentos ótimos foram gerados considerando o modelo linear misto na abordagem dada por Gilmour et al. (1997) e Smith et al. (2001). Essa mesma abordagem foi aplicada para realizar o estudo de simulação e, posteriormente, analisar os dados. O modelo é definido como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_g\mathbf{u}_g + \mathbf{Z}_o\mathbf{u}_o + \boldsymbol{\epsilon} \quad (3.1)$$

em que $\mathbf{y}_{(n \times 1)}$ é o vetor de observações dos dados, sendo $n = n_s \times n_r \times n_c$; $\boldsymbol{\tau}_{(q \times 1)}$ é o vetor de q efeitos fixos de cada local (ambiente) e de blocos com a matriz de delineamento $\mathbf{X}_{(n \times q)}$; $\mathbf{u}_g_{(mn_s \times 1)}$ é o vetor de efeitos aleatórios das linhagens dentro de cada local, com matriz de delineamento $\mathbf{Z}_g_{(n \times mn_s)}$; $\mathbf{u}_o_{(b \times 1)}$ é o vetor de b efeitos aleatórios dos efeitos não genéticos de linhas e colunas dentro de cada local, com matriz de delineamento $\mathbf{Z}_o_{(n \times b)}$; $\boldsymbol{\epsilon}_{(n \times 1)}$ é o vetor de erros aleatórios. Neste estudo, consideraram-se que o vetor

dos efeitos genéticos \mathbf{u}_g é composto apenas pelos efeitos genéticos aditivos, isto é, não consideraram-se os efeitos de dominância e os residuais não aditivos. Assumiu-se que \mathbf{u}_g , \mathbf{u}_o , $\boldsymbol{\epsilon}$ são independentes com distribuição conjunta Gaussiana multivariada com vetor de médias zero e matriz de variâncias e covariâncias

$$\begin{bmatrix} \mathbf{G}_g(\boldsymbol{\sigma}_{g_s}) & 0 & 0 \\ 0 & \mathbf{G}_o(\boldsymbol{\sigma}_o) & 0 \\ 0 & 0 & \mathbf{R}(\boldsymbol{\phi}) \end{bmatrix}.$$

$\mathbf{G}_g(\boldsymbol{\sigma}_{g_s})$ é definido como:

$$\mathbf{G}_g(\boldsymbol{\sigma}_{g_s}) = \mathbf{G}_a \otimes \mathbf{A}_m = \begin{bmatrix} \sigma_{g_{s1}}^2 & \phi_g \sigma_{g_{s1}} \sigma_{g_{s2}} & \phi_g \sigma_{g_{s1}} \sigma_{g_{s3}} \\ \phi_g \sigma_{g_{s1}} \sigma_{g_{s2}} & \sigma_{g_{s2}}^2 & \phi_g \sigma_{g_{s2}} \sigma_{g_{s3}} \\ \phi_g \sigma_{g_{s1}} \sigma_{g_{s3}} & \phi_g \sigma_{g_{s2}} \sigma_{g_{s3}} & \sigma_{g_{s3}}^2 \end{bmatrix} \otimes \mathbf{A}_m,$$

em que $\boldsymbol{\sigma}_{g_s} = (\sigma_{g_{s1}}^2, \sigma_{g_{s2}}^2, \sigma_{g_{s3}}^2, \phi_g)'$ corresponde às variâncias genéticas aditivas dentro de cada ambiente e os elementos fora da diagonal são as covariâncias genéticas entre pares de locais, ϕ_g é a correlação genética entre as mesmas linhagens aplicadas em diferentes locais, \mathbf{G}_a é a matriz de variância genética aditiva e \mathbf{A}_m é uma matriz de relacionamento conhecida de ordem k , em que k corresponde ao número de *test lines* e variedades padrões utilizadas para gerar os delineamentos. $\mathbf{G}_o(\boldsymbol{\sigma}_o) = \mathbf{I}_{n_s} \otimes (\sigma_r^2 \mathbf{I}_{n_r} \oplus \sigma_c^2 \mathbf{I}_{n_c})$ em que $\boldsymbol{\sigma}_o = (\sigma_r^2, \sigma_c^2)'$ e σ_r^2 e σ_c^2 são às variâncias dos efeitos não genéticos de linhas e colunas que foram assumidas iguais em todos os ambientes e cenários; \mathbf{I}_{n_s} , \mathbf{I}_{n_r} e \mathbf{I}_{n_c} são matrizes identidades de ordens n_s , n_r e n_c , respectivamente. Por fim, $\mathbf{R}(\boldsymbol{\phi}) = \bigoplus_{i=1}^3 \sigma_{s_i}^2 \text{AR1}(\phi_r) \otimes \text{AR1}(\phi_c)$, ou seja, uma soma direta das variâncias residuais dentro de cada local ($\sigma_{s_i}^2$) e dos processos separáveis autorregressivos de primeira ordem, como sugerido por Gilmour et al. (1997). Neste caso, $\boldsymbol{\phi} = (\phi_r, \phi_c)'$, sendo, $\phi_r = \phi_c = 0,5$ são os parâmetros de correlações residuais entre linhas e entre colunas dentro de cada local. Na Tabela 3.2 são apresentados os valores dos parâmetros de variâncias para cada um dos cenários considerados neste estudo.

Tabela 3.2. Conjuntos de parâmetros para os efeitos genéticos e não genéticos para gerar os delineamentos ótimos e adotados no estudo de simulação, em que σ_r^2 , σ_c^2 , σ_g^2 , σ_s^2 , são os valores para às variâncias de linhas, colunas, genéticas aditivas e residuais dentro de cada ambiente.

	Cenário 1					Cenário 2				
	σ_r^2	σ_c^2	σ_g^2	σ_s^2	σ_g^2/σ_s^2	σ_r^2	σ_c^2	σ_g^2	σ_s^2	σ_g^2/σ_s^2
Local 1	0,10	0,30	0,80	1,00	0,80	0,10	0,30	0,18	0,30	0,60
Local 2	0,10	0,30	1,00	1,25	0,80	0,10	0,30	0,30	0,50	0,60
Local 3	0,10	0,30	2,00	1,25	1,60	0,10	0,30	0,54	0,90	0,60

Foram considerados três valores para ϕ_g , 0,7, 0,8 e 0,9, caracterizando a dependência entre os efeitos das mesmas linhagens aplicadas em diferentes locais. Para gerar os delineamentos e, posteriormente, realizar o estudo de simulação foi considerado o mesmo critério utilizado durante a simulação do programa de melhoramento em que se adotou somente a variância genética aditiva. Esse procedimento é consequência do fato de que foi simulado um programa de reprodução que usa indivíduos totalmente consanguíneos (exceto na F1). Sendo assim, assumiu-se apenas variação genética aditiva (Lara et al., 2022). Assim, para este estudo de simulação consideraram-se dois cenários, como descrito na Tabela 3.2. Os valores dos parâmetros de variâncias foram assumidos com base nos trabalhos de Cullis et al. (2006), Crossa et al. (2010), Santos (2017), Sukumaran et al. (2017), Goes (2020) e Sermarini et al. (2020).

No primeiro cenário, a razão entre a variância genética aditiva e a variância residual ($\sigma_{g_s}^2/\sigma_s^2$) dos Locais 1 e 2 são iguais, no Local 3 essa razão é o dobro. No segundo cenário, fixou-se a mesma razão

para os três ambientes, sendo igual a 0,6 para avaliar a influência da variância genética aditiva e residual nos valores do ganho genético e na qualidade de seleção do material genético.

Neste estudo, também, foram gerados delineamentos ótimos assumindo-se que os efeitos dos tratamentos (*test lines* e *checks*) são independentes e neste caso, deve-se considerar uma matriz identidade ao invés de considerar a matriz \mathbf{A}_m de relacionamento. O modelo utilizado para gerar os delineamentos é análogo ao apresentado na Equação 3.1, com as mesmas definições, diferenciando-se apenas na estrutura da matriz de variâncias e covariâncias associada aos efeitos genéticos dentro de cada local. Assim, temos:

$$\mathbf{G}_g(\sigma_{g_s}) = \mathbf{G}_a \otimes \mathbf{I}_m = \begin{bmatrix} \sigma_{g_{s1}}^2 & \phi_g \sigma_{g_{s1}} \sigma_{g_{s2}} & \phi_g \sigma_{g_{s1}} \sigma_{g_{s3}} \\ \phi_g \sigma_{g_{s1}} \sigma_{g_{s2}} & \sigma_{g_{s2}}^2 & \phi_g \sigma_{g_{s2}} \sigma_{g_{s3}} \\ \phi_g \sigma_{g_{s1}} \sigma_{g_{s3}} & \phi_g \sigma_{g_{s2}} \sigma_{g_{s3}} & \sigma_{g_{s3}}^2 \end{bmatrix} \otimes \mathbf{I}_m,$$

em que \mathbf{I}_m é uma matriz identidade de ordem m . Em todos os delineamentos consideraram-se cento e oitenta linhagens do estágio PYT para as *test lines* e até quatro variedades de gerações anteriores aos pais para as *checks*. As *test lines* foram rotuladas de 1 até 180 e, as variedades padrões que foram rotuladas de A até D. As *test lines* que foram replicadas em um ambiente, não foram replicadas em outro (Williams et al., 2011). Dessa forma, as *test lines* que foram replicadas totalizaram no máximo quatro observações quando os três locais foram considerados.

Assim, como em Cullis et al. (2020), foi utilizado o critério A para a obtenção do delineamento. Tal critério de otimização busca um delineamento ótimo que minimiza a média das variâncias das diferenças pareadas (*average variance of pairwise differences* - AVPD) dos efeitos do tratamento, dada por,

$$\text{AVPD} = \frac{2}{m-1} \left(\text{tr}(\mathbf{C}^{-1}) - \frac{1}{m} \mathbf{1}' \mathbf{C}^{-1} \mathbf{1} \right), \quad (3.2)$$

em que m é o número de tratamentos, $\text{tr}()$ é o traço da matriz, \mathbf{C} é a matriz de informação dos tratamentos e $\mathbf{1}$ é um vetor de uns. De acordo com Hooks et al. (2009), ao considerar os efeitos aleatórios dos tratamentos, a matriz \mathbf{C} é definida da seguinte forma:

$$\mathbf{C} = \mathbf{Z}'_g (\mathbf{Z}_o \mathbf{G}_o(\sigma_o) \mathbf{Z}'_o + \mathbf{R}(\phi))^{-1} \mathbf{Z}_g + \mathbf{G}_g^{-1}(\sigma_{g_s}) - \mathbf{Z}'_g (\mathbf{Z}_o \mathbf{G}_o(\sigma_o) \mathbf{Z}'_o + \mathbf{R}(\phi))^{-1} \mathbf{X} (\mathbf{X}' (\mathbf{Z}_o \mathbf{G}_o(\sigma_o) \mathbf{Z}'_o + \mathbf{R}(\phi))^{-1} \mathbf{X})^{-1} \mathbf{X}' (\mathbf{Z}_o \mathbf{G}_o(\sigma_o) \mathbf{Z}'_o + \mathbf{R}(\phi))^{-1} \mathbf{Z}_g. \quad (3.3)$$

Mais detalhes sobre essa medida poderão ser consultados Atkinson e Donev (1992), Cullis et al. (2006) e Cullis et al. (2020). Foram considerados 20.000 movimentos para a busca do delineamento. Os delineamentos foram obtidos utilizando os pacotes `owd` (Butler, 2020) no software R (R Core Team, 2020).

3.2.3 Estudo de simulação e medidas das comparações

3.2.3.1 Simulação dos dados

Após a obtenção dos delineamentos, realizou-se um estudo de simulação para a obtenção dos dados com base no modelo linear misto apresentado na Equação 3.1. Os dados foram simulados para cada uma das caracterizações dos delineamentos obtidos, sendo do tipo *grid-plot* ou *p-rep*. Foi assumido na simulação o mesmo conjunto de parâmetros apresentados na seção 3.2.2, foram considerados efeitos aleatórios de *test lines* e das *checks*. Os efeitos dos locais e dos blocos foram fixos. As possíveis combinações dos vetores simulados resultaram em 1000 vetores para os efeitos aleatórios \mathbf{u}_g , \mathbf{u}_o e $\boldsymbol{\epsilon}$, e a partir de uma distribuição normal padrão, esses vetores foram utilizados em todas as simulações para garantir as mesmas condições para todos os delineamentos. Em seguida, foram transformados de tal modo a serem amostras aleatórias de uma distribuição normal com os respectivos parâmetros de variâncias considerados em cada

um dos cenários. Assim, obteve-se (i) 1000 vetores de tamanho 102 $((24+10) \times 3)$ para os efeitos não genéticos de linhas e colunas (\mathbf{u}_o); (ii) 1000 vetores de tamanho 720 $(24 \times 10 \times 3)$ para o efeito residual e para cada ϕ_g obtivemos (iii) 1000 vetores para os efeitos genéticos em cada local (\mathbf{u}_g) em que o tamanho dependerá do delineamento adotado. Desta forma, irá variar entre 552 à 540. A tripla $(\mathbf{u}_o, \mathbf{u}_g, \epsilon)$ obtida pela transformação foi considerada como os verdadeiros efeitos. Em todos os casos, obteve-se 1000 vetores de respostas simulados para cada um dos delineamentos ótimos obtidos.

3.2.3.2 Modelagem estatística e medidas de comparação dos delineamentos

Ajustou-se, para cada conjunto de dados simulados, um modelo conforme a Equação 3.1, para as análises conjuntas e adicionalmente, foram realizadas análises individuais em cada um dos locais, utilizando o método de estimação da máxima verossimilhança restrita (REML), desenvolvido por Patterson e Thompson (1971). Além disso, os dados obtidos dos delineamentos ótimos gerados considerando a matriz de parentesco no modelo do delineamento, também foram analisados considerando um modelo de análise sem a matriz de relacionando, para verificar a influência da matriz na obtenção do delineamento. Para calcular as medidas de comparação registraram-se os EBLUPs, para cada modelo ajustado. Também foram registradas as estimativas dos componentes de variância e seus respectivos erros padrão.

As medidas de comparação consideradas foram: o Ganho Genético Relativo Percebido (*Realized Relative Genetic Gain* - RRGG) que mede o ganho genético percebido e é obtido a partir do *realized genetic gain* (RGG), como a razão entre a média dos melhores $s\%$ EBLUPs e a média dos melhores $s\%$ dos verdadeiros efeitos genéticos, em que $s\%$ é a porcentagem de seleção, aqui assumido 10%, um valor comumente adotado pelos programas brasileiros de melhoramento de plantas (Sermarini et al., 2020; Goes, 2020).

Seguindo Sermarini et al. (2020), também foram calculadas medidas secundárias (i) a precisão genética (correlação entre os verdadeiros efeitos genéticos e seus EBLUPs) e (ii) sucesso de seleção (probabilidade de seleção das verdadeiras melhores $s\% = 10\%$ *test lines*). Para comparar os grupos com os tratamentos (*test lines* e *checks*) repetidos e os não repetidos foi utilizada a média da variância das diferenças entre pares de tratamentos, calculadas a partir da matriz de variâncias para previsões, descrita na Equação 3.2. Para obter essas medidas, utilizou-se o pacote *dae* (Brien, 2021b), do *software* R.

Os estudos de simulação e análises de dados foram realizados no software R (R Core Team, 2020) usando os pacotes *od* (Butler, 2019) para a obtenção dos delineamentos, *dae* para obtenção da anatomia do delineamento (Brien, 2021b), *ggplot2* (Wickham, 2016) e o *asreml-R* (Butler, 2021). Uma das vantagens do pacote *asreml-R* é a capacidade de ajustar modelos lineares mistos eficientes para conjuntos de dados grandes, complexos e desbalanceados, devido ao uso do algoritmo *Average Information* (A_I) e métodos de matriz esparsa para ajustar os modelos. O procedimento usado para obter as estimativas dos parâmetros de variância utilizando o *asreml-R* é apresentado na Seção 3.2.3.3.

3.2.3.3 Estimação dos parâmetros de variância usando o asreml-R

O método de estimação da máxima verossimilhança restrita é um método alternativo utilizado para realizar inferências sobre os componentes de variâncias nos modelos lineares mistos que foi desenvolvido por Patterson e Thompson (1971). No *asreml* a estimação dos parâmetros envolve dois processos que estão relacionados com a estimação dos efeitos fixos e a predição dos efeitos aleatórios. Desta forma, considerando a Equação 3.1, segue que \mathbf{y} tem distribuição normal multivariada, assim:

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\tau}, \mathbf{H}),$$

em que $\mathbf{H} = \mathbf{Z}_g \mathbf{G}(\boldsymbol{\sigma}_g) \mathbf{Z}_g^T + \mathbf{Z}_o \mathbf{G}(\boldsymbol{\sigma}_o) \mathbf{Z}_o^T + \mathbf{R}(\phi)$.

A derivação fornecida a seguir é apresentada em Verbyla (1990) em que \mathbf{y} é transformado utilizando uma matriz não singular $\mathbf{L} = [\mathbf{L}_1 \ \mathbf{L}_2]$ tal que

$$\mathbf{L}_1^T \mathbf{X} = \mathbf{I}_p, \quad \mathbf{L}_2^T \mathbf{X} = \mathbf{0}.$$

Se $\mathbf{y}_j = \mathbf{L}_j^T \mathbf{y}$, com $j = 1, 2$,

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \sim N \left(\begin{bmatrix} \boldsymbol{\tau} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{L}_1^T \mathbf{H} \mathbf{L}_1 & \mathbf{L}_1^T \mathbf{H} \mathbf{L}_2 \\ \mathbf{L}_2^T \mathbf{H} \mathbf{L}_1 & \mathbf{L}_2^T \mathbf{H} \mathbf{L}_2 \end{bmatrix} \right)$$

A distribuição de $\mathbf{L}^T \mathbf{y}$ pode ser dividida em uma distribuição condicional, ou seja, $\mathbf{y}_1 | \mathbf{y}_2$, para estimação de $\boldsymbol{\tau}$, e uma distribuição marginal baseada em \mathbf{y}_2 para estimação de $\boldsymbol{\sigma}_{g_s}$, $\boldsymbol{\sigma}_o$ e $\boldsymbol{\phi}$; Esta última baseia-se na verossimilhança restrita.

A estimação de $\boldsymbol{\tau}$, é obtida igualando \mathbf{y}_1 à sua probabilidade condicional e, é possível obter

$$\hat{\boldsymbol{\tau}} = (\mathbf{X}^T \mathbf{H}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}^{-1} \mathbf{y}.$$

A estimação de $\boldsymbol{\kappa} = [\boldsymbol{\sigma}_{g_s}^T \ \boldsymbol{\sigma}_o^T \ \boldsymbol{\phi}^T]^T$ é baseada no logaritmo da função de verossimilhança restrita,

$$\begin{aligned} l_R &= -\frac{1}{2} [\log(\det \mathbf{L}_2^T \mathbf{H}^{-1} \mathbf{L}_2) + \mathbf{y}_2^T (\mathbf{L}_2^T \mathbf{H} \mathbf{L}_2)^{-1} \mathbf{y}_2] \\ &= -\frac{1}{2} [\log(\det \mathbf{X}^T \mathbf{H}^{-1} \mathbf{X}) + \log \det \mathbf{H} + \mathbf{y}^T \mathbf{P} \mathbf{y}] \end{aligned} \quad (3.4)$$

em que $\mathbf{P} = \mathbf{H}^{-1} - \mathbf{H}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{H}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{H}^{-1}$.

Observe que $\mathbf{y}^T \mathbf{P} \mathbf{y} = (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\tau}})^T \mathbf{H}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\tau}})$. A log-verossimilhança 3.4 depende de \mathbf{X} e não de uma transformação particular não única definida por \mathbf{L} .

$$l_R = -\frac{1}{2} [\log(\det \mathbf{C}) + \log(\det \mathbf{R}) + \log(\det \mathbf{G}_g) + \log(\det \mathbf{G}_o) + \mathbf{y}^T \mathbf{P} \mathbf{y}].$$

Escrevemos

$$\mathbf{P} = \mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{W} \mathbf{C}^{-1} \mathbf{W}^T \mathbf{R}^{-1},$$

com $\mathbf{W} = [\mathbf{X} \ \mathbf{Z}_g \ \mathbf{Z}_o]$. Fazendo $\boldsymbol{\kappa} = [\boldsymbol{\sigma}_{g_s}^T \ \boldsymbol{\sigma}_o^T \ \boldsymbol{\phi}^T]^T$, as estimativas REML de κ_i são obtidas calculando a função escore

$$U(\kappa_i) = \frac{\partial L_R}{\partial \kappa_i} = -\frac{1}{2} [\text{tr}(\mathbf{P} \mathbf{H}_i) - \mathbf{y}^T \mathbf{P} \mathbf{H}_i \mathbf{P} \mathbf{y}] \quad (3.5)$$

igualando-a a zero. Observe que $\mathbf{H}_i = \frac{\partial \mathbf{H}}{\partial \kappa_i}$.

Os elementos da matriz de informação esperada são:

$$-\frac{\partial^2 L_R}{\partial \kappa_i \partial \kappa_j} = \frac{1}{2} \text{tr}(\mathbf{P} \mathbf{H}_{ij}) - \frac{1}{2} \text{tr}[\mathbf{P} \mathbf{H}_i \mathbf{P} \mathbf{H}_j + \mathbf{y}^T \mathbf{P} \mathbf{H}_i \mathbf{P} \mathbf{H}_j \mathbf{P} \mathbf{y} - \frac{1}{2} \text{tr}(\mathbf{y}^T \mathbf{P} \mathbf{H}_{ij} \mathbf{P} \mathbf{y})] \quad (3.6)$$

em que $\mathbf{H}_{ij} = \frac{\partial \mathbf{H}}{\partial \kappa_i \partial \kappa_j}$.

