Optimizing the Allocation of Trials to Sub-Regions in Crop Variety Testing

Maryna Prus

cont. as joint work with L. Filová, H.-P. Piepho and W. Malik

Deutsche Forschungsgemeinschaft

September 11, 2025

UNIVERSITY OF HOHENHEIM

XII Working Seminar on Statistical Methods in Variety Testing

Contents

The Experiment

- P sub-regions
- J locations:
  - J_i locations in sub-region i
  - $J = \sum_{i=1}^{P} J_i$
- L blocks in each location
- K genotypes in each block
- H years, typically $H = 3$

To optimize: $J_1, \dots, J_P$

Overview

- One year experiment, genotype effects uncorrelated

  $\rightarrow$ P. & Piepho (2021)

- **Multi-annual** experiment, genotype effects uncorrelated

  $\rightarrow$ P. & Piepho (2024)

- One year experiment, genotype effects **correlated**

  $\rightarrow$ P. (2025)

- One year experiment, genotype effects **correlated**, **large** number of genotypes

  $\rightarrow$ Bodnar & P. (2026)

Here: ***Multi-annual*** *experiment, genotype effects* ***correlated****,* ***large*** *number of genotypes*

Linear Mixed Model

$$Y_{ijhkl} = \mu_i + \eta_h + \lambda_{ij} + \beta_{ih} + \alpha_{ik} + \omega_{hk} + \delta_{ijh} + \gamma_{ijk} + \tau_{ihk} + \phi_{ijhk} + b_{ijhl} + \varepsilon_{ijhkl}$$

- $i = 1, \dots P$, P - number of sub-regions
- $j = 1, \dots, J_i$, J_i - number of locations in i-th sub-region
  - $J = \sum_{i=1}^{P} J_i$ - total number of locations
- $l = 1, \dots, L$, L - number of blocks in each location
- $k = 1, \dots, K$, K - number of genotypes in each block
- $h = 1, \dots, H$, H - number of years

Linear Mixed Model

$$Y_{ijhkl} = \mu_i + \eta_h + \lambda_{ij} + \beta_{ih} + \boldsymbol{\alpha}_{ik} + \omega_{hk} + \delta_{ijh} + \gamma_{ijk} + \tau_{ihk} + \phi_{ijhk} + b_{ijhl} + \varepsilon_{ijhkl}$$

- μ_i - mean (fixed) effect of i-th sub-region
- η_h - effect of h-th year, $\mathrm{var}(\eta_h) = \sigma_\eta^2$
- λ_{ij} - effect of j-th location within i-th sub-region, $\mathrm{var}(\lambda_{ij}) = \sigma_\lambda^2$
- β_{ih} - effect of i-th sub-region in h-th year, $\mathrm{var}(\beta_{ih}) = \sigma_\beta^2$
- $\boldsymbol{\alpha}_{ik}$ - effect of genotype k in sub-region i,
  $\mathrm{Cov}(\boldsymbol{\alpha}) = \mathbf{U}$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \dots, \boldsymbol{\alpha}_K^\top)^\top$, $\boldsymbol{\alpha}_k = (\alpha_{1k}, \dots, \alpha_{Pk})^\top$
  Often $\mathbf{U} = \mathbf{N} \otimes \mathbf{V}$

Linear Mixed Model

$$Y_{ijhkl} = \mu_i + \eta_h + \lambda_{ij} + \beta_{ih} + \alpha_{ik} + \omega_{hk} + \delta_{ijh} + \gamma_{ijk} + \tau_{ihk} + \phi_{ijhk} + b_{ijhl} + \varepsilon_{ijhkl}$$

- ω_{kh} - effect of genotype k in h-th year, $\mathrm{var}(\omega_{kh}) = \sigma^2_\omega$
- δ_{ijh} - effect of location j within sub-region i in h-th year, $\mathrm{var}(\delta_{ijh}) = \sigma^2_\delta$
- γ_{ijk} - effect of genotype k in location j within sub-region i, $\mathrm{var}(\gamma_{ijk}) = \sigma^2_\gamma$
- τ_{ikh} - effect of genotype k within sub-region i in h-th year, $\mathrm{var}(\tau_{ikh}) = \sigma^2_\tau$
- ϕ_{ijhk} - effect of genotype k in location j in h-th year, $\mathrm{var}(\phi_{ijhk}) = \sigma^2_\phi$

Linear Mixed Model

$$Y_{ijhkl} = \mu_i + \eta_h + \lambda_{ij} + \beta_{ih} + \alpha_{ik} + \omega_{hk} + \delta_{ijh} + \gamma_{ijk} + \tau_{ihk} + \phi_{ijhk} + b_{ijhl} + \varepsilon_{ijhkl}$$

