Optimal Allocation of Trials to Sub-Regions in Multi-Environment Crop Variety Testing

Maryna Prus[1]

joint work with L. Filová[2], H.-P. Piepho[1] and W.A. Malik[1]

[1]Univesity of Hohenheim, Stuttgart [2]Comenius Univesity, Bratislava

UNIVERSITY OF HOHENHEIM

June 25, 2026

DFG Deutsche Forschungsgemeinschaft

Contents

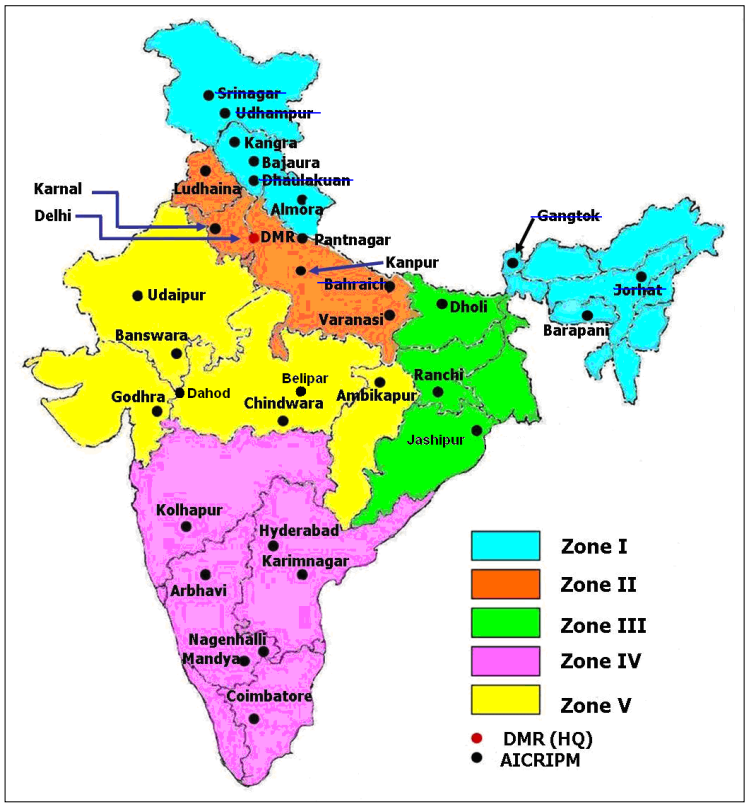
Introduction

- P sub-regions (zones), $P = 5$ on map
- J locations:
  - J_i locations in sub-region (zone) i
  - $J = J_1 + J_2 + \cdots + J_P$
- L blocks in each location
- K genotypes in each block
- H years, typically $H = 3$

To optimize: $J_1, \ldots, J_P$

How many locations in which sub-region?

Breeding sub-regions (zones) of maize in India



Kleinknecht et al. (2013), Crop Science

Linear Mixed Model

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

- μ_i - mean (fixed) effect of i-th sub-region
- η_h - effect of h-th year, $\mathrm{var}(\eta_h) = \sigma^2_\eta$
- λ_{ij} - effect of j-th location within i-th sub-region, $\mathrm{var}(\lambda_{ij}) = \sigma^2_\lambda$
- β_{ih} - effect of i-th sub-region in h-th year, $\mathrm{var}(\beta_{ih}) = \sigma^2_\beta$
- 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, $\mathrm{var}(\phi_{ijhk}) = \sigma^2_\phi$
- b_{ijhl} - effect of block l in location j in h-th year, $\mathrm{var}(b_{ijhl}) = \sigma^2_b$
- ε_{ijhkl} - observational error, $\mathrm{var}(\varepsilon_{ijhkl}) = \sigma^2$
- *All random effects and observational errors are uncorrelated and have zero mean*