Os elementos da matriz de informação esperada são

$$E \left(\frac{\partial^2 L_R}{\partial \kappa_i \partial \kappa_j} \right) = \frac{1}{2} \text{tr}(\mathbf{P} \mathbf{H}_i \mathbf{P} \mathbf{H}_j). \quad (3.7)$$

Dada uma estimativa inicial $\boldsymbol{\kappa}^{(0)}$, uma atualização $\boldsymbol{\kappa}, \boldsymbol{\kappa}^{(1)}$ usando o algoritmo *Fisher-score* (FS) é:

$$\boldsymbol{\kappa}^{(1)} = \boldsymbol{\kappa}^{(0)} + \mathbf{I}(\boldsymbol{\kappa}^{(0)}, \boldsymbol{\kappa}^{(0)})^{-1} \mathbf{U}(\boldsymbol{\kappa}^{(0)}) \quad (3.8)$$

em que $U(\boldsymbol{\kappa}^{(0)})$ é a função escore (3.5) e $\mathbf{I}(\boldsymbol{\kappa}^{(0)}, \boldsymbol{\kappa}^{(0)})$ é a matriz de informação esperada (3.7) de $\boldsymbol{\kappa}$ avaliada em $\boldsymbol{\kappa}^{(0)}$.

A matriz utilizada pelo algoritmo *Average Information* denotada por \mathbf{I}_A é obtida calculando a média (3.6) e (3.7) e aproximando $\mathbf{y}^T \mathbf{P} \mathbf{H}_{ij} \mathbf{P} \mathbf{y}$ por sua esperança, $\text{tr}(\mathbf{P} \mathbf{H}_{ij})$ nos casos em que $\mathbf{P} \mathbf{H}_{ij} \neq 0$. Para modelos de componentes de variância (isto é, aqueles lineares em relação às variâncias em \mathbf{H}), os termos em \mathbf{I}_A são médias exatas daqueles em (3.6) e (3.7). A ideia básica é usar $\mathbf{I}_A(\kappa_i, \kappa_j)$ no lugar da matriz de informação esperada em (3.8) para atualizar $\boldsymbol{\kappa}$.

Os elementos de \mathbf{I}_A são

$$\mathbf{I}_A = \frac{1}{2} \mathbf{y}^T \mathbf{P} \mathbf{H}_i \mathbf{P} \mathbf{H}_j \mathbf{P} \mathbf{y}.$$

A matriz \mathbf{I}_A é a soma de quadrados residual e matriz de produtos de

$$\mathbf{y} = [\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_n]$$

em que \mathbf{y}_i é a variável de 'trabalho' para $\boldsymbol{\kappa}_i$ e é dada por

$$\begin{aligned} \mathbf{y}_i &= \mathbf{H}_i \mathbf{P} \mathbf{y} \\ &= \mathbf{H}_i \mathbf{R}^{-1} \tilde{\boldsymbol{\epsilon}}, \end{aligned}$$

em que $\tilde{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\tau}} - \mathbf{Z}_g \tilde{\mathbf{u}}_g - \mathbf{Z}_o \tilde{\mathbf{u}}_o$. E $\hat{\boldsymbol{\tau}}$, $\tilde{\mathbf{u}}_g$, e $\tilde{\mathbf{u}}_o$ são as soluções para os parâmetros.

3.3 Resultados e discussão

Nas Figuras 3.1 e 3.2, são apresentados os croquis para os delineamentos *p*-rep 3 (Tabela 3.1), considerando-se os cenários 1 e 2, indicados na Tabela 3.2, respectivamente. Em ambas as figuras, nos croquis apresentados assumiu-se $\phi_g = 0,7$ e em (a) tem-se a dependência entre os tratamentos e em (b) independência. De maneira geral, para todos os delineamentos considerados, os tratamentos estão aleatoriamente distribuídos na área experimental e não foram observadas grandes frequências de *test lines* ocupando parcelas vizinhas ou parcelas na fronteira da área experimental, isto é, na região localizada no limite estabelecido pelo bloco e nas bordaduras. No entanto, percebeu-se que quão maiores os parâmetros associados às variâncias genética e residual dentro de cada local, maior concentração de *checks* ocupando parcelas vizinhas, porém melhor é a distribuição das *test lines* na área experimental. No segundo cenário, observou-se um comportamento contrário, quão menores tais parâmetros, maior foi a incidência de *test lines* em parcelas próximas.

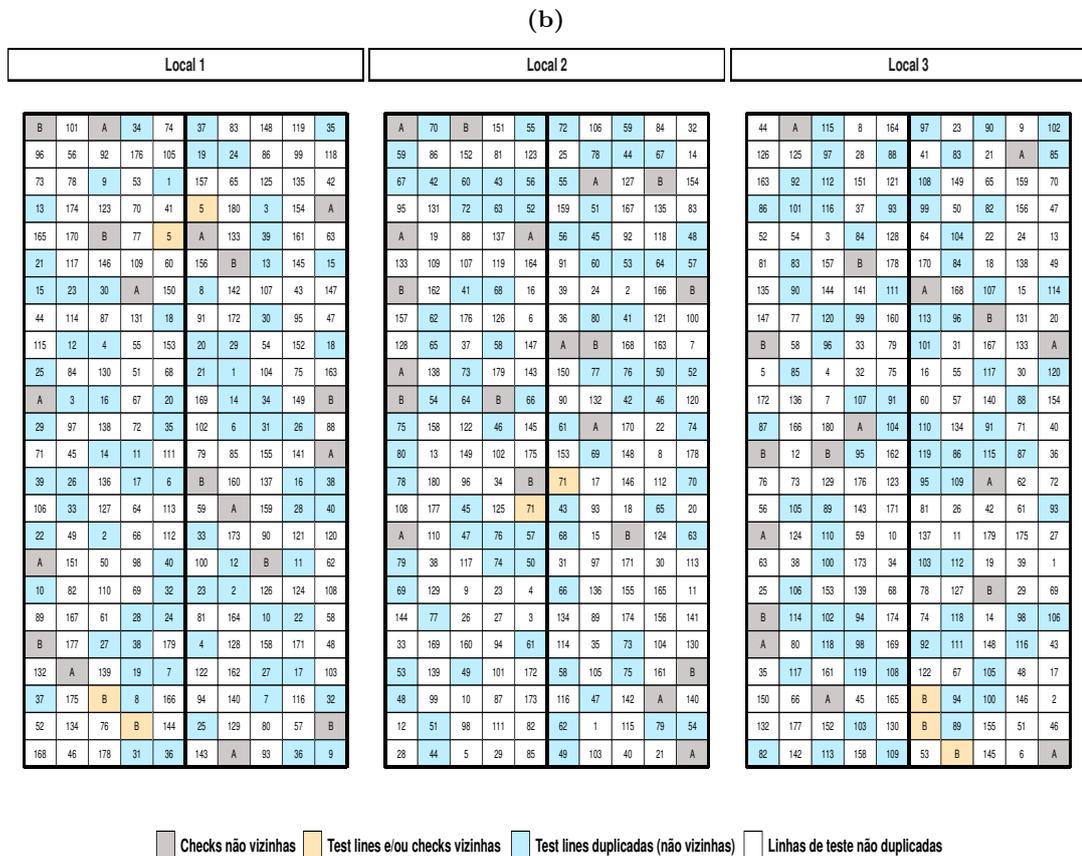
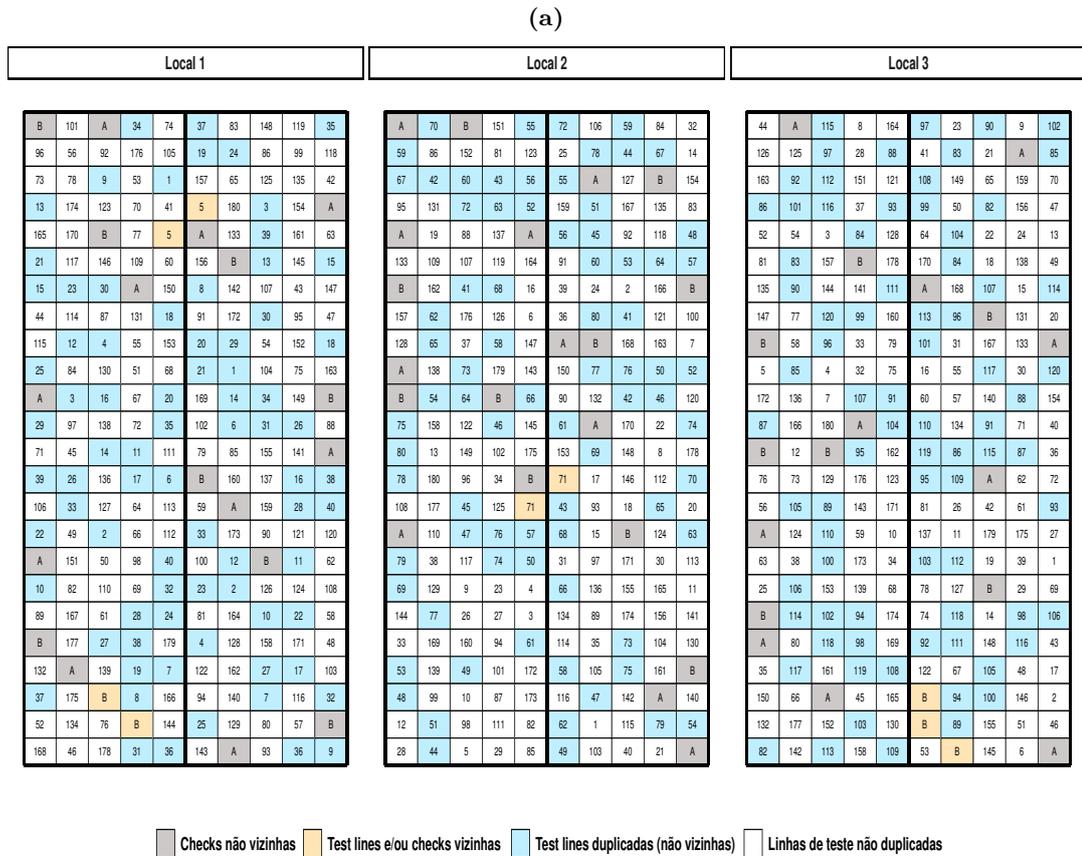


Figura 3.1. Croquis do p -rep 3 descrito na Tabela 3.1 para o cenário 1 em que foram considerados $p=22\%$, 140 *test lines* não duplicadas, 40 *test lines* duplicadas, 2 *checks* repetidas 10 vezes para $\phi_g = 0, 7$. Em (a) os delineamentos foram gerados considerando a suposição de dependência entre os tratamentos (*test lines* e *checks*) em (b) os delineamentos foram gerados considerando os tratamentos independentes.

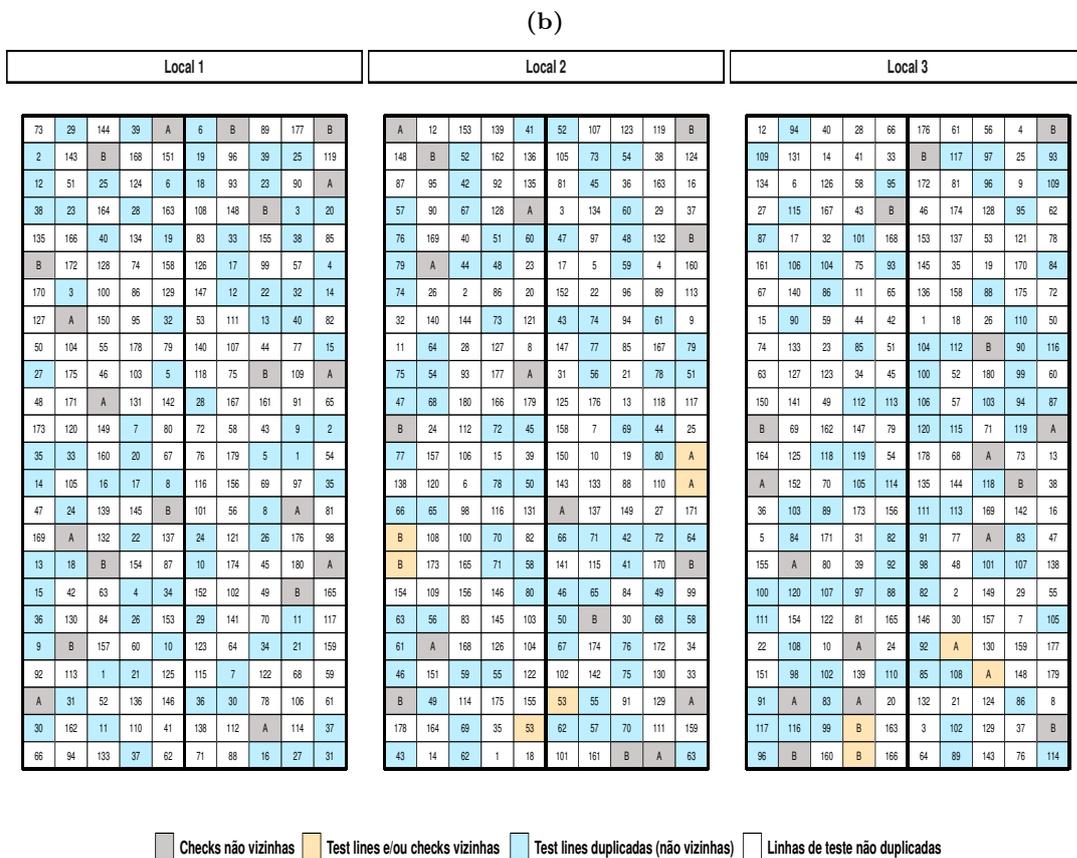
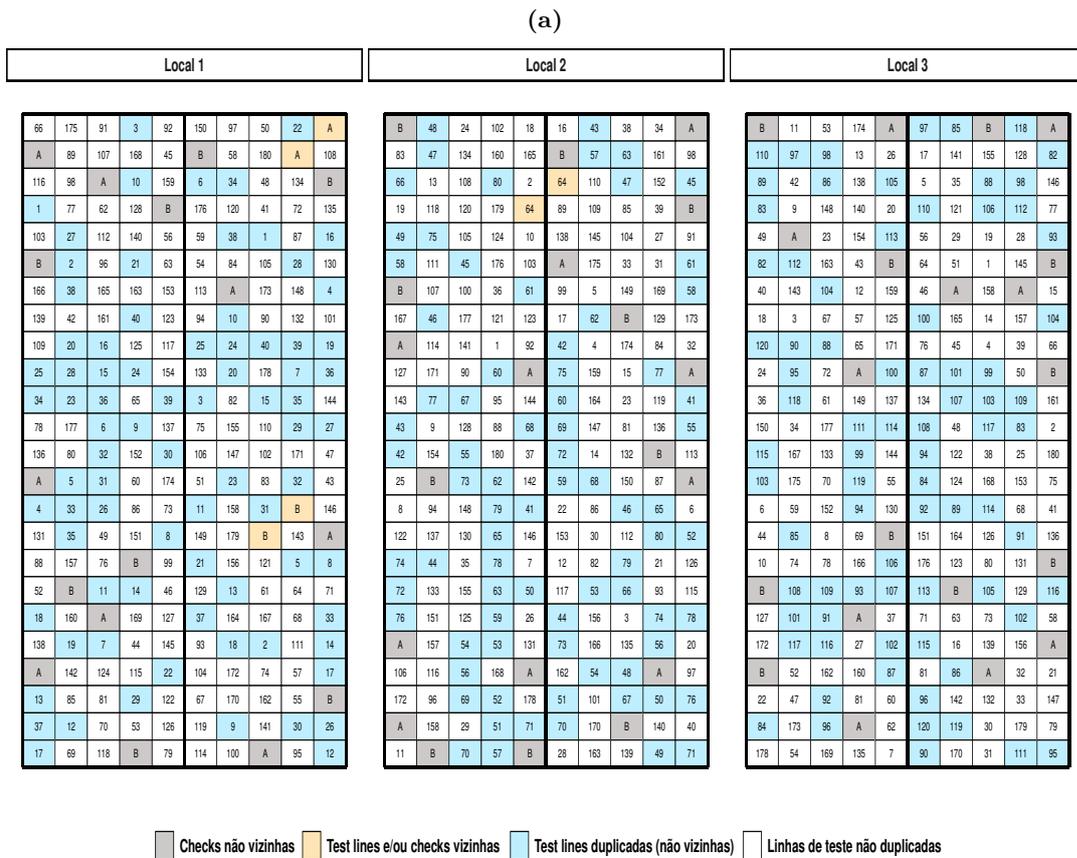


Figura 3.2. Croquis do p -rep 3 descrito na Tabela 3.1 para o cenário 2 em que foram considerados $p=22\%$, 140 *test lines* não duplicadas, 40 *test lines* duplicadas, 2 *checks* repetidas 10 vezes para $\phi_g = 0,7$. Em (a) os delineamentos foram gerados considerando a suposição de dependência entre os tratamentos (*test lines* e *checks*) em (b) os delineamentos foram gerados considerando os tratamentos independentes.

As Figuras 3.3 a 3.5 apresentam os resultados obtidos do estudo de simulação das médias e dos desvios padrão das medidas do RRG, da precisão genética e do sucesso da seleção da análise conjunta e individual dos locais para os delineamentos ótimos obtidos considerando as suposições de dependência ou independência entre os tratamentos para o primeiro e segundo cenários, respectivamente. Além disso, também são apresentados os resultados das medidas utilizadas quando não é considerada a matriz de relacionamento no modelo de análise. Note que, graficamente em termos de média, as suposições com relação à obtenção dos delineamentos pouco interferem nos resultados apresentados. Em relação à comparação entre os delineamentos (i) os maiores resultados encontrados para o ganho genético, precisão genética e o sucesso de seleção no primeiro cenário e no segundo cenário foram obtidos com mais frequência para o p -rep 5 no Local 3, sendo: 0,95, 0,92, 72%, 0,91, 0,88 e 66% no segundo, respectivamente. As Figuras A.9-A.14 (localizadas no Apêndice) apresentam as porcentagens da superioridade de cada um dos delineamentos, considerando todas as medidas de comparação e em ambos os cenários e constatou-se que em todos os casos (locais, análises individual e conjunta e cenários): (ii) o delineamento *grid-plot* apresentou as menores porcentagens de superioridade em relação ao delineamento p -rep e o delineamento p -rep 5 apresentou as maiores porcentagens de superioridade; (iii) quanto maior a razão entre variância genética e residual dentro dos locais, menor é a porcentagem de superioridade do *grid-plot* em relação a qualquer delineamento p -rep; (iv) quando $p=22\%$ percebe-se que mesmo apresentando diferentes quantidades de *checks* e diferentes números de repetição, não há um comportamento padrão relacionado à superioridade do delineamento p -rep. Cullis et al. (2006), Santos (2017), Goes (2020) e Hoefler et al. (2020), também, observaram a superioridade do p -rep em relação ao *grid-plot*.

Ainda em relação às Figuras 3.3 a 3.5, para os valores atribuídos aos parâmetros de variâncias associados aos efeitos genéticos e residuais, observou-se que: (i) quanto maiores são os valores atribuídos a esses parâmetros, maiores são os resultados obtidos para as medidas de comparação; (ii) quanto maior a razão entre variância genética e residual dentro dos locais (cenário 1) maiores são os resultados obtidos para as medidas utilizadas, sendo observados os maiores resultados no Local 3; (iii) quando as razões das variâncias genéticas e residuais são iguais (cenário 2), os locais tendem a apresentar resultados próximos.

Percebeu-se que em relação ao primeiro cenário, os Locais 2 e 3 para os quais foram fixados valores iguais para todos os parâmetros de variância, exceto para o parâmetro associado à variância genética, quanto maior é esta variância, maiores são as medidas de comparação. Outros estudos também observaram resultados análogos (Cullis et al., 2006; Santos, 2017; Goes, 2020; Sermarini et al., 2020). No cenário 2, consideraram-se diferentes valores para as variâncias genéticas e residuais dos locais, porém, a mesma razão entre estas variâncias. Neste caso, observou-se que os valores das medidas de comparação são próximos em todos os locais. Dessa forma, percebe-se que o ganho genético, precisão genética e o sucesso de seleção são afetados pela variância genética de forma mais evidente se os valores dos outros parâmetros são iguais. Porém, deve-se considerar que, neste cenário, todas as variâncias residuais dos locais foram maiores do que as variâncias genéticas e estudos revelam que esta característica pode afetar o ganho genético.

Cruz Neto et al. (2016) observaram em estudos realizados com maracujá que nos casos em que a variância genética é menor do que residual, a herdabilidade tende a apresentar baixos resultados ocasionando menores ganhos genéticos. de Oliveira et al. (2015) observaram em uma população de seringueira, a predominância da variância genética sobre a residual indicando alta variabilidade genética o que resulta em altos ganhos genético. Santos (2017) analisou 21 experimentos de cana-de-açúcar fornecidos pelo Centro de Tecnologia Canavieira (CTC - Empresa Brasileira de Cana-de-Açúcar) e destacou que quando a variância genética é maior que a variância residual, há uma melhor seleção das *test lines*. Diante do exposto, observa-se que os resultados encontrados neste estudo corroboram com outros disponíveis na literatura.