- b_{ijhl} - effect of block l in location j in h-th year, $\mathrm{var}(b_{ijhl}) = \sigma_b^2$
- ε_{ijhkl} - observational error, $\mathrm{var}(\varepsilon_{ijhkl}) = \sigma^2$
- All random effects and observational errors are uncorrelated and have zero mean

LMM here: *Cross-classification* - same locations each year

Often used: *Nested model* - locations nested within year

$\rightarrow$ particular case, without λ_{ij} and γ_{ijk}

Optimization Problem

Search for optimal numbers of locations $J_1, \dots, J_P$ for

- prediction of genotype effects

$$\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_K$$

- prediction of pairwise linear contrasts

$$\boldsymbol{\theta}^{k,k'} = \boldsymbol{\alpha}_k - \boldsymbol{\alpha}_{k'}, \quad k \neq k'$$

for all pairs of genotypes

for given total number of locations J

Mean Squared Error (MSE) Matrix

MSE matrix for BLUP $\hat{\boldsymbol{\alpha}}$ of genotype effects:

$$\mathrm{Cov}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = \left\{ \frac{1}{c}(\mathbf{I}_K - \frac{1}{K}\mathbf{1}_K\mathbf{1}_K^\top) \otimes [(\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{R}]^{-1} + \mathbf{U}^{-1} \right\}^{-1}$$

$$\mathbf{F} = \text{block-diag}(\mathbf{1}_{J_1}, \ldots, \mathbf{1}_{J_P})$$

$$c = \sigma_\gamma^2 + \frac{1}{H}\left(\sigma_\phi^2 + \frac{1}{L}\sigma^2\right)$$

$$\mathbf{R} = \frac{\sigma_\tau^2}{cH}\mathbf{I}_P + \frac{\sigma_\omega^2}{cH}\mathbf{1}_P\mathbf{1}_P^\top$$

Experimental Design

Exact design:

$$\xi := \begin{pmatrix} x_1 & \dots & x_P \\ J_1 & \dots & J_P \end{pmatrix}$$

$x_1, \dots, x_P$ - sub-regions

Approximate design:

$$\xi := \begin{pmatrix} x_1 & \dots & x_P \\ w_1 & \dots & w_P \end{pmatrix}, \quad w_i = J_i / J$$

$$\sum_{i=1}^{P} w_i = 1 \qquad \& \qquad w_i \geq 0$$

Experimental Design

Moment matrix:

$$\mathbf{M}(\xi) = \mathrm{diag}(w_1, \ldots, w_P)$$

For exact designs

$$\mathbf{M}(\xi) = \frac{1}{J}\mathbf{F}^\top\mathbf{F}, \quad \mathbf{F} = \text{block-diag}(\mathbf{1}_{J_1}, \ldots, \mathbf{1}_{J_P})$$

Search for optimal weights w_i^* to minimize

MSE matrix of $\hat{\alpha}$ or $\hat{\theta}$

A-Criterion for Genotype Effects

A-criterion for prediction of genotype effects $\boldsymbol{\alpha}$

$$\Phi_A = \mathrm{tr}\left(\mathrm{Cov}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\right)$$

$$\Phi_A(\xi) = const + \frac{c}{J}\,\mathrm{tr}\left[\left(\mathbf{I}_K \otimes \mathbf{M}(\xi) + \mathbf{B}\right)^{-1}\mathbf{H}\right]$$

$$\mathbf{B} = \left[\mathbf{I}_K \otimes \tilde{\mathbf{R}} + (\mathbf{T} \otimes \mathbf{I}_P)\,\tilde{\mathbf{U}}\,(\mathbf{T} \otimes \mathbf{I}_P)\right]^{-1} \quad \& \quad \mathbf{H} = \mathbf{B}\,(\mathbf{T}\otimes\mathbf{I}_P)\,\tilde{\mathbf{U}}^2(\mathbf{T}\otimes\mathbf{I}_P)\,\mathbf{B}$$

$$\mathbf{T} = \mathbf{I}_K - \frac{1}{K}\mathbf{1}_K\mathbf{1}_K^\top \quad \& \quad \tilde{\mathbf{U}} = J/c\,\mathbf{U} \quad \& \quad \tilde{\mathbf{R}} = J\,\mathbf{R}$$