Linear Mixed Model

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

$\boldsymbol{\alpha}_{ik}$ - effect of genotype k in sub-region i

$\boldsymbol{\alpha}_k = (\alpha_{1k}, \ldots, \alpha_{Pk})^\top$ - vector of genotype×sub-region effects of genotype k

$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \ldots, \boldsymbol{\alpha}_K^\top)^\top$ - vector of genotype×sub-reg. effects of all genotypes

$\mathrm{Cov}(\boldsymbol{\alpha}) = \mathsf{U}$, $\mathsf{U} = \mathsf{N} \otimes \mathsf{V}$

N - Kinship matrix with dimension $K \times K$

No marker data $\rightarrow$ $\mathsf{N} = \mathsf{I}_K$

$$\mathsf{N} = \begin{pmatrix} n_{11} & n_{12} & \cdots & n_{1K} \\ n_{21} & n_{22} & \cdots & n_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ n_{K1} & n_{K2} & \cdots & n_{KK} \end{pmatrix}$$

$n_{kk}\mathsf{V} = \mathrm{Cov}(\boldsymbol{\alpha}_k)$, matrix with dimension $P \times P$

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

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

Optimization Problem

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

under fixed total number of locations J

- for prediction of genotype effects

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

- for prediction of pairwise linear contrasts

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

Mean Squared Error (MSE) Matrix

Search for optimal numbers of locations J_i^* to minimize

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

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

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

$$\mathsf{M}(\xi) = \mathrm{diag}\left(\tfrac{J_1}{J}, \ldots, \tfrac{J_P}{J}\right), \qquad \xi := (J_1, \ldots, J_P)$$

$$c = \sigma_\gamma^2 + \frac{1}{H}\left(\sigma_\phi^2 + \frac{1}{L}\sigma^2\right) \quad \& \quad \mathsf{R} = \frac{\sigma_\tau^2}{cH}\mathsf{I}_P + \frac{\sigma_\omega^2}{cH}1_P 1_P^\top$$

A-Criterion for Genotype Effects

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

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

$$\Phi_A(\xi) = \mathrm{tr}\left[(\mathsf{I}_K \otimes \mathsf{M}(\xi) + \mathsf{B})^{-1}\,\mathsf{H}\right], \qquad \mathsf{M}(\xi) = \mathrm{diag}\left(\tfrac{J_1}{J}, \ldots, \tfrac{J_P}{J}\right)$$

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

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

K-Bayesian linear criterion (P., 2025, Stat. Neerl.)

→ For **small** number of genotypes computation by OptimalDesign package in R, MISOCP (Harman & Filová, 2025)

→ For **large** number of genotypes computation by modified version of algorithm proposed in Harman et al. (2015), Appl. Stoch. Models Bus. Ind.

Example

$P = 5$ sub-regions, $H = 3$ years

Variances and covariances *(Kleinknecht et al., 2013, Crop Science)*:

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

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

Kinship N: computed from two datasets with 80 and 698 genotypes

Covariance matrix of genotype effects: $\mathsf{U} = \mathsf{N} \otimes \mathsf{V}$

Optimal Designs for A-Criterion for Genotype Effects

Optimal and highly efficient numbers of locations per sub-region

J	$K = 80$						$K = 698$					
	J_1	J_2	J_3	J_4	J_5	MSE_{Tr}	J_1	J_2	J_3	J_4	J_5	MSE_{Tr}
10	4	1	1	3	1	17007	2	2	2	2	2	1967370718
20	7	2	3	7	1	12757	4	4	4	4	4	1966922683
40	13	6	7	12	2	9935	8	8	8	8	8	1966696245
100	28	17	20	27	8	7513	20	20	20	20	20	1966559528

Is balanced design always optimal for large number of genotypes? $\rightarrow$ **NO**

Counterexample *(P. et al., 2026)*:

Several families of genotypes with CS (compound symmetry) covariance structure within family and no correlation between families

Literature

Harman, R., Bachratá, A., and Filová, L. (2015). Construction of efficient experimental designs under multiple resource constraints. *Applied Stochastic Models in Business and Industry*

Harman, R. and Filová, L. (2025). 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., Filová, L., Piepho, H.-P., and Malik, W.A. (2026). Optimal allocation of trials to sub-regions in crop variety testing with multiple years and correlated genotype effects.
Preprint available under `http://arxiv.org/abs/2604.27831`

DFG Project

This research is funded by DFG via the project

"Optimal Design for Multi-Environment Trials"

led by M.P.

link: *https://gepris.dfg.de/gepris/projekt/533007746*

Thank you for your attention!