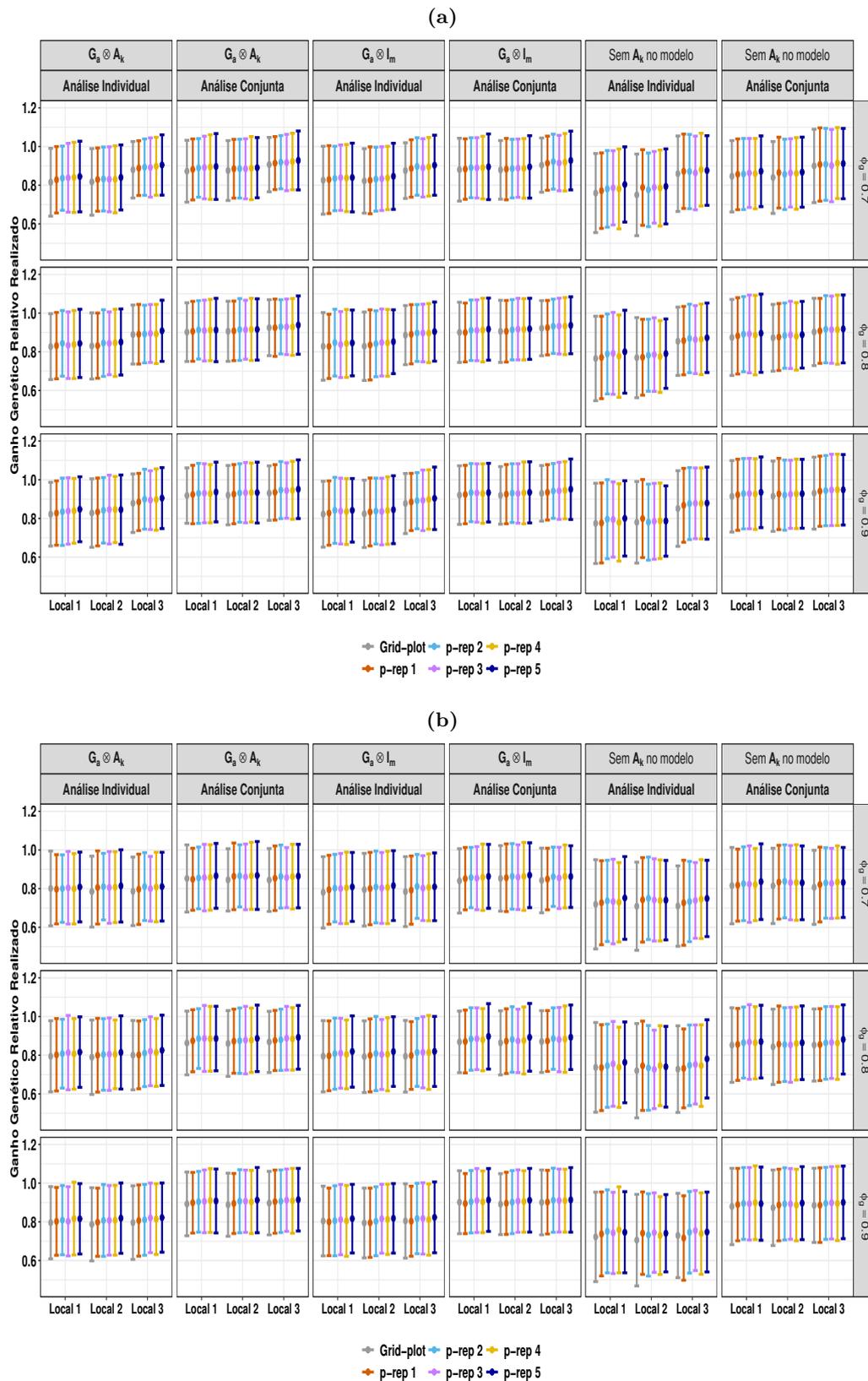


Figura 3.3. Média \pm desvio padrão do Ganho Genético Relativo Realizado, da análise individual e conjunta, para os delineamentos gerados considerando o primeiro (a) e segundo (b) cenário. O painel intitulado “Sem A_k no modelo” refere-se a não inclusão da matriz de relacionamento no modelo de análise. O símbolo \otimes é o produto direto da matriz de variância genética aditiva G_a e a matriz de relacionamento A_k .

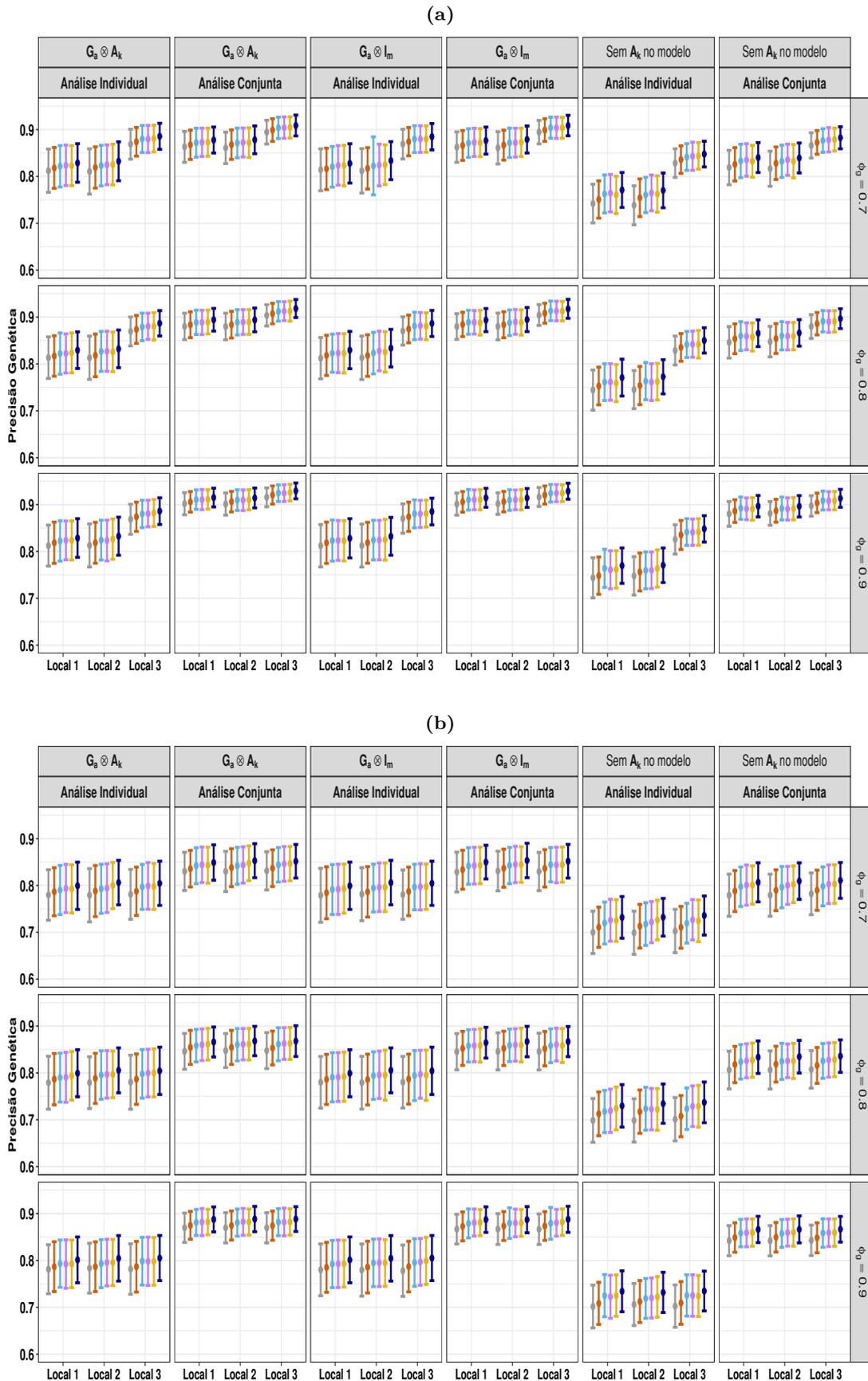


Figura 3.4. Média \pm desvio padrão da precisão genética, da análise individual e conjunta, para os delineamentos gerados considerando o primeiro (a) e segundo (b) cenário. O painel intitulado “Sem \mathbf{A}_k no modelo” refere-se a não inclusão da matriz de relacionamento no modelo de análise. O símbolo \otimes é o produto direto da matriz de variância genética aditiva \mathbf{G}_a e a matriz de relacionamento \mathbf{A}_k .

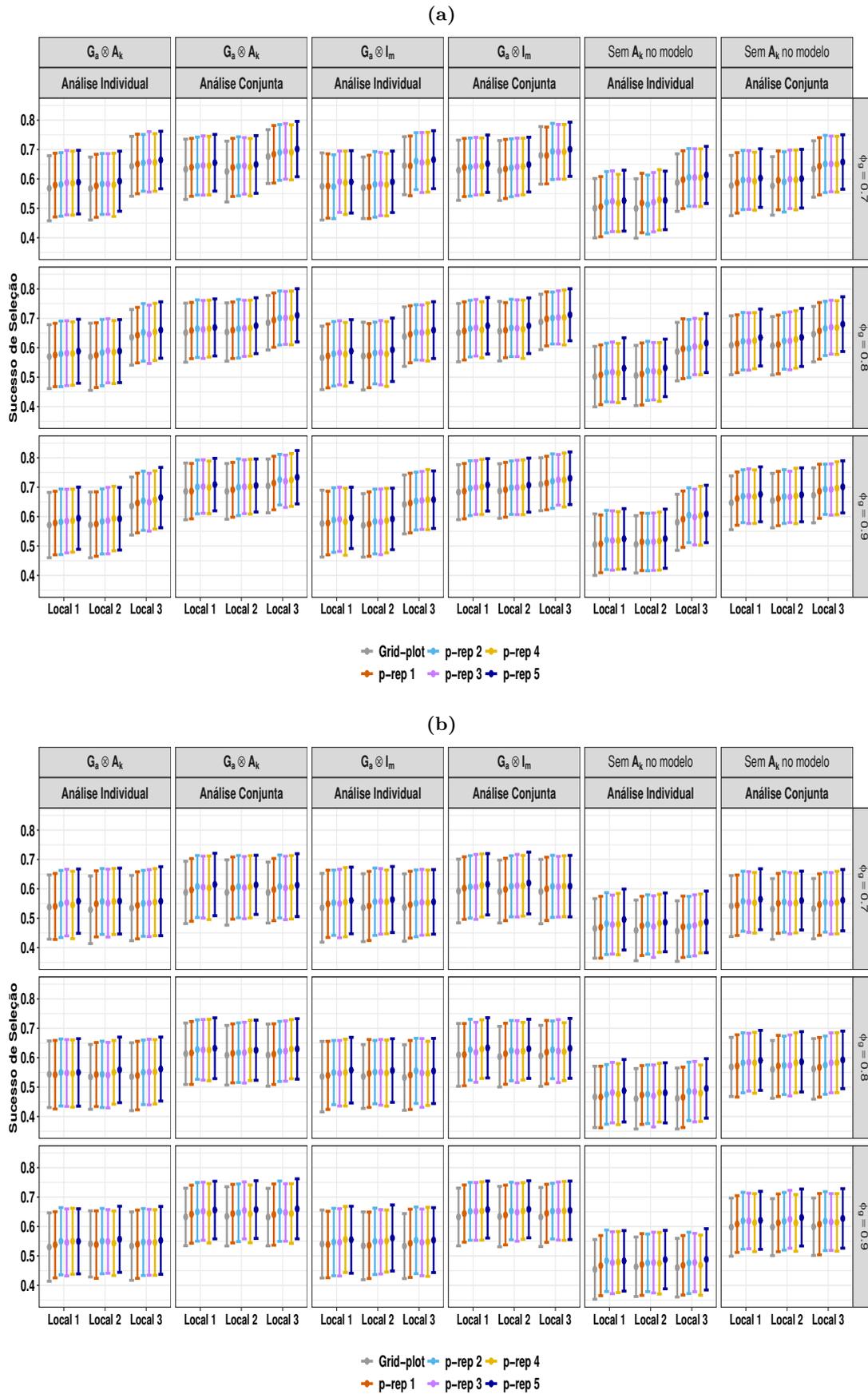


Figura 3.5. Média \pm desvio padrão do sucesso de seleção, da análise individual e conjunta, para os delineamentos gerados considerando o primeiro (a) e segundo (b) cenário. O painel intitulado “Sem \mathbf{A}_k no modelo” refere-se a não inclusão da matriz de relacionamento no modelo de análise. O símbolo \otimes é o produto direto da matriz de variância genética aditiva \mathbf{G}_a e a matriz de relacionamento \mathbf{A}_k .

De modo geral, a análise conjunta apresentou valores mais altos relacionando às medidas de comparação em todos os cenários e locais, como já demonstrado em Santos (2017) e em Tanaka (2020). Além disso, de forma evidente, percebe-se que quando a matriz de relacionamento não é inserida no modelo de análise, os resultados são menores em todas as medidas de comparação, e, também, foi observada neste caso uma maior variabilidade. Os resultados indicaram que quando se utiliza a matriz de parentesco no modelo de análise há um ganho de pelo menos uma *test lines* verdadeiramente boa que é selecionada pelo modelo. Outros estudos demonstram que a precisão genética e o ganho genético melhoram quando a matriz de parentesco é inserida nas análises (Pérez-Rodríguez et al., 2017; Borges da Silva et al., 2021).

Nas Figuras 3.6 a 3.13, são apresentadas as médias das estimativas dos componentes de variância com os seus respectivos erros padrão provenientes dos modelos ajustados para análise individual e conjunta dos locais dos delineamentos ótimos obtidos para cenários 1 e 2, respectivamente. Observa-se novamente que as suposições consideradas para gerar os delineamentos não influenciaram nos valores relacionados aos componentes de variância. Notou-se que, em todos os cenários e locais a variância entre linhas tende a ser superestimada (Figuras 3.6 (a) e 3.10 (a)). Às variâncias que apresentaram maior proximidade entre os valores estimados pelos modelos ajustados e o seu verdadeiro valor na análise conjunta, foram: de colunas (Figuras 3.6 (b) e 3.10 (b)), dos efeitos genéticos em cada local (Figura 3.7 (a) e 3.11 (a)), residuais por local (Figuras 3.7 (b) e 3.11 (b)), correlação genética entre as mesmas linhagens aplicadas em diferentes locais (Figuras 3.9 e 3.13 (e)). Os componentes de variâncias da correlação residual entre linhas e entre colunas no mesmo local apresentaram subestimação (Figuras 3.9 (a) e (b) e 3.9 (a) e (b)).

Curiosamente, observou-se que quanto maior a variância residual, maior é o erro padrão. Dessa forma, em todos os cenários, no Local 3 os valores obtidos para as estimativas dos componentes de variância são menos precisos. Possivelmente, esse fato é uma consequência dos altos valores atribuídos para os parâmetros de variâncias genética e residual, respectivamente. Essa variabilidade diminui à medida que ϕ_g aumenta. Por fim, observa-se um aumento nos erros padrão quando a matriz de relacionamento não é inserida no modelo de análise, indicando que as estimativas dos componentes de variância são menos precisas tanto nas análises individuais quanto na análise conjunta. Borges da Silva et al. (2021) observaram que os componentes de variância são menos enviesados quando a matriz de relacionamento é considerada.

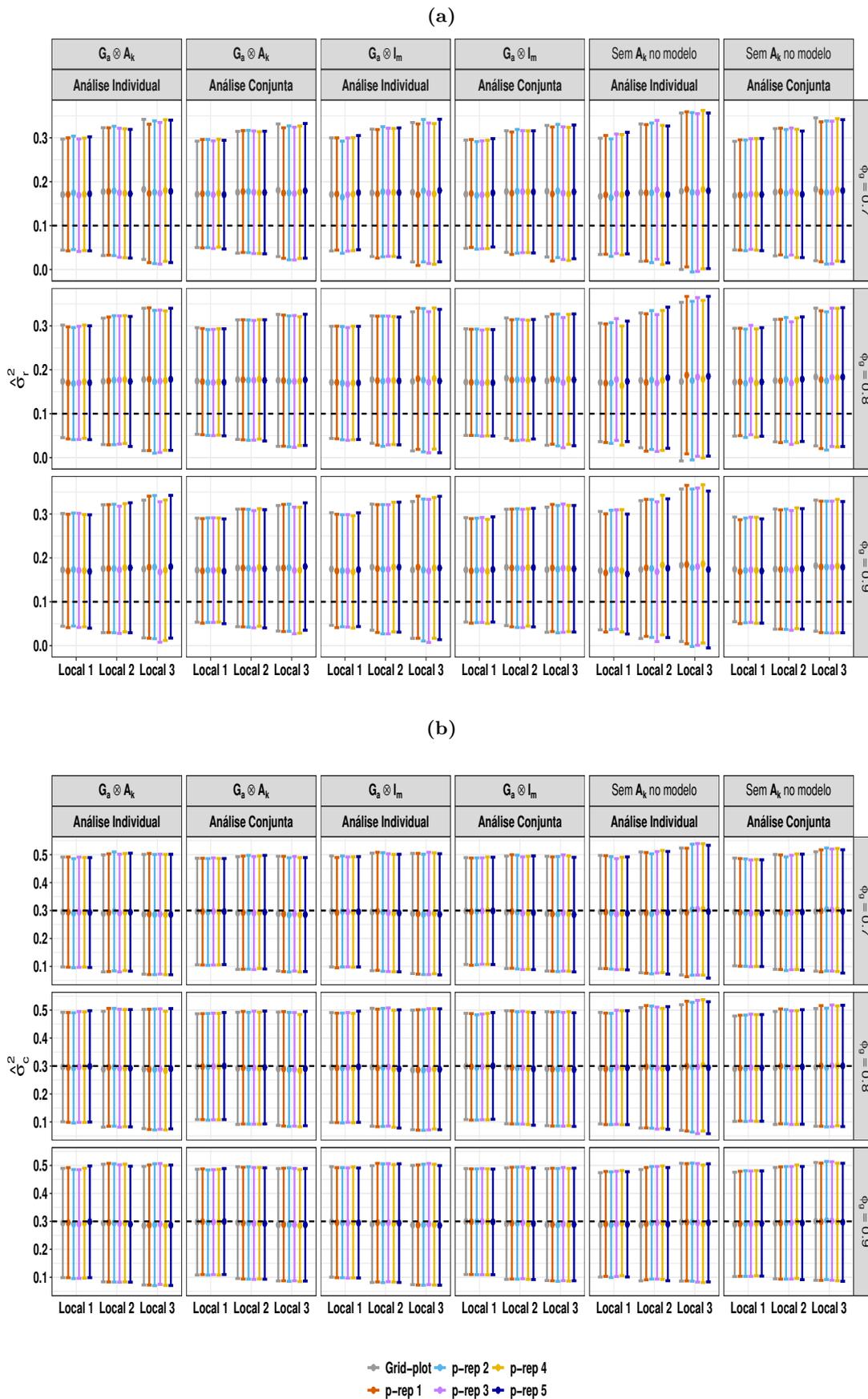


Figura 3.6. Média \pm erro padrão das estimativas dos parâmetros relacionados às variâncias entre linhas (a) e colunas (b), da análise individual e conjunta, para os delineamentos gerados considerando o primeiro cenário. O painel intitulado “Sem \mathbf{A}_k no modelo” refere-se a não inclusão da matriz de relacionamento no modelo de análise. O símbolo \otimes é o produto direto da matriz de variância genética aditiva \mathbf{G}_a e a matriz de relacionamento \mathbf{A}_k .

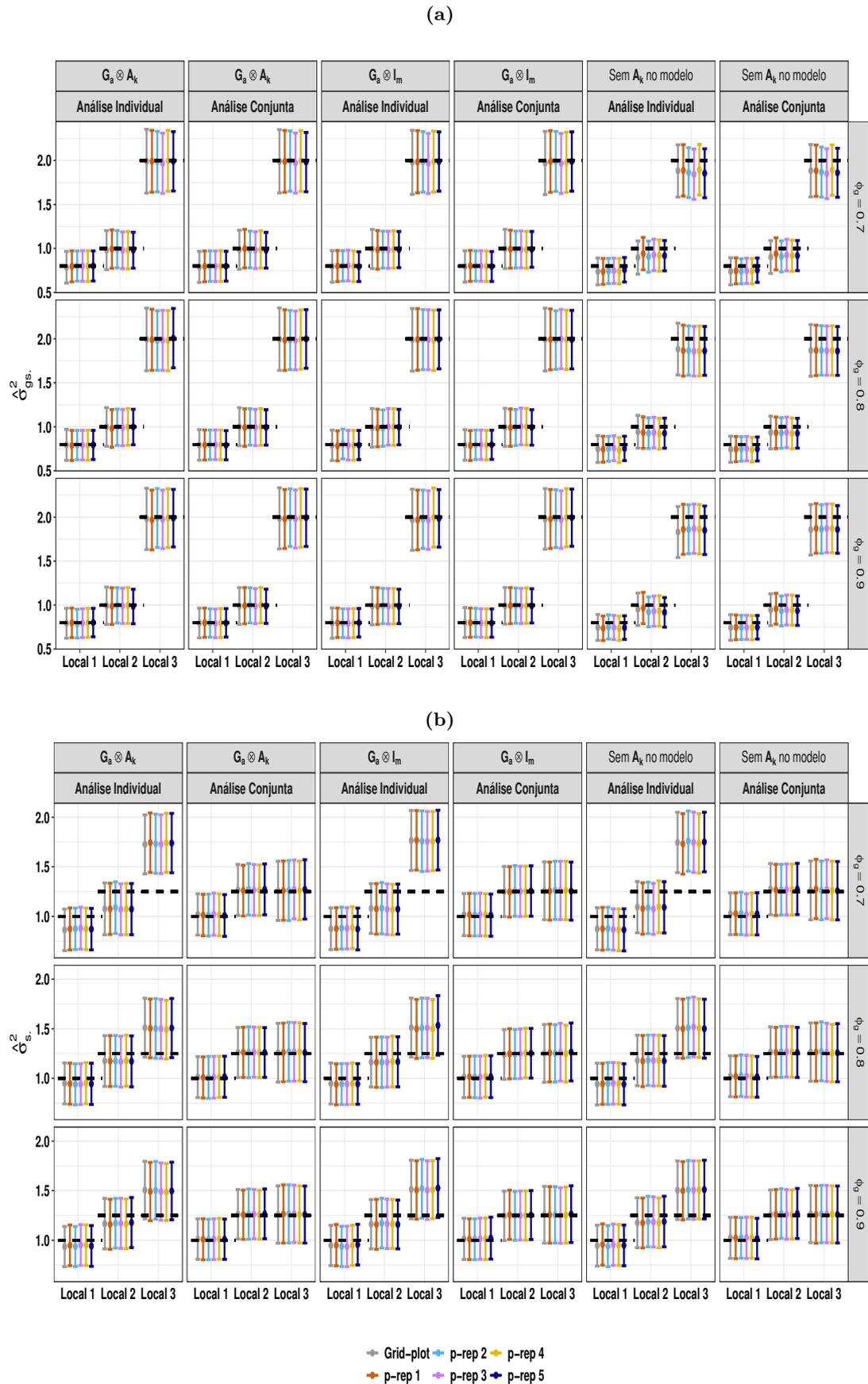


Figura 3.7. Média \pm erro padrão das estimativas dos parâmetros relacionados às variâncias genéticas dos diferentes locais (a) e residual dos locais (b), da análise individual e conjunta, para os delineamentos gerados considerando o primeiro cenário. O painel intitulado “Sem \mathbf{A}_k no modelo” refere-se a não inclusão da matriz de relacionamento no modelo de análise. O símbolo \otimes é o produto direto da matriz de variância genética aditiva \mathbf{G}_a e a matriz de relacionamento \mathbf{A}_k .