K-Bayesian linear criterion but! $\mathbf{H}$ *singular*

Computational method proposed in P. (2025) cannot be directly used

A-Criterion for Pairwise Linear Contrasts

A-criterion for prediction of pairwise linear contrasts

$$\Phi_A = \mathrm{tr}\left(\mathrm{Cov}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right)$$

$$\Phi_A(\xi) = const + \frac{c}{J}\,\mathrm{tr}\left[\left(\mathbf{I}_K \otimes \mathbf{M}(\xi) + \mathbf{B}\right)^{-1} \tilde{\mathbf{H}}\right]$$

$$\mathbf{B} = \left[\mathbf{I}_K \otimes \tilde{\mathbf{R}} + (\mathbf{T} \otimes \mathbf{I}_P)\,\tilde{\mathbf{U}}\,(\mathbf{T} \otimes \mathbf{I}_P)\right]^{-1} \quad \& \quad \tilde{\mathbf{H}} = \mathbf{B}\,(\mathbf{T}\otimes\mathbf{I}_P)\,\tilde{\mathbf{U}}(\mathbf{T}\otimes\mathbf{I}_P)\tilde{\mathbf{U}}(\mathbf{T}\otimes\mathbf{I}_P)\,\mathbf{B}$$

$$\mathbf{T} = \mathbf{I}_K - \frac{1}{K}\mathbf{1}_K\mathbf{1}_K^\top \quad \& \quad \tilde{\mathbf{U}} = J/c\,\mathbf{U} \quad \& \quad \tilde{\mathbf{R}} = J\,\mathbf{R}$$

$\rightarrow$ $\tilde{\mathbf{H}}$ *singular*

$\rightarrow$ *Computational method has to be extended*

Example *(Kleinknecht et al. (2013))*

$P = 5$ sub-regions, $H = 3$ years, $J = 20$ locations

Variances:

$$\sigma_\omega^2 = 31 \quad \& \quad \sigma_\tau^2 = 18 \quad \& \quad \sigma_\gamma^2 = 160 \quad \& \quad \sigma_\phi^2 + \frac{1}{L}\sigma^2 = 333$$

Covariance matrix of genotype effects: $\mathbf{U} = \mathbf{N} \otimes \mathbf{V}$, $\mathbf{N}$ is AR(1)

$$\mathbf{N} = (\mathbf{N}_{ij})_{i,j=1,\ldots,K} \qquad \& \qquad \mathbf{N}_{ij} = \rho^{|i-j|}$$

$$\mathbf{V} = \begin{pmatrix} 567 & 254 & 239 & 485 & 328 \\ 254 & 155 & 118 & 240 & 162 \\ 239 & 118 & 155 & 226 & 153 \\ 485 & 240 & 226 & 488 & 310 \\ 328 & 162 & 153 & 310 & 215 \end{pmatrix}$$

Optimal Designs for A-Criterion for Genotype Effects

Optimal numbers of locations per sub-region

ρ	H	K	Approximate design ξ_a^*					Exact design ξ_e^*				
			w_1	w_2	w_3	w_4	w_5	J_1	J_2	J_3	J_4	J_5
1/3	2	5	0.42	0.05	0.09	0.39	0.05	8	1	2	8	1
		10	0.41	0.05	0.11	0.38	0.05	8	1	2	8	1
	6	5	0.36	0.09	0.16	0.34	0.05	7	2	3	7	1
		10	0.34	0.11	0.17	0.33	0.05	7	2	3	7	1
2/3	2	5	0.45	0.05	0.05	0.40	0.05	9	1	1	8	1
		10	0.42	0.05	0.10	0.38	0.05	8	1	2	8	1
	6	5	0.40	0.05	0.12	0.38	0.05	8	1	2	8	1
		10	0.36	0.10	0.15	0.34	0.05	7	2	3	7	1

$\rho = 0 \quad \rightarrow \quad \mathbf{N} = \mathbf{I}_K \quad \rightarrow$ same OD as in P. & Piepho (2024)

Computation using OptimalDesign Package in R

Literature

Harman, R. and Filová, L. (2019). Package 'OptimalDesign'. `https://cran.rproject.org/web/packages/OptimalDesign/index.html`.

Kleinknecht, K., Möhring, J., Singh, K., Zaidi, P., Atlin, G., and Piepho, H.-P. (2013). Comparison of the performance of BLUE and BLUP for zoned indian maize data. *Crop Science*, 53, 1384-1391.

Prus, M. (2025). Computing optimal allocation of trials to subregions in crop-variety testing in case of correlated genotype effects. *Statistica Neerlandica*

Prus, M. and Piepho, H.-P. (2021). Optimizing the allocation of trials to sub-regions in multi-environment crop variety testing. *Journal of Agricultural, Biological and Environmental Statistics*.

Literature

Prus, M. and Piepho, H.-P. (2024). Optimizing the allocation of trials to sub-regions in crop variety testing with multiple years and locations. *Journal of Agricultural, Biological and Environmental Statistics*

Bodnar T. and Prus, M. (2026) Dimension reduction for optimal design problems with Kronecker-product structure: Application in crop variety testing. *Submitted manuscript.*

Thank you for your attention!