UNIVERSITY OF HOHENHEIM

Linear Mixed Model

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

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}

Counterexample: Block-diagonal N with CS Blocks

$$\mathsf{N} = \mathbb{I}_f \otimes \left(b_1 \mathbb{I}_m + b 1_m 1_m^\top \right), \quad b = \sigma_\alpha^2 \, r \quad \& \quad b_1 = \sigma_\alpha^2 (1 - r)$$

- f - number of families (groups) of genotypes
- m - number of genotypes per family, i.e. $mf = K$
- σ_α^2 - variance related coefficient within one family
- r - covariance related coefficient within one family
- No covariance between families

Counterexample: Block-diagonal N with CS Blocks

Optimal designs for block-diagonal kinship matrix N with CS blocks for $K = 900$ genotypes and $J = 40$ locations

r	f	m	J_1	J_2	J_3	J_4	J_5	MSE_{Tr}
1/2	6	150	14	4	6	15	1	231627
	15	60	15	4	6	13	2	151125
	60	15	14	5	6	13	2	118707
1/4	6	150	14	5	7	13	1	178343
	15	60	14	5	7	12	2	140899
	60	15	13	6	7	13	1	125245

Computation by OptimalDesign package, MISOCP

Optimal designs are not balanced for $K = 900$

Optimal Designs under Additional Linear Constraints

Sets of additional linear constraints

C_1 $J_i \geq 2, \quad i = 1, \ldots, 5$ - minimum number of locations fixed

C_2 $J/4 \geq J_i \geq 2, \quad i = 1, \ldots, 5$
- minimum and maximum numbers of locations fixed

C_3 $J_i \geq 2, \quad i = 1, \ldots, 5 \quad \& \quad 40J_1 + 44J_2 + 50J_3 + 65J_4 + 60J_5 \leq 50J$
- minimum number of locations and maximum cost fixed

C_4 $J/3 \geq J_i \geq 1, \quad i = 1, \ldots, 5 \quad \&$
$40J_1 + 44J_2 + 50J_3 + 55J_4 + 60J_5 \leq 50J$
- minimum and maximum numbers of locations and maximum cost fixed

Optimal Designs under Additional Linear Constraints

Optimal and highly efficient designs for A-criterion for 80 genotypes under additional constraint sets $C_1 - C_4$

J	J_1	J_2	J_3	J_4	J_5	MSE_{Tr}	J_1	J_2	J_3	J_4	J_5	MSE_{Tr}
	C_1						C_2					
20	7	2	3	6	2	12866	5	3	5	5	2	13278
40	13	6	7	12	2	9935	10	8	9	10	3	10104
100	28	17	20	27	8	7513	25	19	22	25	9	7538
	C_3						C_4					
20	9	2	2	5	2	13013	6	3	4	6	1	12904
40	15	6	6	11	2	9972	13	6	7	12	2	9935
100	33	17	19	24	7	7531	29	16	19	28	8	7513

C_1 $J_i \geq 2$ C_2 $J/4 \geq J_i \geq 2$

C_3 $J_i \geq 2$ & $40J_1 + 44J_2 + 50J_3 + 65J_4 + 60J_5 \leq 50J$

C_4 $J/3 \geq J_i \geq 1$ & $40J_1 + 44J_2 + 50J_3 + 55J_4 + 60J_5 \leq 50J$