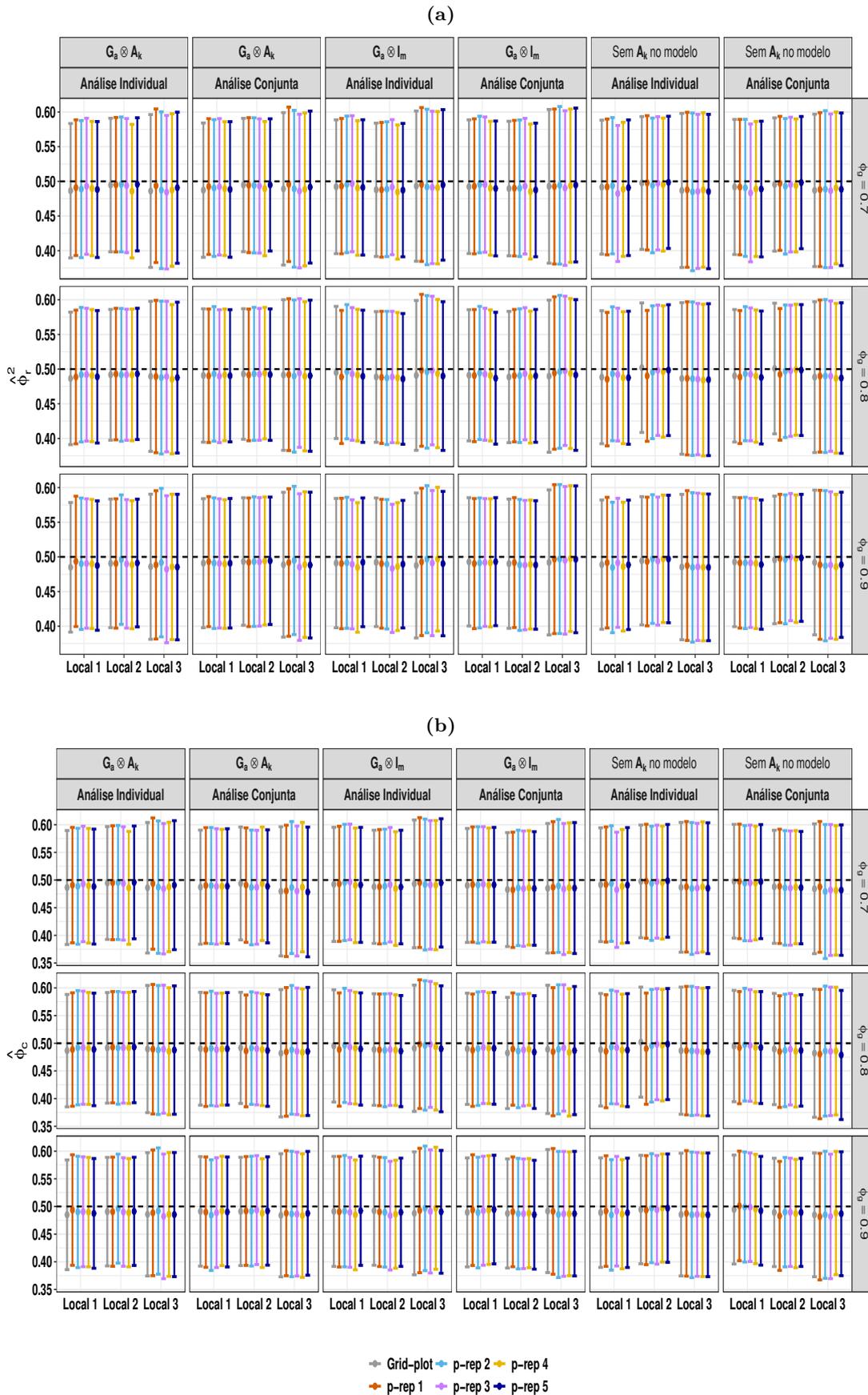


Figura 3.8. Média \pm erro padrão das estimativas dos parâmetros relacionados à correlação residual entre linhas (a), e colunas (b), da análise individual e conjunta, para os delineamentos gerados considerando o primeiro cenário. O painel intitulado “Sem \mathbf{A}_k no modelo” refere-se a não inclusão da matriz de relacionamento no modelo de análise. O símbolo \otimes é o produto direto da matriz de variância genética aditiva \mathbf{G}_a e a matriz de relacionamento \mathbf{A}_k .

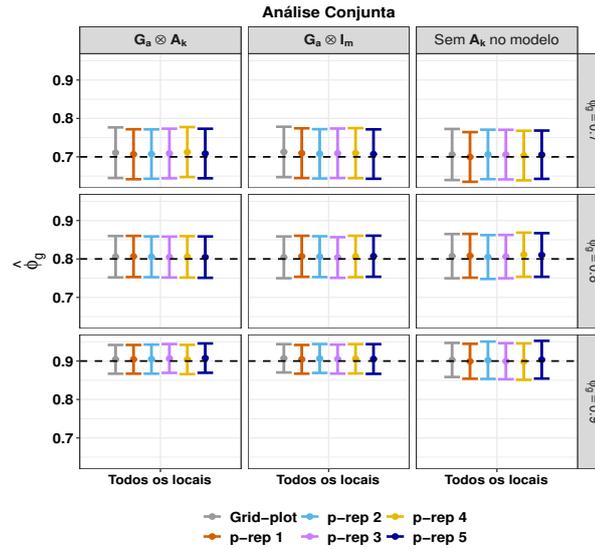


Figura 3.9. Média \pm erro padrão das estimativas dos parâmetros relacionados à correlação genética (entre efeitos da mesma *test line* em diferentes locais), da análise conjunta, para os delineamentos gerados considerando o primeiro cenário. O painel intitulado “Sem \mathbf{A}_k no modelo” refere-se a não inclusão da matriz de relacionamento no modelo de análise. O símbolo \otimes é o produto direto da matriz de variância genética aditiva \mathbf{G}_a e a matriz de relacionamento \mathbf{A}_k .

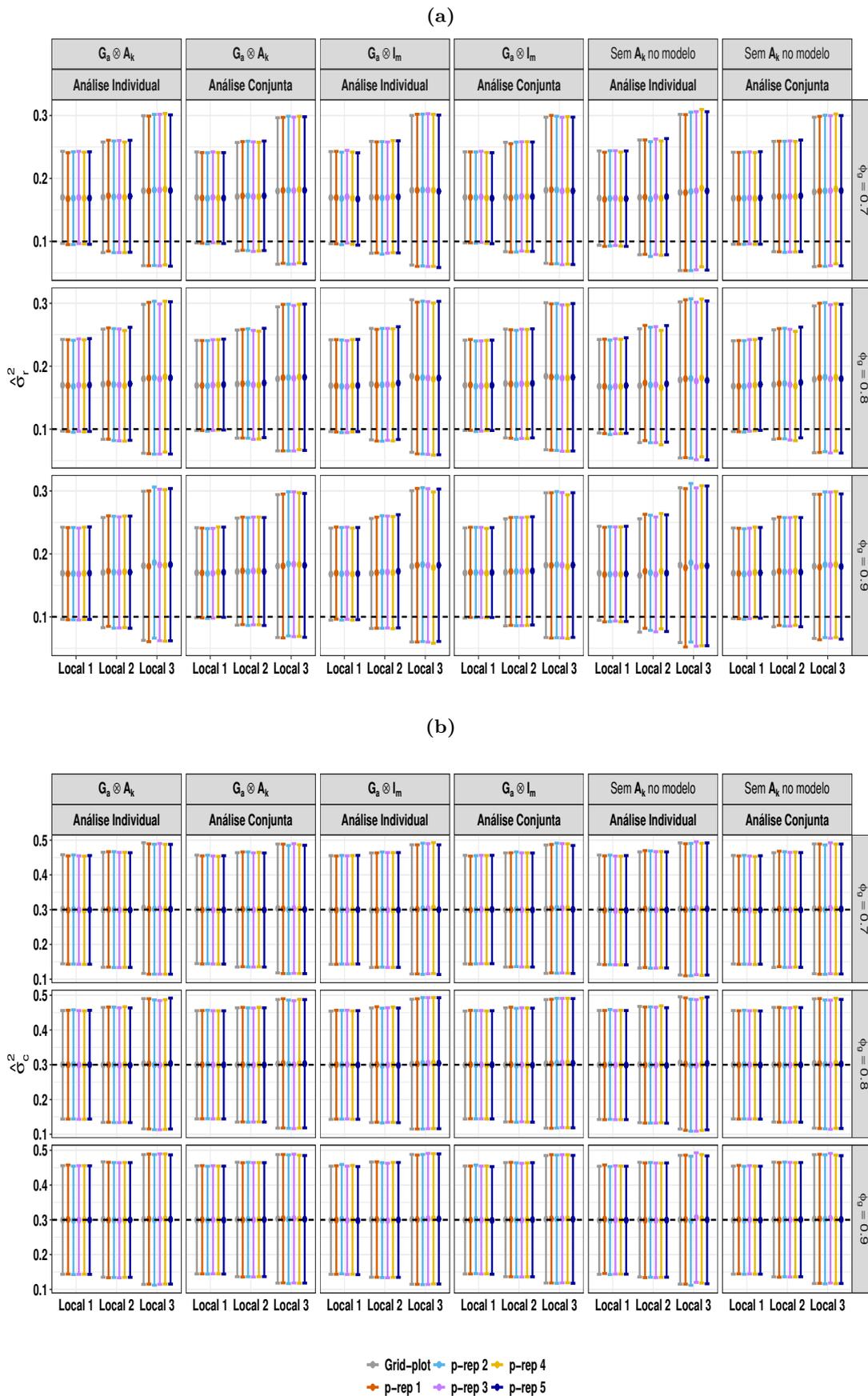


Figura 3.10. Média \pm erro padrão das estimativas dos parâmetros relacionados às variâncias entre linhas (a) e colunas (b), da análise individual e conjunta, para os delineamentos gerados considerando o segundo cenário. O painel intitulado “Sem \mathbf{A}_k no modelo” refere-se a não inclusão da matriz de relacionamento no modelo de análise. O símbolo \otimes é o produto direto da matriz de variância genética aditiva \mathbf{G}_a e a matriz de relacionamento \mathbf{A}_k .

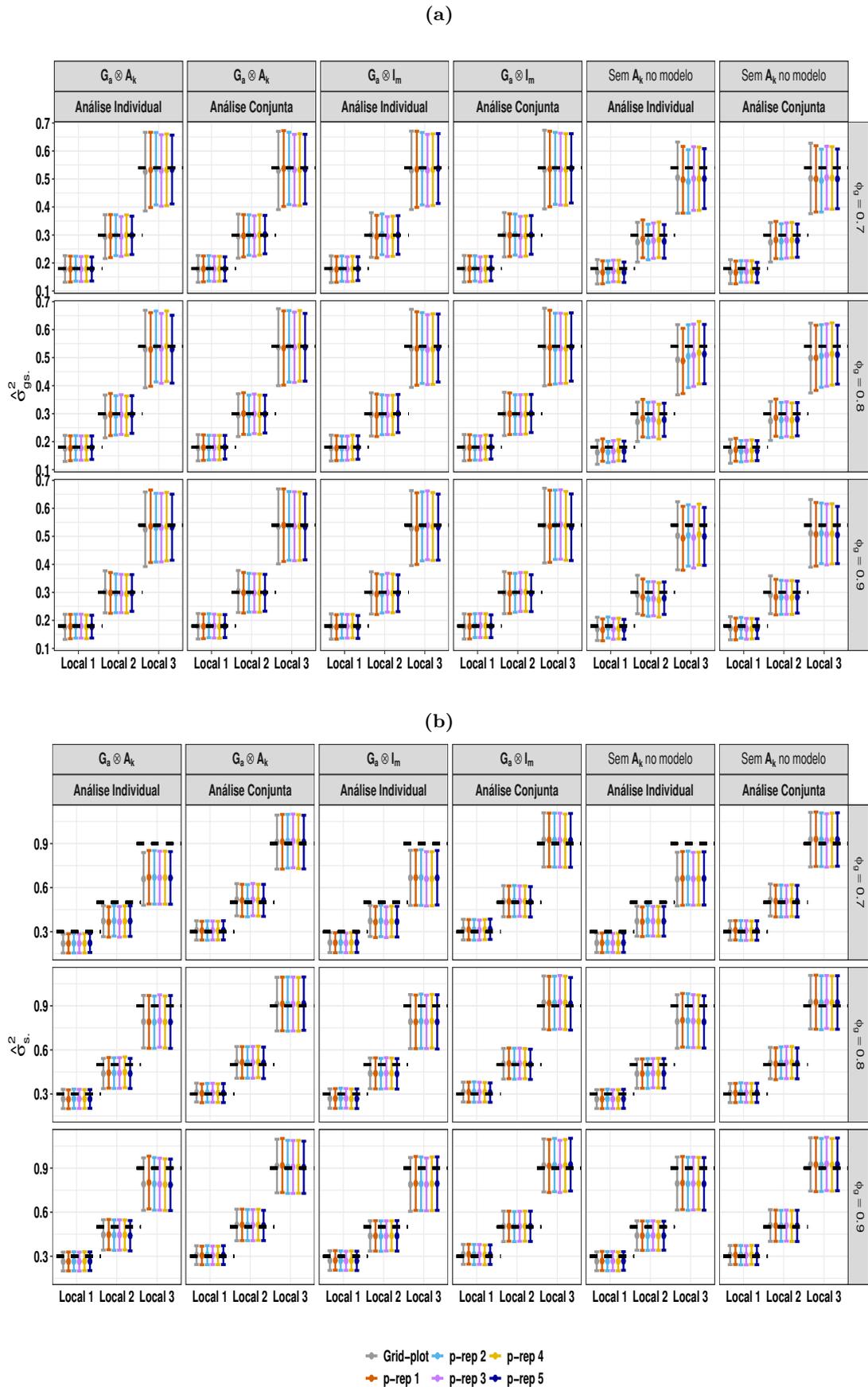


Figura 3.11. Média \pm erro padrão das estimativas dos parâmetros relacionados às variâncias genéticas dos diferentes locais (a) e residual dos locais (b), da análise individual e conjunta, para os delineamentos gerados considerando o segundo cenário. O painel intitulado “Sem A_k no modelo” refere-se a não inclusão da matriz de relacionamento no modelo de análise. O símbolo \otimes é o produto direto da matriz de variância genética aditiva G_a e a matriz de relacionamento A_k .

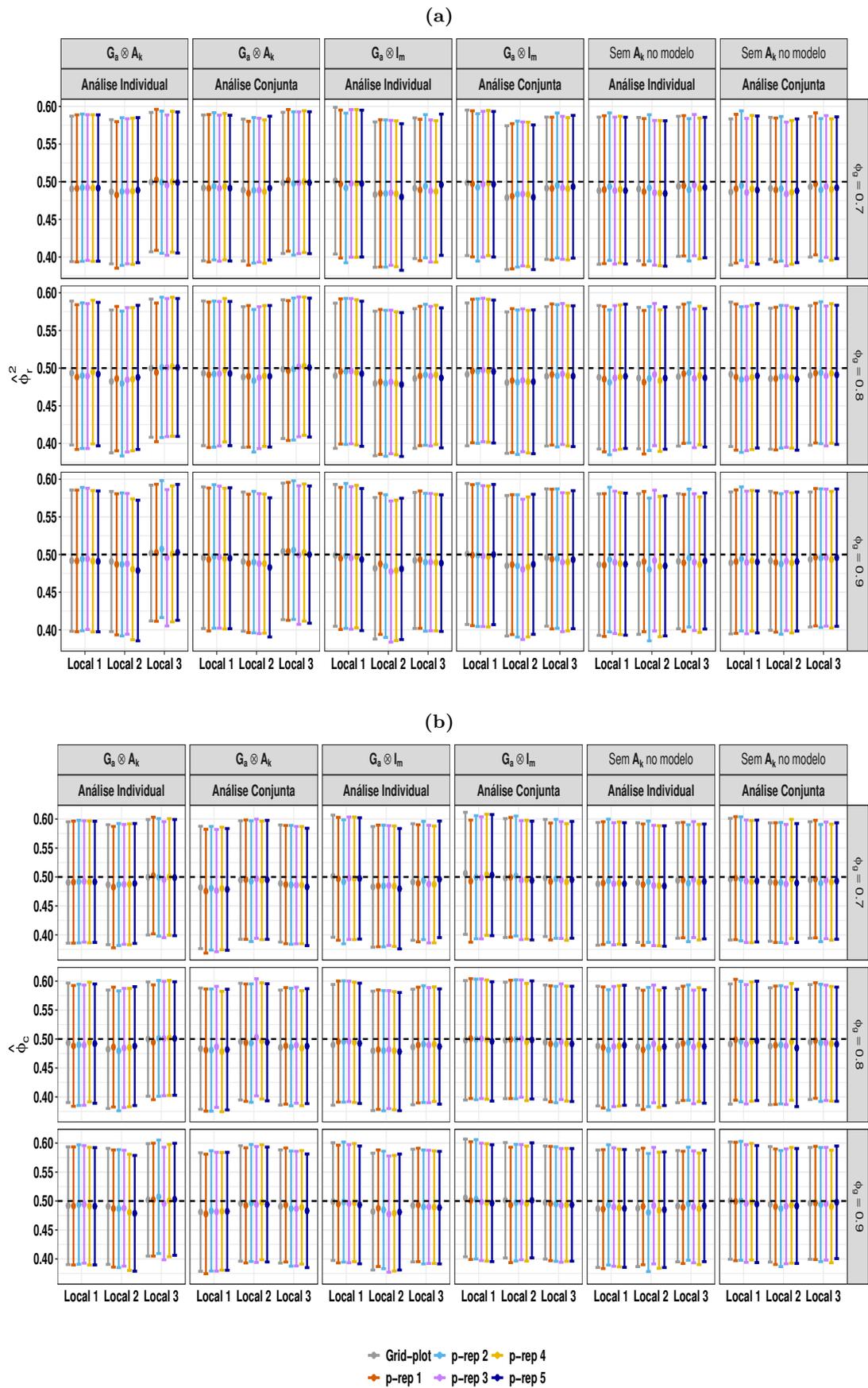


Figura 3.12. Média \pm erro padrão das estimativas dos parâmetros relacionados à correlação residual entre linhas (a), e colunas (b), da análise individual e conjunta, para os delineamentos gerados considerando o segundo cenário. O painel intitulado “Sem A_k no modelo” refere-se a não inclusão da matriz de relacionamento no modelo de análise. O símbolo \otimes é o produto direto da matriz de variância genética aditiva G_a e a matriz de relacionamento A_k .

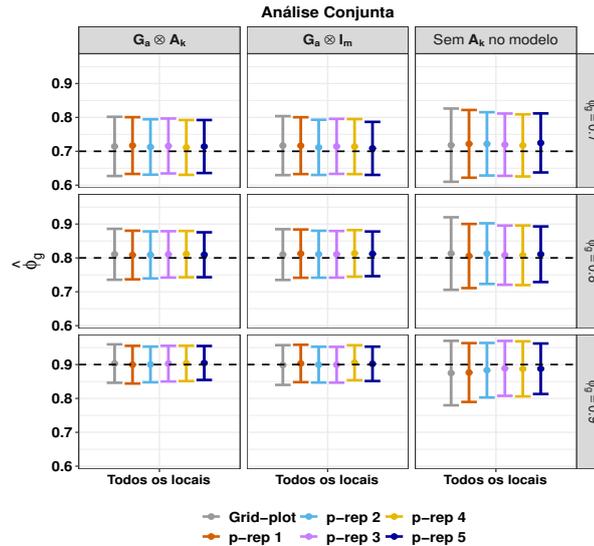


Figura 3.13. Média \pm erro padrão das estimativas dos parâmetros relacionados à correlação genética (entre efeitos da mesma *test line* em diferentes locais), da análise conjunta, para os delineamentos gerados considerando o segundo cenário. O painel intitulado “Sem \mathbf{A}_k no modelo” refere-se a não inclusão da matriz de relacionamento no modelo de análise. O símbolo \otimes é o produto direto da matriz de variância genética aditiva \mathbf{G}_a e a matriz de relacionamento \mathbf{A}_k .

De acordo com os resultados apresentados nas Tabelas 3.3 e 3.4, (i) as *test lines* duplicadas apresentaram os menores valores médios referentes às médias das variâncias das diferenças entre pares de tratamentos quando comparados às *test lines* não duplicadas, o que é esperado, uma vez que a variabilidade do material genético dentro do grupo referente as *test lines* duplicadas é menor; (ii) a variação da porcentagem $p\%$ de *test lines* duplicadas pouco afeta os valores dessa medida, porém observa-se uma tendência levemente crescente nos resultados à medida que $p\%$ aumenta; (iii) comparando os dois cenários percebe-se que a razão entre as variâncias genética e residual por local pouco afetam os valores dessa medida. Porém, a variância residual dos locais tende a afetá-las, quanto maior é essa variância, maiores são os valores das médias das variâncias para as diferenças entre pares de tratamentos; (v) quanto maior a correlação genética entre os mesmos genótipos aplicados em diferentes locais menores são os valores. Ao relacionar esses resultados com o RRG, observa-se que há um aumento do ganho genético à medida que ϕ_g aumenta e quando $\phi_g = 0,9$ as médias das variâncias para as diferenças entre pares de tratamentos são menores quando comparadas aos outros valores assumidos para este parâmetro.

Tabela 3.3. Média das variâncias das diferenças entre pares de tratamentos para os delineamentos ótimos considerando as suposições de dependência e independência entre os tratamentos para o primeiro cenário, em que a razão entre a variância genética aditiva e a variância residual ($\sigma_{g_s}^2/\sigma_s^2$) dos Locais 1 e 2 são iguais e no Local 3 essa razão é o dobro em relação aos Locais 1 e 2.

	Local 1		Local 2		Local 3	
Relação de dependência entre os tratamentos						
$\phi_g = 0,7$						
	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas
Grid-plot	-	0,537	-	0,675	-	0,676
<i>p</i> -rep 1	0,304	0,551	0,491	0,676	0,486	0,679
<i>p</i> -rep 2	0,343	0,554	0,476	0,678	0,469	0,683
<i>p</i> -rep 3	0,346	0,555	0,454	0,689	0,452	0,694
<i>p</i> -rep 4	0,349	0,555	0,451	0,691	0,449	0,695
<i>p</i> -rep 5	0,355	0,563	0,440	0,707	0,443	0,709
$\phi_g = 0,8$						
	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas
Grid-plot	-	0,474	-	0,533	-	0,534
<i>p</i> -rep 1	0,274	0,481	0,429	0,591	0,425	0,594
<i>p</i> -rep 2	0,301	0,485	0,408	0,597	0,407	0,599
<i>p</i> -rep 3	0,302	0,486	0,401	0,600	0,401	0,603
<i>p</i> -rep 4	0,305	0,486	0,394	0,605	0,392	0,606
<i>p</i> -rep 5	0,312	0,491	0,391	0,614	0,387	0,618
$\phi_g = 0,9$						
	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas
Grid-plot	-	0,385	-	0,482	-	0,483
<i>p</i> -rep 1	0,220	0,388	0,326	0,480	0,327	0,481
<i>p</i> -rep 2	0,240	0,393	0,322	0,485	0,322	0,486
<i>p</i> -rep 3	0,242	0,394	0,311	0,490	0,315	0,490
<i>p</i> -rep 4	0,245	0,393	0,311	0,490	0,313	0,491
<i>p</i> -rep 5	0,249	0,399	0,311	0,499	0,311	0,500
Relação de independência entre os tratamentos						
$\phi_g = 0,7$						
	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas
Grid-plot	-	0,541	-	0,679	-	0,682
<i>p</i> -rep 1	0,307	0,551	0,485	0,676	0,487	0,681
<i>p</i> -rep 2	0,346	0,554	0,466	0,685	0,472	0,686
<i>p</i> -rep 3	0,343	0,557	0,457	0,689	0,454	0,692
<i>p</i> -rep 4	0,346	0,557	0,447	0,693	0,452	0,693
<i>p</i> -rep 5	0,356	0,561	0,445	0,703	0,448	0,702
$\phi_g = 0,8$						
	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas
Grid-plot	-	0,474	-	0,593	-	0,597
<i>p</i> -rep 1	0,272	0,481	0,418	0,591	0,422	0,593
<i>p</i> -rep 2	0,303	0,485	0,408	0,598	0,409	0,598
<i>p</i> -rep 3	0,302	0,486	0,396	0,604	0,398	0,605
<i>p</i> -rep 4	0,305	0,486	0,394	0,604	0,396	0,607
<i>p</i> -rep 5	0,313	0,491	0,389	0,615	0,389	0,617
$\phi_g = 0,9$						
	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas
Grid-plot	-	0,387	-	0,483	-	0,485
<i>p</i> -rep 1	0,221	0,390	0,320	0,482	0,325	0,483
<i>p</i> -rep 2	0,245	0,390	0,326	0,482	0,326	0,483
<i>p</i> -rep 3	0,242	0,395	0,315	0,490	0,316	0,491
<i>p</i> -rep 4	0,246	0,393	0,313	0,490	0,314	0,492
<i>p</i> -rep 5	0,251	0,397	0,311	0,498	0,311	0,499

Tabela 3.4. Média das variâncias das diferenças entre pares de tratamentos para os delineamentos ótimos considerando as suposições de dependência e independência entre os tratamentos para o segundo cenário em que a razão entre a variância genética aditiva e a variância residual ($\sigma_{g_s}^2/\sigma_{s_s}^2$) é a mesma razão para os três ambientes.

	Local 1		Local 2		Local 3	
Relação de dependência entre os tratamentos						
$\phi_g = 0,7$						
	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas
Grid-plot	-	0,164	-	0,273	-	0,490
<i>p</i> -rep 1	0,095	0,167	0,197	0,273	0,354	0,490
<i>p</i> -rep 2	0,104	0,169	0,187	0,278	0,337	0,497
<i>p</i> -rep 3	0,104	0,169	0,183	0,279	0,327	0,501
<i>p</i> -rep 4	0,107	0,168	0,179	0,281	0,321	0,502
<i>p</i> -rep 5	0,109	0,171	0,176	0,289	0,320	0,514
$\phi_g = 0,8$						
	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas
Grid-plot	-	0,142	-	0,238	-	0,427
<i>p</i> -rep 1	0,084	0,145	0,168	0,238	0,302	0,428
<i>p</i> -rep 2	0,093	0,147	0,165	0,242	0,294	0,434
<i>p</i> -rep 3	0,092	0,147	0,161	0,243	0,282	0,437
<i>p</i> -rep 4	0,094	0,147	0,157	0,244	0,282	0,439
<i>p</i> -rep 5	0,094	0,150	0,156	0,250	0,279	0,451
$\phi_g = 0,9$						
	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas
Grid-plot	-	0,115	-	0,192	-	0,346
<i>p</i> -rep 1	0,066	0,117	0,130	0,193	0,235	0,346
<i>p</i> -rep 2	0,073	0,118	0,130	0,195	0,232	0,350
<i>p</i> -rep 3	0,074	0,119	0,126	0,197	0,227	0,353
<i>p</i> -rep 4	0,075	0,119	0,125	0,198	0,224	0,356
<i>p</i> -rep 5	0,076	0,121	0,125	0,202	0,224	0,363
Relação de independência entre os tratamentos						
$\phi_g = 0,7$						
	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas
Grid-plot	-	0,164	-	0,274	-	0,493
<i>p</i> -rep 1	0,095	0,167	0,197	0,274	0,349	0,493
<i>p</i> -rep 2	0,105	0,169	0,193	0,276	0,342	0,497
<i>p</i> -rep 3	0,106	0,169	0,185	0,279	0,333	0,500
<i>p</i> -rep 4	0,105	0,169	0,182	0,279	0,324	0,504
<i>p</i> -rep 5	0,109	0,171	0,181	0,285	0,324	0,510
$\phi_g = 0,8$						
	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas
Grid-plot	-	0,143	-	0,239	-	0,431
<i>p</i> -rep 1	0,083	0,146	0,170	0,240	0,301	0,431
<i>p</i> -rep 2	0,091	0,148	0,165	0,243	0,298	0,435
<i>p</i> -rep 3	0,092	0,147	0,161	0,243	0,289	0,438
<i>p</i> -rep 4	0,093	0,148	0,161	0,244	0,286	0,439
<i>p</i> -rep 5	0,095	0,149	0,158	0,249	0,282	0,448
$\phi_g = 0,9$						
	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas	Duplicadas	Não duplicadas
Grid-plot	-	0,116	-	0,194	-	0,350
<i>p</i> -rep 1	0,066	0,118	0,130	0,195	0,233	0,351
<i>p</i> -rep 2	0,075	0,119	0,130	0,196	0,235	0,353
<i>p</i> -rep 3	0,074	0,119	0,127	0,197	0,230	0,355
<i>p</i> -rep 4	0,075	0,119	0,127	0,198	0,227	0,356
<i>p</i> -rep 5	0,076	0,121	0,125	0,202	0,226	0,363

Os resultados apresentados não mostraram grandes interferências nos valores relacionados ao RRG, a precisão genética, o sucesso da seleção e as estimativas dos componentes de variância ao analisar os delineamentos gerados considerando a suposição de dependência e independência dos tratamentos no modelo para gerar o delineamento. Desse modo, nas condições assumidas neste estudo, observou-se

que qualquer suposição acerca da relação entre os tratamentos podem ser consideradas para a obtenção desses delineamentos em ensaios multi-ambientais. Porém, quanto maior os valores dos parâmetros de variância, associados às variâncias genética e residual dos locais, melhor é a distribuição das *test lines* na área experimental e melhores são os resultados das medidas de comparação. Desta forma, entende-se que em um estudo de simulação, os valores para os parâmetros de variâncias devem ser escolhidos com cautela, pois estes valores podem influenciar no ganho genético e proporcionar uma melhora na seleção. Esses estudos podem auxiliar os programas de melhoramento no processo seletivo para recomendação de novos materiais comerciais (Cruz Neto et al., 2016).

3.4 Conclusão

Neste estudo, foram gerados delineamentos ótimos para ensaios de campo multi-ambientais do tipo *grid-plot* e *p-rep* espacialmente otimizados, para uma área experimental composta de 24 linhas por 10 colunas e 180 *test lines*, dois blocos e três locais. Os delineamentos foram gerados considerando duas suposições para os tratamentos, isto é, considerando a relação de dependência e independência dos tratamentos. A matriz de parentesco utilizada foi proveniente de um estudo de simulação de um programa completo de melhoramento de trigo por 40 anos de seleção fenotípica convencional para rendimento considerando o *pedrigree* das linhagens no estágio PYT do 39º ano e conhecendo seis gerações anteriores. Além disso, foram considerados dois cenários para as variâncias genética e residual. Foram utilizados modelos lineares mistos considerando variação espacial nos erros de parcelas com e sem matriz de parentesco para gerar os delineamentos. Para o estudo de simulação e análise dos dados, foram utilizados modelos lineares mistos com variação espacial nos erros de parcelas e a inclusão da matriz de parentesco. As análises dos conjuntos de dados simulados foram conduzidas de forma individual e conjunta para os ensaios nos três locais.

Os resultados indicaram que não foram observadas diferenças nas medidas de comparação ao considerar as suposições de dependência ou independência entre os efeitos de linhagens no modelo para gerar o delineamento. Em relação às medidas do ganho genético, da precisão genética e do sucesso de seleção apresentaram poucas diferenças dentro de cada cenário e de cada local. No entanto, observou-se que o delineamento *p-rep* sem *checks* apresentou os maiores resultados de ganho genético e qualidade de seleção quando comparado aos demais delineamentos. Dessa forma, pode-se concluir que o número de *checks*, bem como, seus números de repetições, com uma porcentagem $p\%$ de *test lines* duplicadas também podem afetar o ganho genético e a qualidade de seleção. Sendo assim, recomenda-se os delineamentos parcialmente replicados sem *checks* para realizar estudos dessa natureza. Porém, se for necessária a presença de *checks*, recomendam-se utilizar os delineamentos *p-rep* que apresentem não apenas as *test lines*, mas também às variedades de interesse com maior quantidade de diferentes *checks* e menor número de repetição ao invés dos delineamentos *grid-plot*, visto que, esse último apresentou pior desempenho.

Adicionalmente, notou-se que quando são fixados valores iguais para os parâmetros de variâncias associados aos efeitos não genéticos e residuais, quanto maior a variância genética melhores são o ganho genético e a qualidade de seleção. Observou-se, também, que quanto maior a correlação entre os mesmos genótipos aplicados em diferentes locais, melhores são os resultados. Além disso, quando se consideraram valores mais altos para os parâmetros de variâncias genéticas e residual por local, foi possível notar uma melhor aleatorização das variedades em teste na área experimental e melhores resultados para as medidas utilizadas. Por fim, constataram-se melhores resultados quando a matriz de relacionamento é inserida no modelo de análise, como esperado.

Referências

- Atkinson, A. C. e Donev, A. N. (1992). *Optimum experimental designs*.
- Borges da Silva, É. D., Xavier, A., e Faria, M. V. (2021). Joint modeling of genetics and field variation in plant breeding trials using relationship and different spatial methods: A simulation study of accuracy and bias. *Agronomy*, 11(7):1397.
- Brien, C. (2021b). dae: Functions useful in the design and anova of experiments. R package version 3.1-37.
- Butler, D. (2019). od: Generate optimal experimental designs. R package version 2.0.0.
- Butler, D. (2020). odw: Generate optimal experimental designs. r package version 2.1.4.
- Butler, D. (2021). asreml: Fits the linear mixed model. R package version 4.1.0.149.
- Crossa, J., Campos, G. d. l., Pérez, P., Gianola, D., Burgueno, J., Araus, J. L., Makumbi, D., Singh, R. P., Dreisigacker, S., Yan, J., et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186(2):713–724.
- Cruz Neto, A. J. d., Rosa, R. C. C., Oliveira, E. J. d., Sampaio, S. R., Santos, I. S. d., Souza, P. U., Passos, A. R., e Jesus, O. N. d. (2016). Genetic parameters, adaptability and stability to selection of yellow passion fruit hybrids. *Crop Breeding and Applied Biotechnology*, 16(4):321–329.
- Cullis, B. R., Smith, A. B., Cocks, N. A., e Butler, D. G. (2020). The design of early-stage plant breeding trials using genetic relatedness. *Journal of Agricultural, Biological and Environmental Statistics*, 25(4):553–578.
- Cullis, B. R., Smith, A. B., e Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *Journal of agricultural, biological, and environmental statistics*, 11(4):381.
- de Oliveira, A. L. B., Gouvêa, L. R. L., Verardi, C. K., Silva, G. A. P., e de Gonçalves, P. S. (2015). Genetic variability and predicted genetic gains for yield and laticifer system traits of rubber tree families. *Euphytica*, 203(2):285–293.
- Endelman, J. B., Atlin, G. N., Beyene, Y., Semagn, K., Zhang, X., Sorrells, M. E., e Jannink, J.-L. (2014). Optimal design of preliminary yield trials with genome-wide markers. *Crop Science*, 54(1):48–59.
- Gaynor, R. C., Gorjanc, G., Bentley, A. R., Ober, E. S., Howell, P., Jackson, R., Mackay, I. J., e Hickey, J. M. (2017). A two-part strategy for using genomic selection to develop inbred lines. *Crop Science*, 57(5):2372–2386.
- Gaynor, R. C., Gorjanc, G., e Hickey, J. M. (2021). Alphasimr: an r package for breeding program simulations. *G3 Gene/Genomes/Genetics*, 11(jkaa07).
- Gilmour, A. R., Cullis, B. R., e Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics*, pages 269–293.
- Goes, A. d. L. (2020). *Delineamentos ótimos para experimentos com cana-de-açúcar*. PhD thesis, Universidade de São Paulo.
- Hoefler, R., González-Barrios, P., Bhatta, M., Nunes, J. A., Berro, I., Nalin, R. S., Borges, A., Covarrubias, E., Diaz-Garcia, L., Quincke, M., et al. (2020). Do spatial designs outperform classic experimental designs? *Journal of Agricultural, Biological and Environmental Statistics*, 25(4):523–552.

- Hooks, T., Marx, D., Kachman, S., e Pedersen, J. (2009). Optimality criteria for models with random effects. *Revista Colombiana de Estadística*, 32(1):17–31.
- Kempton, R. (1984). The design and analysis of unreplicated field trials. *Vortraege fuer Pflanzenzuechtung (Germany)*.
- Lara, L. A. d. C., Pocrnic, I., Oliveira, T. d. P., Gaynor, R. C., e Gorjanc, G. (2022). Temporal and genomic analysis of additive genetic variance in breeding programmes. *Heredity*, 128(1):21–32.
- Merrick, L. F., Herr, A. W., Sandhu, K. S., Lozada, D. N., e Carter, A. H. (2022). Optimizing plant breeding programs for genomic selection. *Agronomy*, 12(3):714.
- Oakey, H., Verbyla, A., Pitchford, W., Cullis, B., e Kuchel, H. (2006). Joint modeling of additive and non-additive genetic line effects in single field trials. *Theoretical and Applied Genetics*, 113(5):809–819.
- Oakey, H., Verbyla, A. P., Cullis, B. R., Wei, X., e Pitchford, W. S. (2007). Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. *Theoretical and Applied Genetics*, 114(8):1319–1332.
- Paget, M., Alspach, P., Anderson, J., Genet, R., Braam, W., e Apiolaza, L. (2017). Replicate allocation to improve selection efficiency in the early stages of a potato breeding scheme. *Euphytica*, 213(9):1–15.
- Patterson, H. D. e Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3):545–554.
- Pérez-Rodríguez, P., Crossa, J., Rutkoski, J., Poland, J., Singh, R., Albizu, A. L., Autrique, E., De los Campos, G., Burgueño, J., e Dreisigacker, S. (2017). Single-step genomic and pedigree genotype x environment interaction models for predicting wheat lines in international environments. *Plant Genome*, 10(2):np.
- Piepho, H., Möhring, J., Melchinger, A., e Büchse, A. (2008). Blup for phenotypic selection in plant breeding and variety testing. *Euphytica*, 161(1-2):209–228.
- Piepho, H.-P., Vo-Thanh, N., e Tobias, R. (2020). Generating experimental designs for estimation of genetically related treatment effects using sas. *Agronomy Journal*, 112(5):3929–3940.
- Santos, A. d. (2017). *Design and analysis of sugarcane breeding experiments: a case study*. PhD thesis, Universidade de São Paulo.
- Sermarini, R. A., Brien, C., Demétrio, C. G. B., e dos Santos, A. (2020). Impact on genetic gain from using misspecified statistical models in generating p-rep designs for early generation plant-breeding experiments. *Crop Science*, 60(6):3083–3095.
- Shah, K. R. e Sinha, B. (2012). *Theory of optimal designs*, volume 54. Springer Science & Business Media.
- Smith, A., Cullis, B., e Thompson, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*, 57(4):1138–1147.
- Sukumaran, S., Crossa, J., Jarquín, D., e Reynolds, M. (2017). Pedigree-based prediction models with genotype x environment interaction in multi-environment trials of cimmyt wheat.
- Tanaka, E. (2020). Simple outlier detection for a multi-environmental field trial. *Biometrics*, 76(4):1374–1382.

- Verbyla, A. P. (1990). A conditional derivation of residual maximum likelihood. *Australian Journal of Statistics*, 32(2):227–230.
- Verges, V. L. e Van Sanford, D. A. (2020). Genomic selection at preliminary yield trial stage: Training population design to predict untested lines. *Agronomy*, 10(1):60.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*.
- Williams, E., Piepho, H.-P., e Whitaker, D. (2011). Augmented p-rep designs. *Biometrical Journal*, 53(1):19–27.

4 CONSIDERAÇÕES FINAIS

Neste estudo, foram avaliados delineamentos *grid-plot* e *p*-rep espacialmente otimizados para 180 *test lines* em uma área experimental de 24 linhas por 10 colunas e 2 blocos para ensaios multi-ambientais (três ambientes), com o objetivo de investigar o ganho genético e a qualidade de seleção considerando medidas do ganho genético relativo percebido, da precisão genética e do sucesso de seleção ao assumir diferentes caracterizações para esses delineamentos. Isto é, considerando diferentes números de *checks* (0, 1, 2, 4) com repetições para as mesmas (0, 5, 10, 15, 20) e diferentes porcentagens de *test-lines* repetidas (11%, 22% e 33%), sob diferentes suposições a respeito do modelo para gerar o delineamento. Também foram considerados diferentes cenários ao atribuir valores diferentes para os parâmetros de variâncias associados aos efeitos genéticos e residual por local. As análises dos conjuntos de dados simulados foram realizadas de forma individual e conjunta para os ensaios nos três locais.

No Capítulo 2, verificou-se a efetividade dos delineamentos gerados sob a hipótese de efeitos de *test-lines* fixos e aleatórios. De modo geral, não foram identificadas diferenças entre as medidas de comparação ao considerar efeitos de tratamento fixos ou aleatórios no modelo de delineamento. No entanto, houve uma distribuição uniforme das *test lines* replicadas e/ou dos tratamentos na área experimental ao assumir efeitos de *test lines* fixos. Por esta razão, é recomendado assumir efeitos de *test lines* aleatórios para gerar delineamentos ótimos. Adicionalmente, observou-se que, para as análises individuais e conjuntas, os delineamentos *p*-rep foram ligeiramente superiores aos delineamentos *grid-plot* na seleção do material genético para todos os cenários explorados. Entre os delineamentos *p*-rep, o que não contém *checks* resultou no maior ganho genético. Além disso, quanto maior a razão entre as variâncias genéticas e residuais e a correlação entre as mesmas linhagens em locais diferentes, melhores serão os resultados.

Estendendo o estudo, o Capítulo 3 teve como objetivo gerar delineamentos ótimos considerando as suposições de dependência ou independência entre os tratamentos (*test-lines* e *checks*). Quando os tratamentos foram assumidos não independentes, a matriz de parentesco utilizada foi proveniente de um estudo de simulação de um programa completo de melhoramento de trigo por 40 anos de seleção fenotípica convencional para rendimento considerando o *pedrigree* das linhagens no estágio PYT do 39º ano e conhecendo seis gerações anteriores. Além disso, foram considerados dois cenários para as variâncias genética e residual. Em ambas as situações (análises individuais e conjuntas, com e sem a matriz de relacionamento no modelo do delineamento), na grande maioria das vezes, os delineamentos *p*-rep mostraram superioridade em relação aos delineamentos *grid-plot*. Mas não foram observadas diferenças em relação às medidas de comparação (RRGG, precisão genética e sucesso de seleção) ao considerar as suposições sobre a relação entre os tratamentos no modelo para o delineamento. De modo geral, notou-se que quando são fixados valores iguais para os parâmetros de variâncias associados aos efeitos não genéticos e residuais, quanto maior a variância genética melhores são o ganho genético e a qualidade de seleção, como esperado. Observou-se também que quanto maior a correlação entre os mesmos genótipos aplicados em diferentes locais, melhores são os resultados. Por fim, quando considerou-se valores mais altos para os parâmetros de variâncias genéticas e residual por local, foi possível notar uma melhor aleatorização das variedades em teste na área experimental e melhores resultados para as medidas utilizadas.

De modo geral, os resultados apresentados em ambos os capítulos indicaram que o delineamento parcialmente replicado sem *checks* apresentou os melhores resultados e pode ser considerado para ensaios nas fases iniciais de programas de melhoramento genético de planta. Porém, caso seja necessário o uso de *checks* (delineamento 6 ou *p*-rep 5), recomenda-se o delineamento *p*-rep com pequeno número de repetições para as *checks*, ao invés dos delineamentos *grid-plot*. Este estudo apresenta algumas limitações. Aqui, foram considerados apenas dados simulados adotando conjuntos de parâmetros cujos valores foram estimados em estudos desenvolvidos por programas de melhoramento de plantas e, em ambos os capítulos, considerou-se o mesmo tamanho de área experimental e a mesma suposição em relação ao

número de *test lines* e *checks* em avaliação. Desta forma, outros resultados poderiam ser encontrados, se fossem assumidos, por exemplo, valores mais altos para os parâmetros de variâncias não genéticos. Além disso, a matriz de relacionamento utilizada no Capítulo 3 foi para apenas um tipo de cultivar (o trigo). Pesquisas futuras podem ser desenvolvidas considerando ensaios de uniformidade e outras matrizes de relacionamento de diferentes cultivares para analisar situações semelhantes às realizadas neste estudo.

APÊNDICES

No material suplementar são apresentadas as tabelas de média das variâncias das diferenças entre pares de tratamentos para os delineamentos ótimos que foram gerados considerando as hipóteses de efeitos aleatórios e efeitos fixos de *test lines* no modelo para gerar o delineamento, respectivamente. Também são apresentados os *layouts* para os Delineamentos 1, 2, 3, 5 e 6, que são delineamentos otimizados com ϕ_g de 0,9 para tratamentos de efeitos aleatórios e fixos. Além disso, também são apresentadas estimativas dos componentes de variância para todos os delineamentos estudados considerando as análises conjuntas e individuais, e ambas as hipóteses quanto ao efeito das *test lines*. No Apêndice II, são apresentados os gráficos de setores indicando a porcentagem de superioridade dos delineamentos espacialmente otimizados gerados considerando a suposição de dependência e independência das linhagens. Além disso, também são apresentadas as porcentagens de superioridade dos delineamentos quando não é considerado no modelo de análise a matriz de relacionamento. Este material suplementar contém duas tabelas e quatorze figuras.

Apêndice I

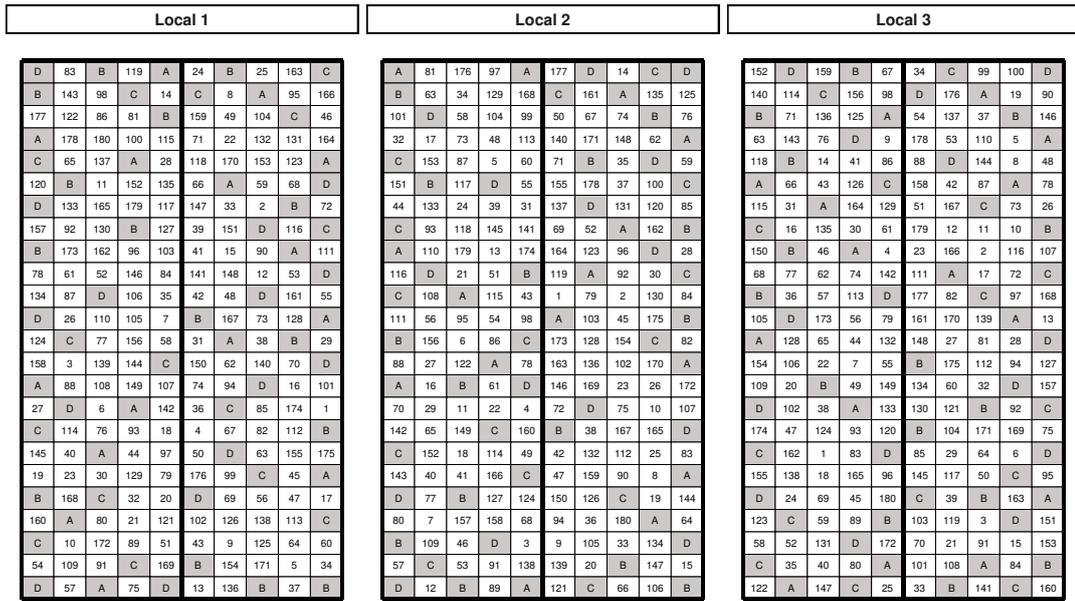
Tabela A.1. Média das variâncias das diferenças entre os pares de tratamento para os delineamentos ótimos que foram gerados considerando as hipóteses de que os *test lines* são de efeitos aleatórios.

	Local 1		Local 2		Local 3	
	Duplicadas	Não Duplicadas	Duplicadas	Não Duplicadas	Duplicadas	Não Duplicadas
$\phi_g = 0,7$						
Delineamento 1	-	0,546	-	0,682	-	0,683
Delineamento 2	0,369	0,549	0,464	0,685	0,471	0,685
Delineamento 3	0,340	0,557	0,470	0,683	0,471	0,687
Delineamento 4	0,349	0,557	0,457	0,690	0,461	0,690
Delineamento 5	0,366	0,547	0,458	0,680	0,465	0,678
Delineamento 6	0,366	0,549	0,458	0,684	0,461	0,686
$\phi_g = 0,8$						
Delineamento 1	-	0,477	-	0,596	-	0,597
Delineamento 2	0,320	0,481	0,401	0,601	0,408	0,601
Delineamento 3	0,318	0,484	0,397	0,604	0,405	0,603
Delineamento 4	0,317	0,481	0,398	0,601	0,403	0,600
Delineamento 5	0,318	0,479	0,398	0,596	0,405	0,595
Delineamento 6	0,319	0,481	0,400	0,601	0,403	0,600
$\phi_g = 0,9$						
Delineamento 1	-	0,477	-	0,596	-	0,597
Delineamento 2	0,250	0,392	0,315	0,489	0,324	0,489
Delineamento 3	0,254	0,393	0,318	0,490	0,319	0,490
Delineamento 4	0,251	0,392	0,315	0,489	0,318	0,489
Delineamento 5	0,252	0,390	0,316	0,485	0,321	0,484
Delineamento 6	0,253	0,393	0,317	0,491	0,319	0,489

Tabela A.2. Média das variâncias das diferenças entre pares de tratamentos para os delineamentos ótimos que foram gerados considerando as suposições de que os efeitos de tratamentos são fixos.

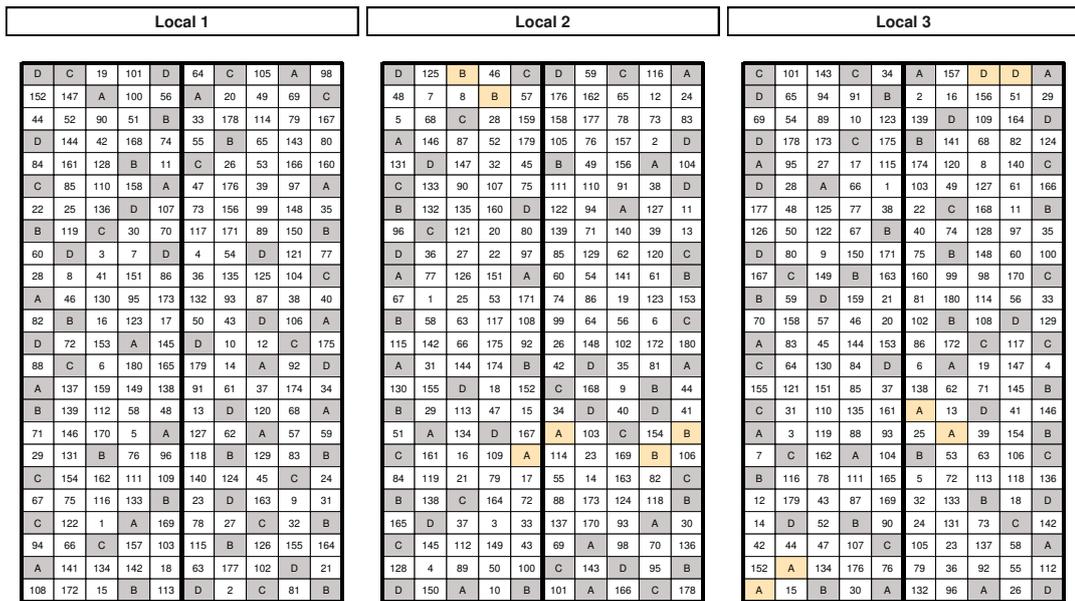
	Local 1		Local 2		Local 3	
	Duplicadas	Não Duplicadas	Duplicadas	Não Duplicadas	Duplicadas	Não Duplicadas
$\phi_g = 0,7$						
Delineamento 1	-	0,545	-	0,683	-	0,685
Delineamento 2	0,356	0,552	0,453	0,691	0,444	0,693
Delineamento 3	0,360	0,559	0,446	0,700	0,447	0,702
Delineamento 4	0,354	0,558	0,440	0,698	0,445	0,700
Delineamento 5	0,357	0,555	0,443	0,692	0,444	0,694
Delineamento 6	0,348	0,568	0,440	0,705	0,445	0,704
$\phi_g = 0,8$						
Delineamento 1	-	0,476	-	0,596	-	0,596
Delineamento 2	0,316	0,482	0,396	0,602	0,393	0,605
Delineamento 3	0,313	0,489	0,392	0,612	0,393	0,612
Delineamento 4	0,311	0,488	0,390	0,609	0,390	0,610
Delineamento 5	0,309	0,486	0,388	0,605	0,388	0,606
Delineamento 6	0,307	0,497	0,386	0,619	0,390	0,617
$\phi_g = 0,9$						
Delineamento 1	-	0,385	-	0,482	-	0,482
Delineamento 2	0,250	0,391	0,310	0,489	0,311	0,489
Delineamento 3	0,252	0,396	0,313	0,495	0,312	0,497
Delineamento 4	0,249	0,394	0,310	0,493	0,310	0,494
Delineamento 5	0,249	0,393	0,312	0,490	0,312	0,490
Delineamento 6	0,249	0,400	0,308	0,503	0,309	0,502

(a)



Checks não vizinhas Test lines não repetidas

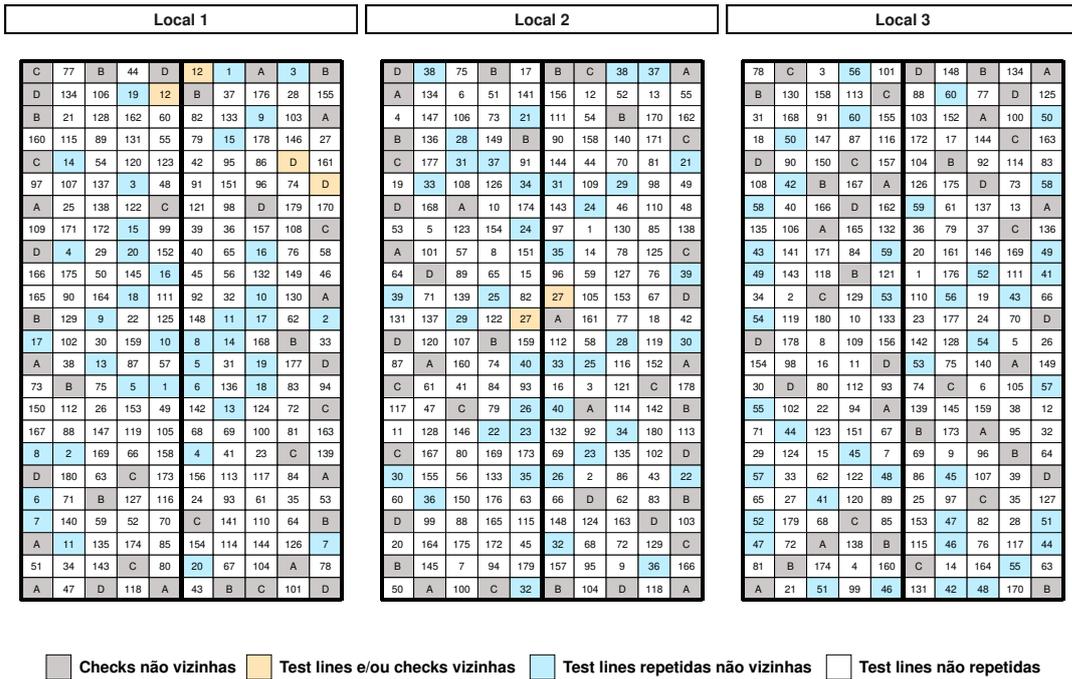
(b)



Checks não vizinhas Checks vizinhas Test lines não repetidas

Figura A.1. Layouts para o Delineamento 1 descritos na Tabela 3.1 onde consideramos $p=0\%$, 180 test lines não duplicadas, 4 checks repetidas 15 vezes para $\phi_g = 0,9$, ao assumir (a) efeitos aleatórios de test lines e (b) efeitos fixos de test lines no modelo para gerar o delineamento.

(a)



(b)

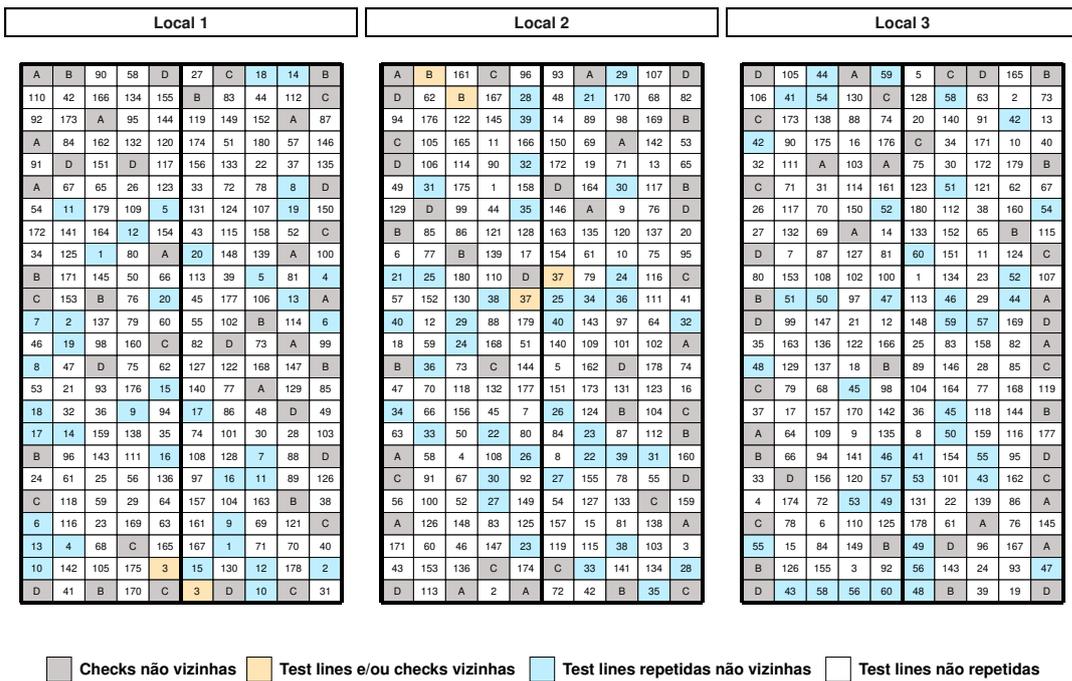
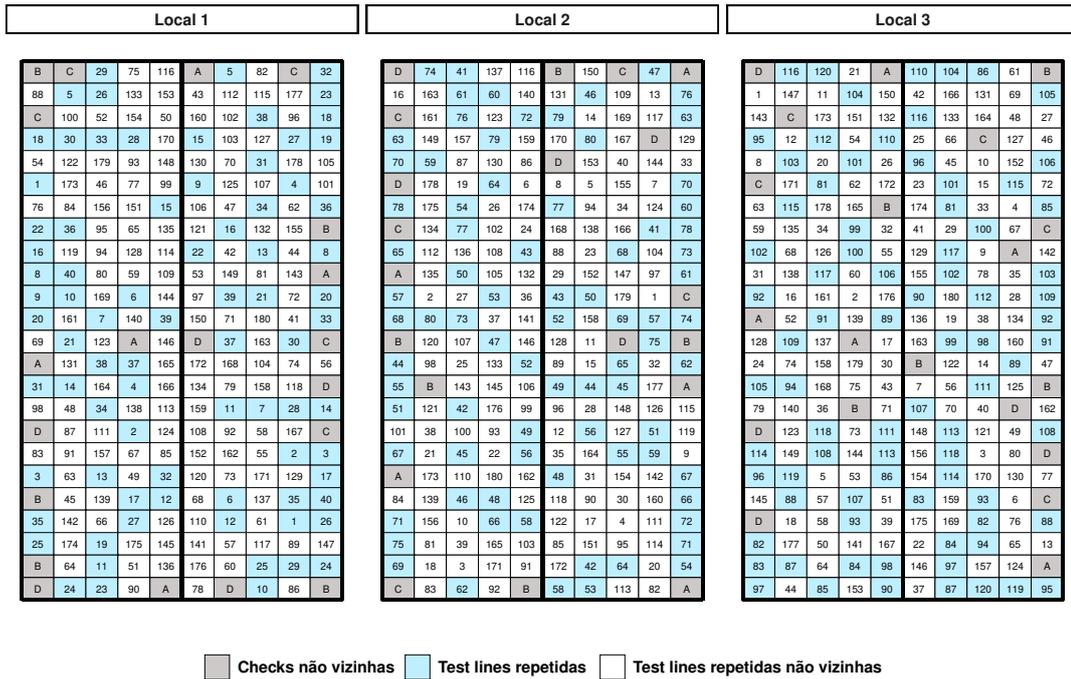


Figura A.2. Layouts para o Delineamento 2 descritos na Tabela 3.1 em que consideramos $p=11\%$, 160 test lines não duplicadas, 20 test lines duplicadas, 4 checks repetidas 10 vezes para $\phi_g = 0,9$, ao assumir (a) efeitos aleatórios de test lines e (b) efeitos fixos de test lines no modelo para gerar o delineamento.

(a)



(b)

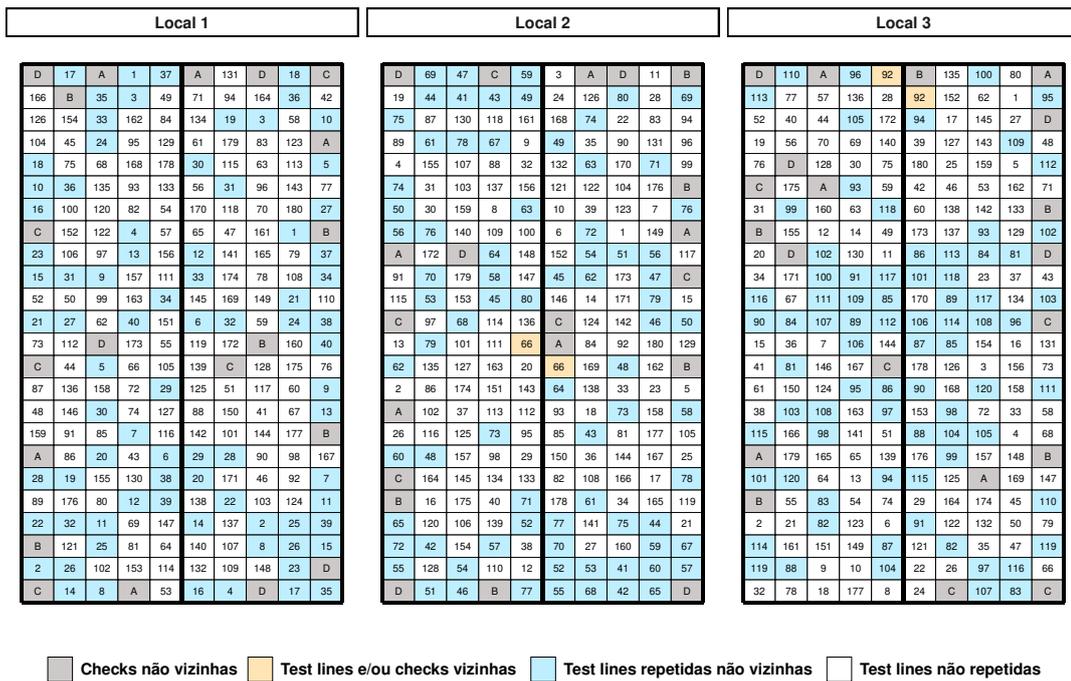
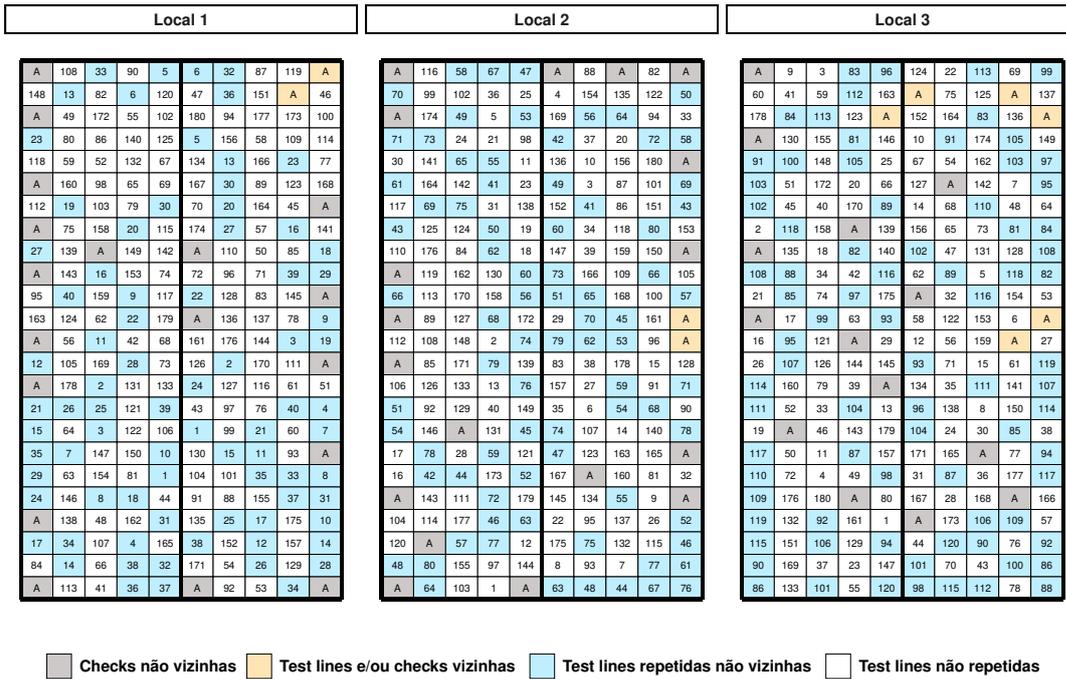


Figura A.3. Layouts para o Delineamento 3 descritos na Tabela 3.1 consideramos $p=22\%$, 140 test lines não duplicadas, 40 linhas de teste duplicadas, 4 verificações repetidas 5 vezes para $\phi_g = 0,9$, ao assumir (a) efeitos aleatórios de test lines e (b) efeitos fixos de test lines no modelo para gerar o delineamento.

(a)



(b)

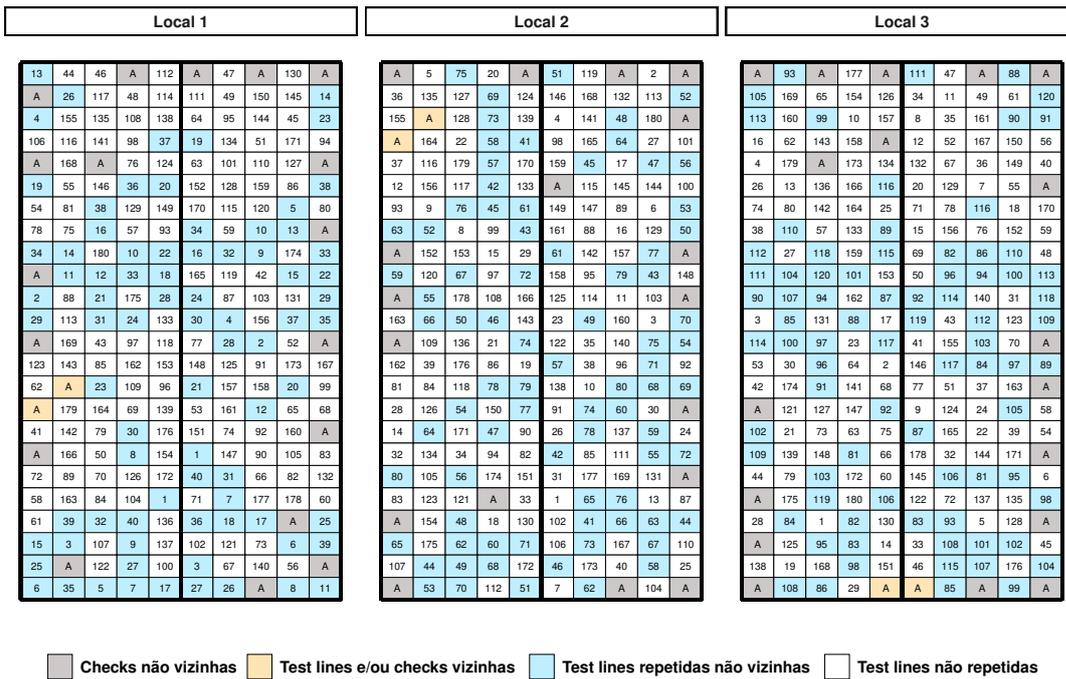
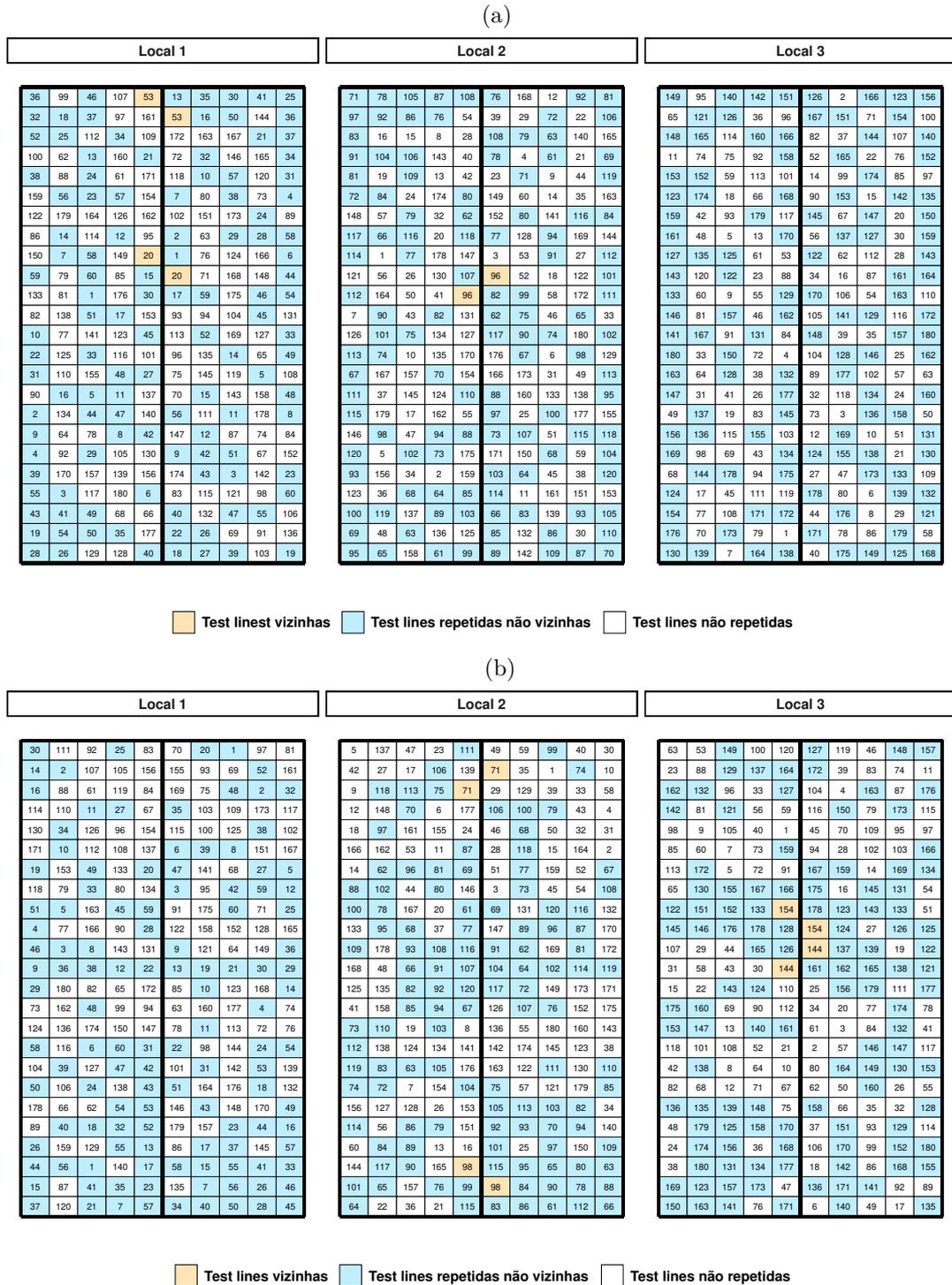


Figura A.4. Layouts para o Delineamento 5 descritos na Tabela 3.1 consideramos $p= 22\%$, 140 test lines não duplicadas, 40 test lines duplicadas, 1 checks repetidas 20 vezes para $\phi_g = 0,9$, ao assumir (a) efeitos aleatórios de test lines e (b) efeitos fixos de test lines no modelo para gerar o delineamento.



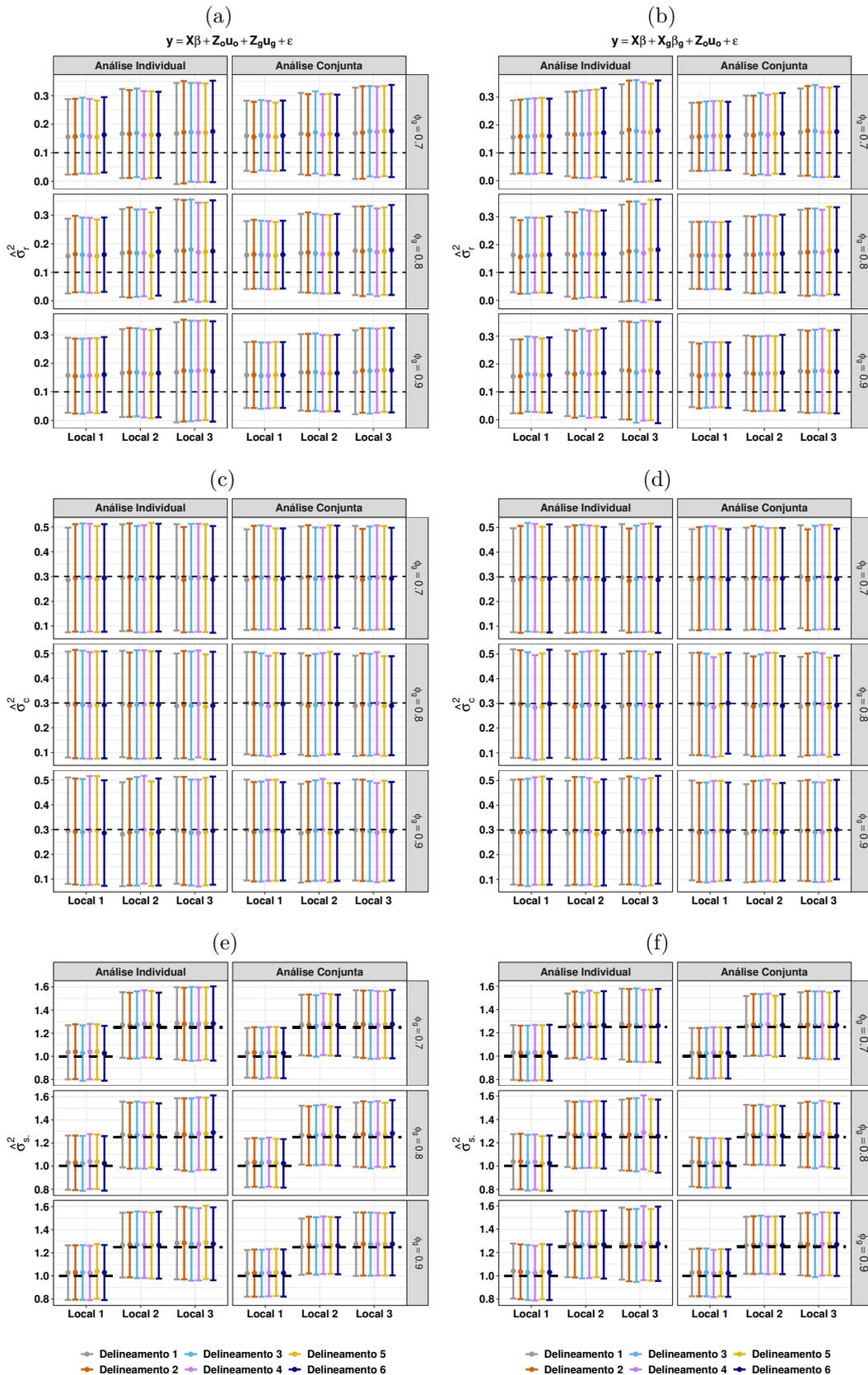


Figura A.6. Média \pm erro padrão das estimativas dos parâmetros relacionados às variâncias entre ((a) e (b)) linhas, ((c) e (d)) colunas e ((d) e (e)) resíduos do local, ao assumir (((a), (c) e (e))) efeitos aleatórios de *test lines* e ((b), (d) e (f)) efeitos fixos de *test lines* no modelo para gerar o delineamento, para as análises individuais e conjuntas dos diferentes delineamentos ótimos obtidos. A linha tracejada preta indica o valor esperado (suposto) para cada parâmetro.

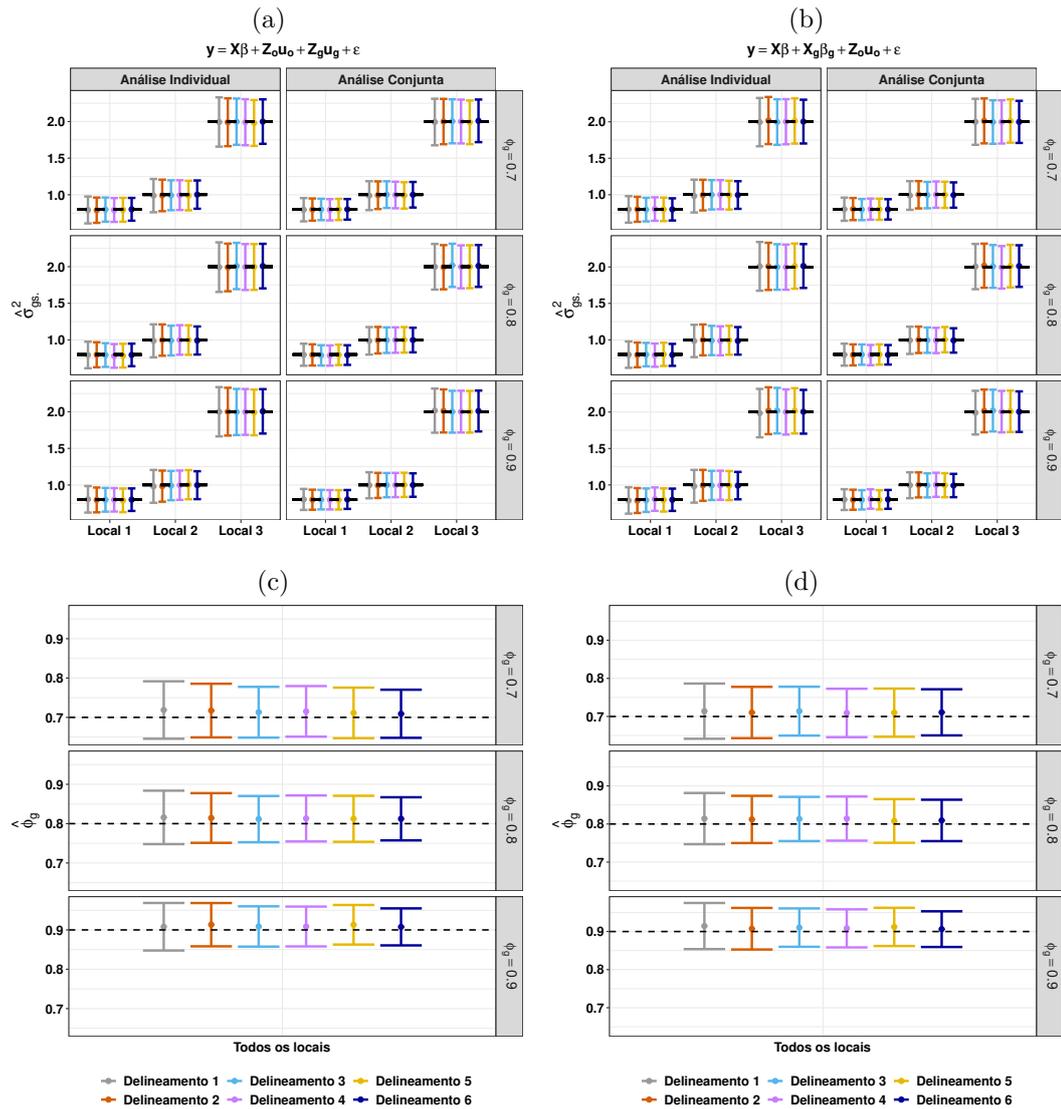


Figura A.7. Média \pm erro padrão das estimativas dos parâmetros relacionados às variâncias genéticas ((a) e (b)) e a correlação genética ((entre os efeitos da *test lines* em local diferente) ((c) e (d)) das diferentes áreas, ao assumir ((a) e (c)) efeitos aleatórios de *test lines* e ((b) e (d)) efeitos fixos de *test lines* no modelo para gerar o delineamento, para as análises individuais e conjuntas dos diferentes delineamentos ótimos obtidos. A linha tracejada preta indica o valor esperado (suposto) para cada parâmetro.

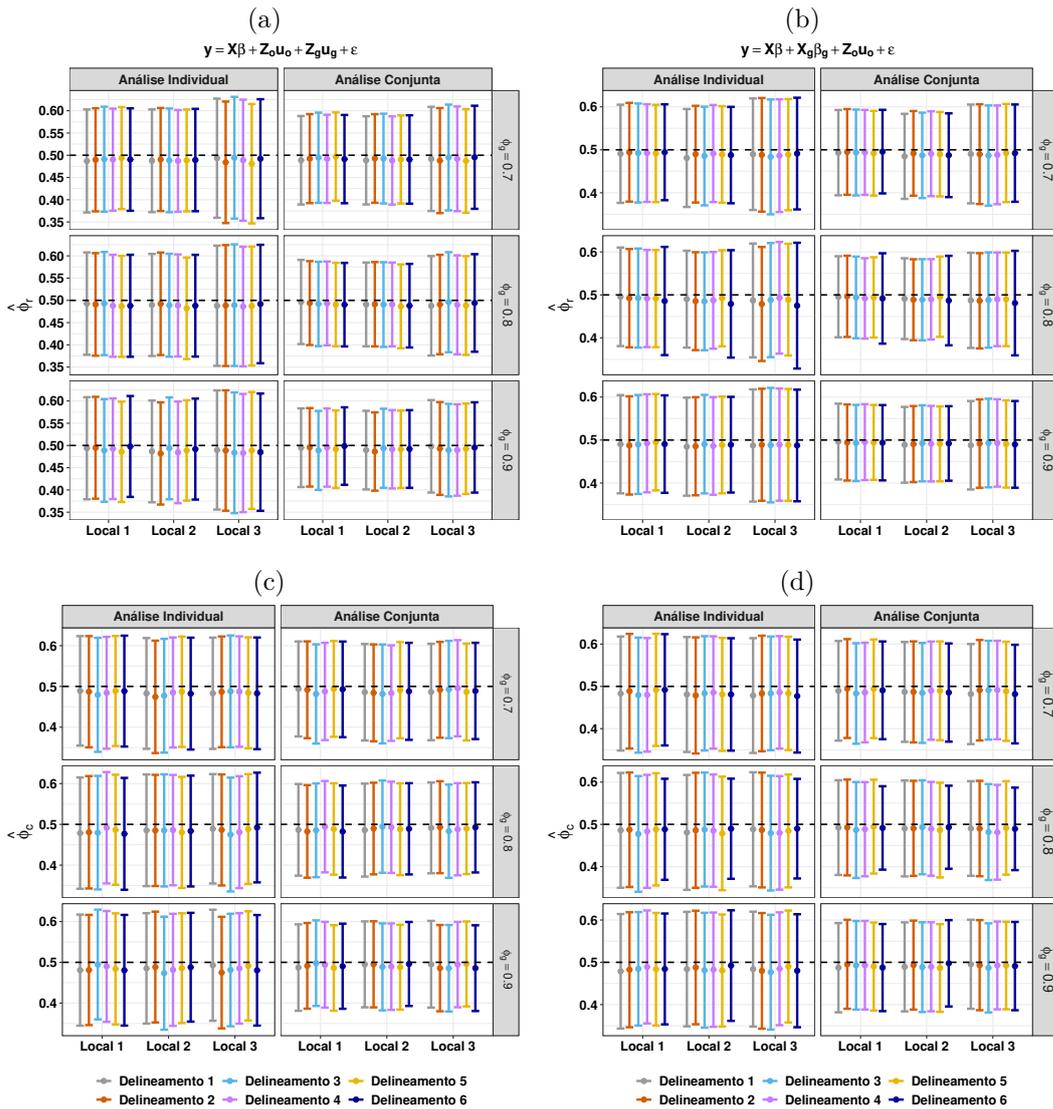


Figura A.8. Média \pm erro padrão das estimativas dos parâmetros relacionados a ((a) e (b)) correlação residual entre as linhas e ((c) e (d)) correlação residual entre colunas, ao assumir (((a) e (c))) efeitos aleatórios de *test lines* e ((b) e (d)) efeitos fixos de *test lines* no modelo para gerar o delineamento, para as análises individuais e conjuntas dos diferentes delineamentos ótimos obtidos. A linha tracejada preta indica o valor esperado (suposto) para cada parâmetro.

Apêndice II

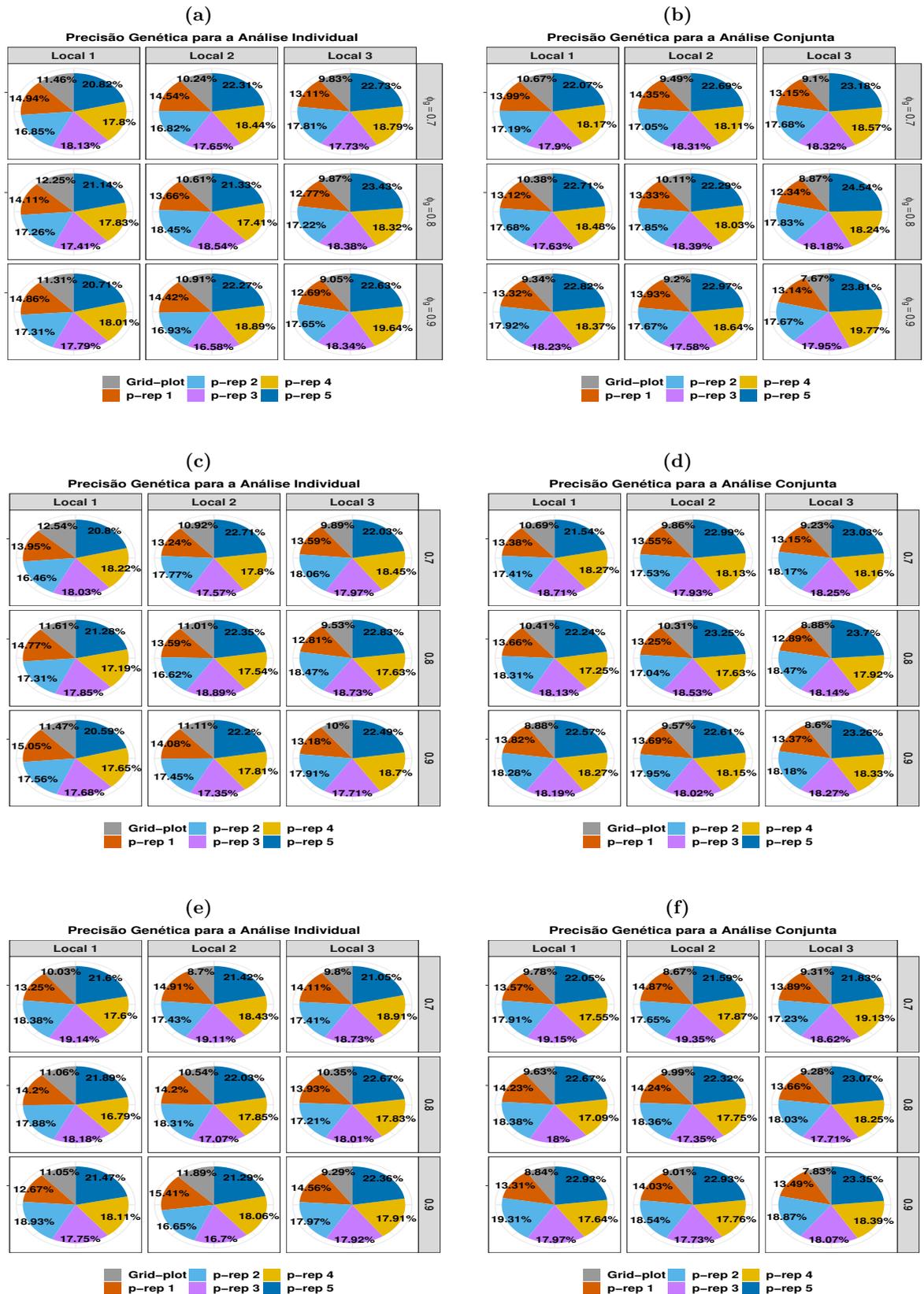


Figura A.9. Valores das porcentagens das comparações dos delineamentos para a precisão genética referente ao primeiro cenário para a análise individual ((a), (c) e (e)) e conjunta ((b), (d) e (f)). Os painéis (a) e (b) referem-se aos resultados obtidos dos delineamentos gerados considerando a matriz de relacionamento; (c) e (d) referem-se aos resultados obtidos dos delineamentos gerados sem a inclusão da matriz de relacionamento e (e) e (f) aos resultados obtidos dos modelos de análise que não incluiu a matriz de relacionamento.

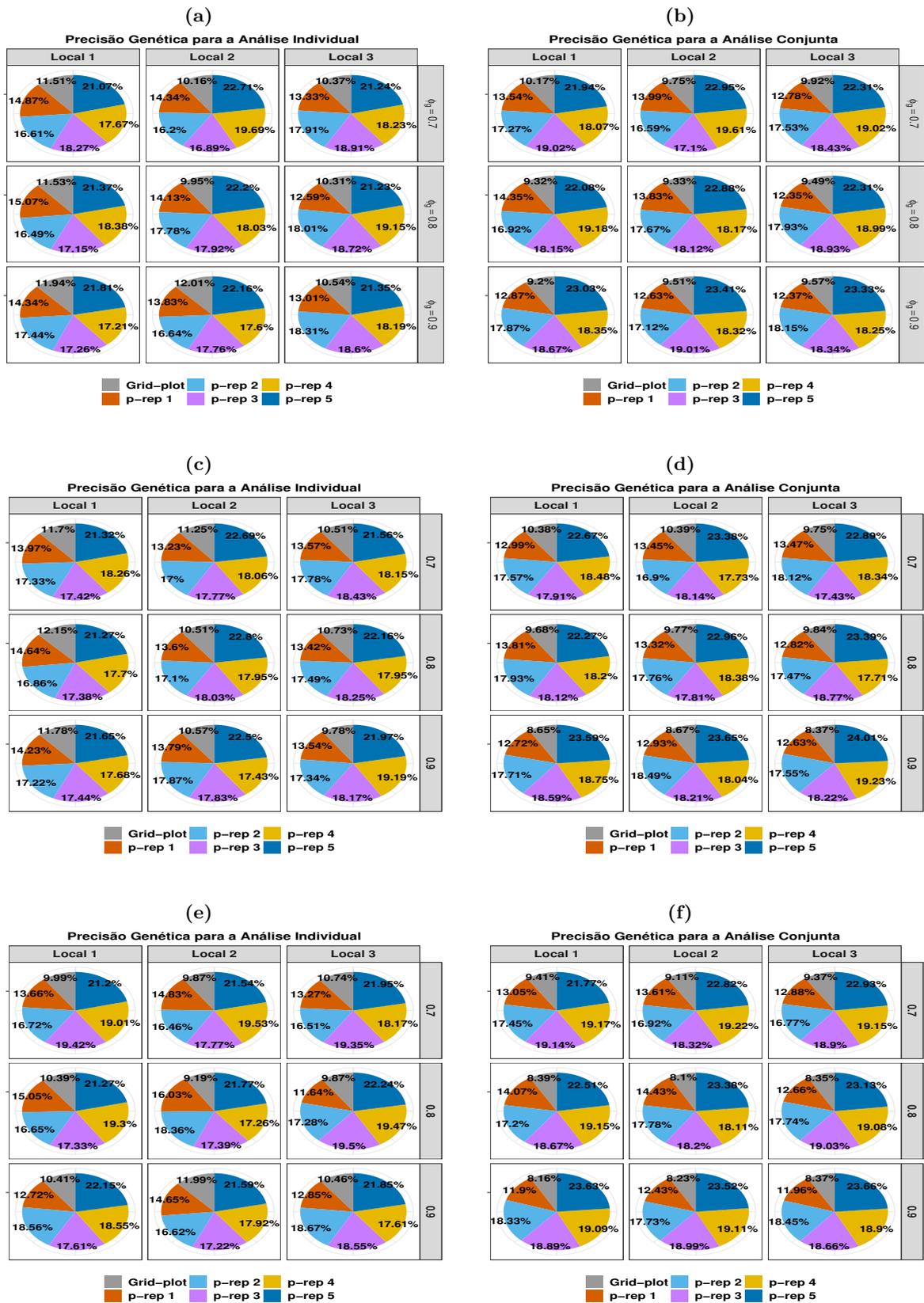


Figura A.10. Valores das porcentagens das comparações dos delineamentos para a precisão genética referente ao segundo cenário para a análise individual ((a), (c) e (e)) e conjunta ((b), (d) e (f)). Os painéis (a) e (b) referem-se aos resultados obtidos dos delineamentos gerados considerando a matriz de relacionamento; (c) e (d) referem-se aos resultados obtidos dos delineamentos gerados sem a inclusão da matriz de relacionamento e (e) e (f) aos resultados obtidos dos modelos de análise que não incluiu a matriz de relacionamento.

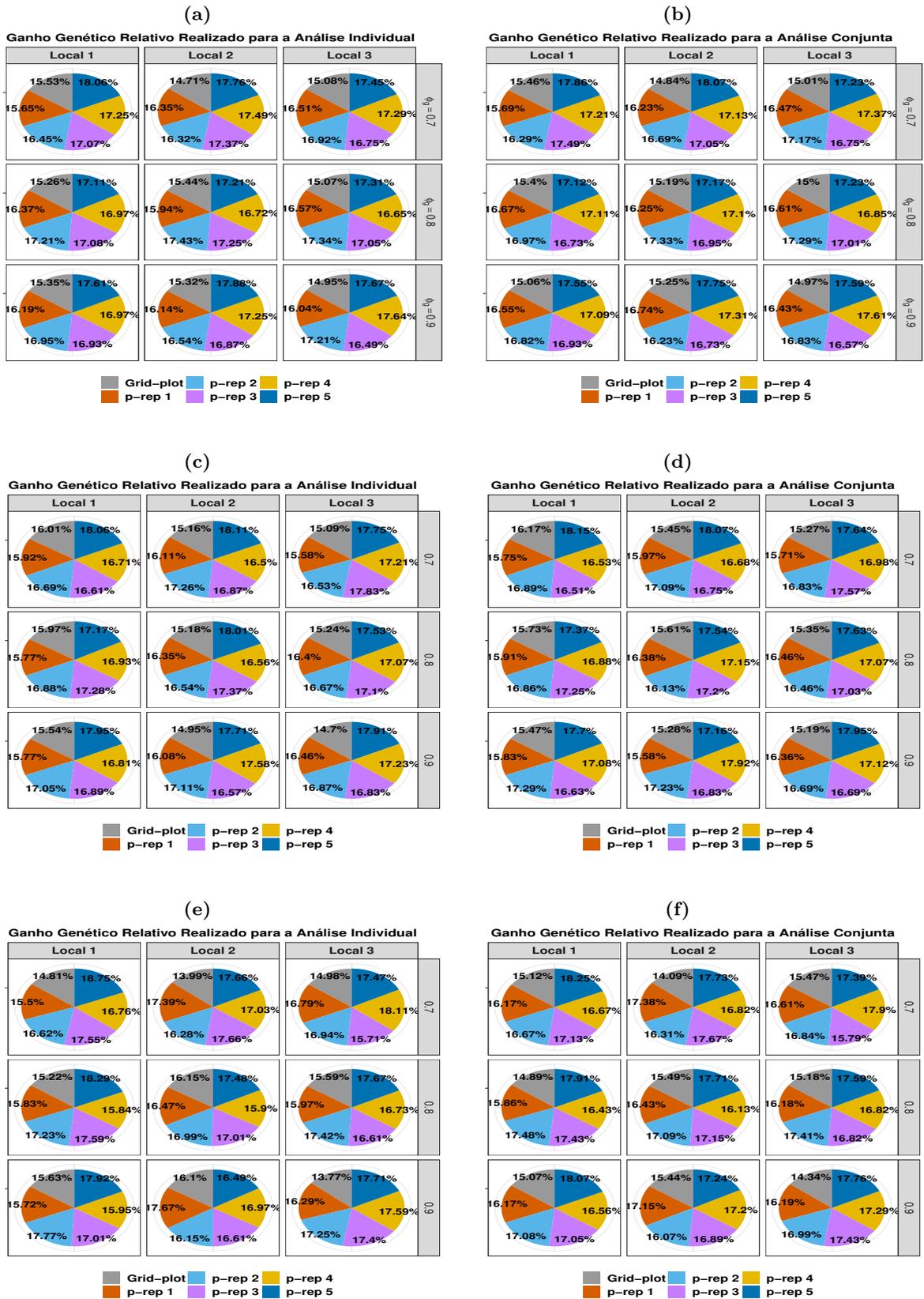


Figura A.11. Valores das porcentagens das comparações dos delineamentos para a Ganho Genético Relativo Realizado referente ao primeiro cenário para a análise individual ((a), (c) e (e)) e conjunta ((b), (d) e (f)). Os painéis (a) e (b) referem-se aos resultados obtidos dos delineamentos gerados considerando a matriz de relacionamento; (c) e (d) referem-se aos resultados obtidos dos delineamentos gerados sem a inclusão da matriz de relacionamento e (e) e (f) aos resultados obtidos dos modelos de análise que não incluiu a matriz de relacionamento.

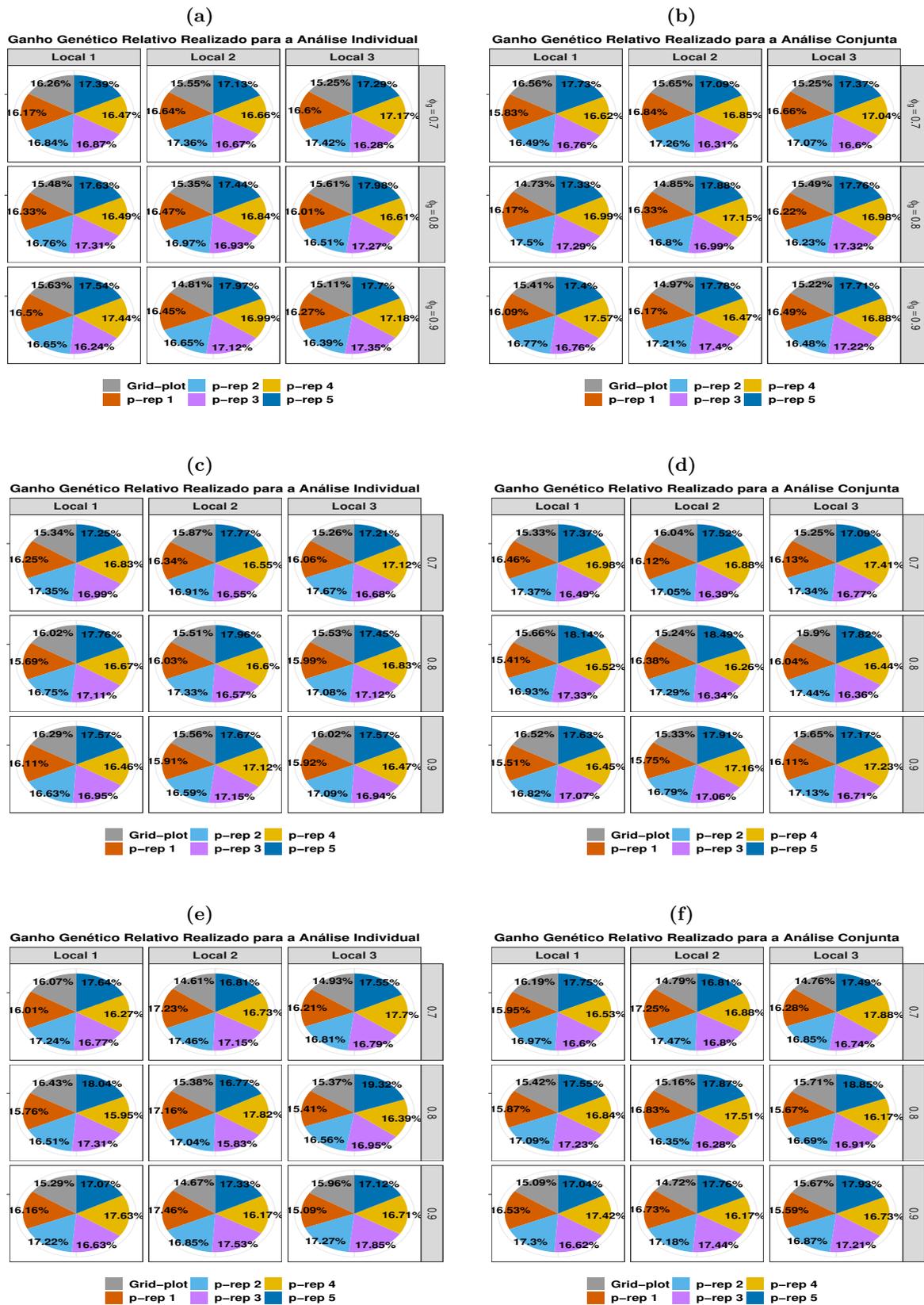


Figura A.12. Valores das porcentagens das comparações dos delineamentos para a Ganho Genético Relativo Realizado referente ao segundo cenário para a análise individual ((a), (c) e (e)) e conjunta ((b), (d) e (f)). Os painéis (a) e (b) referem-se aos resultados obtidos dos delineamentos gerados considerando a matriz de relacionamento; (c) e (d) referem-se aos resultados obtidos dos delineamentos gerados sem a inclusão da matriz de relacionamento e (e) e (f) aos resultados obtidos dos modelos de análise que não incluiu a matriz de relacionamento.

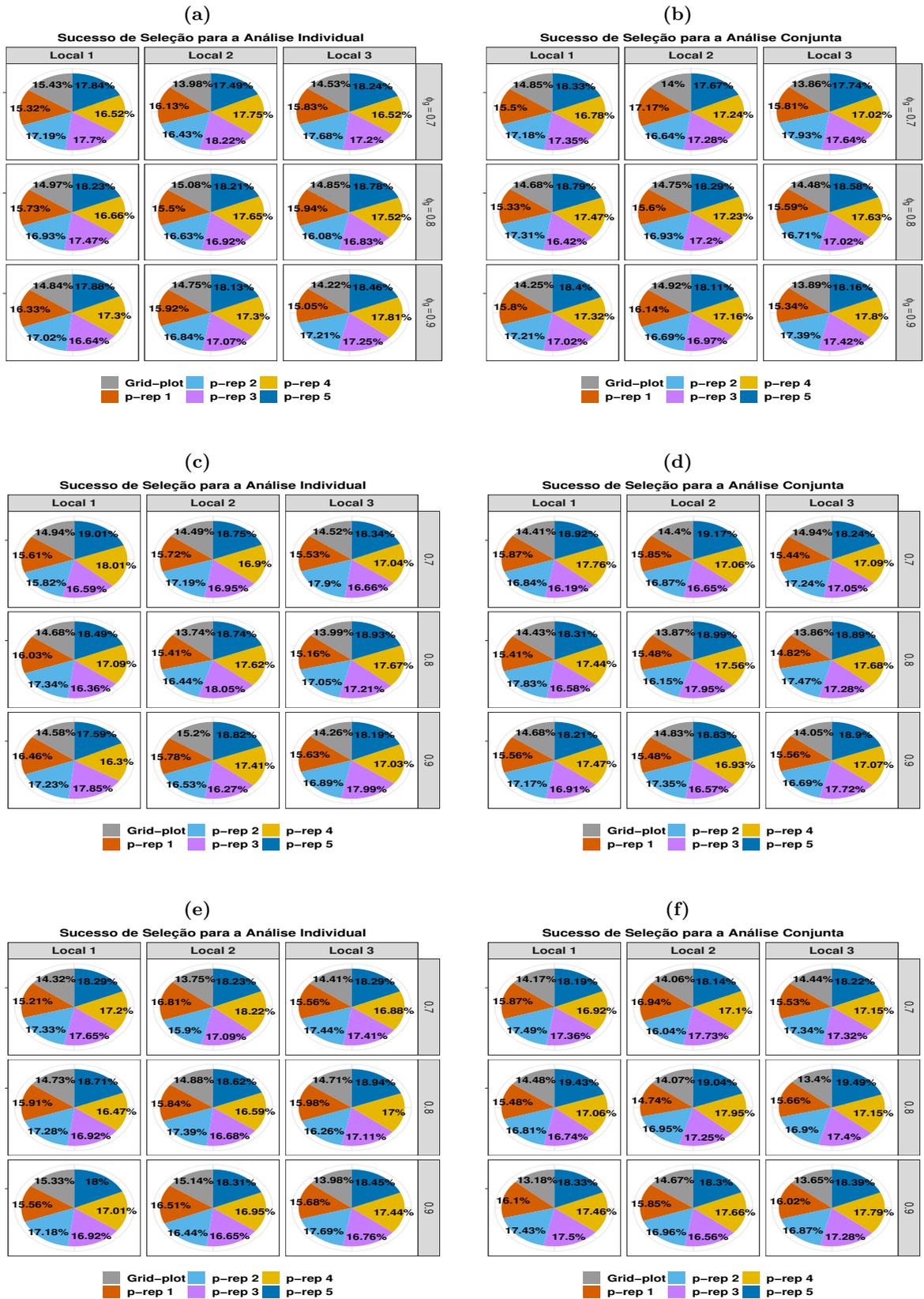


Figura A.13. Valores das porcentagens das comparações dos delineamentos para a Sucesso de Seleção referente ao primeiro cenário para a análise individual ((a), (c) e (e)) e conjunta ((b), (d) e (f)). Os painéis (a) e (b) referem-se aos resultados obtidos dos delineamentos gerados considerando a matriz de relacionamento; (c) e (d) referem-se aos resultados obtidos dos delineamentos gerados sem a inclusão da matriz de relacionamento e (e) e (f) aos resultados obtidos dos modelos de análise que não incluiu a matriz de relacionamento.

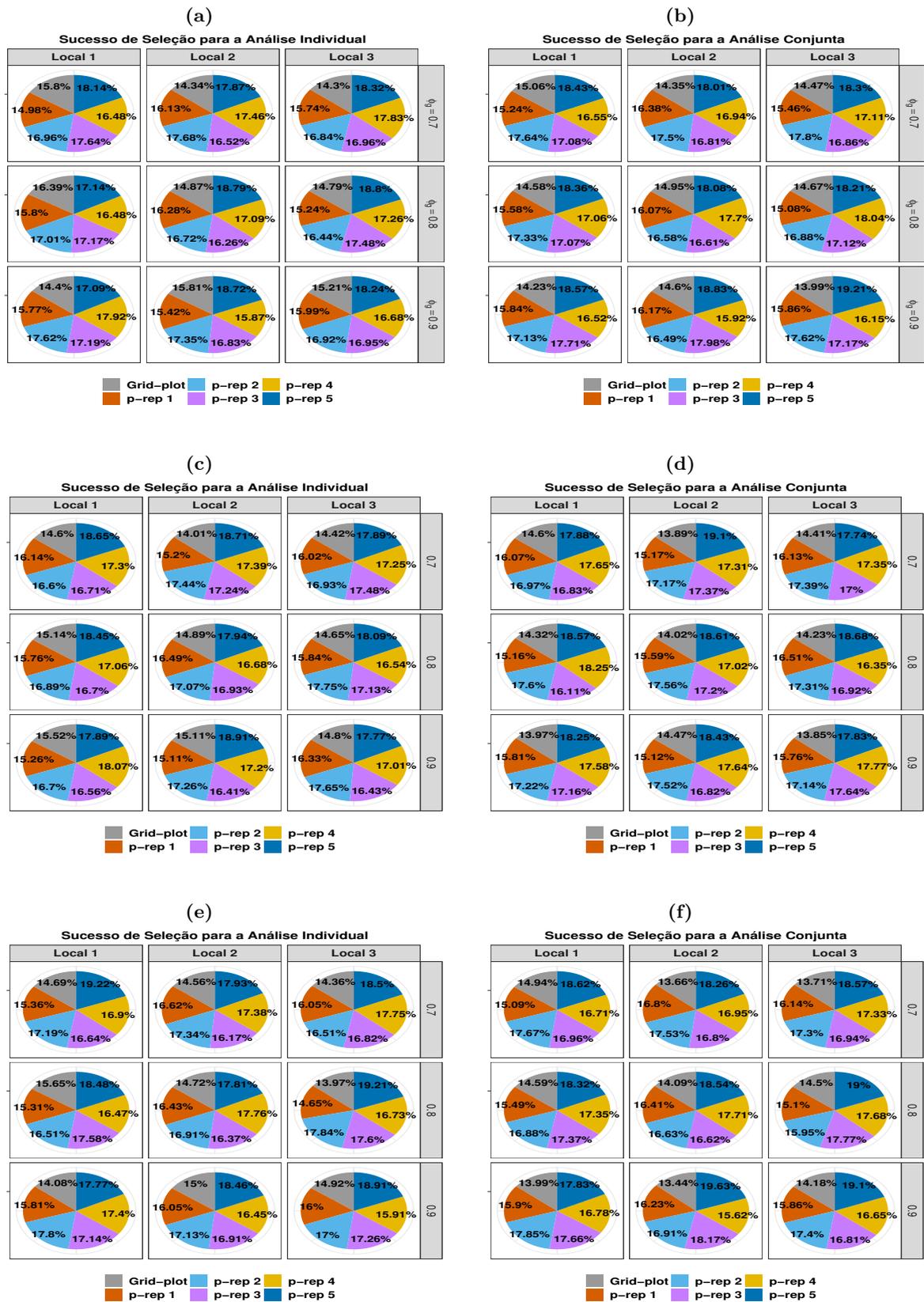


Figura A.14. Valores das porcentagens das comparações dos delineamentos para a Sucesso de Seleção referente ao segundo cenário para a análise individual ((a), (c) e (e)) e conjunta ((b), (d) e (f)). Os painéis (a) e (b) referem-se aos resultados obtidos dos delineamentos gerados considerando a matriz de relacionamento; (c) e (d) referem-se aos resultados obtidos dos delineamentos gerados sem a inclusão da matriz de relacionamento e (e) e (f) aos resultados obtidos dos modelos de análise que não incluiu a matriz de relacionamento.