

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

Maryna Prus and Hans-Peter Piepho



December 2, 2020



Contents

- 1 Model Specification
- 2 Best Linear Unbiased Prediction
- 3 Optimal design
- 4 Real Data Example
- 5 Literature

Experiment

- P sub-regions
- J locations:
 - J_i locations in sub-region i
 - $J = \sum_{i=1}^P J_i$
- K genotypes in each location
- L replications per genotype in each location
 - typically $L = 2$

To optimize: J_1, \dots, J_P

Linear mixed model

$$Y_{ijkl} = \mu_i + \alpha_{ik} + \lambda_{ij} + \gamma_{ijk} + b_{ijl} + \varepsilon_{ijkl}$$

- $i = 1, \dots, P$, P - number of sub-regions
- $j = 1, \dots, J_i$, J_i - number of locations in i -th sub-region
- $k = 1, \dots, K$, K - number of genotypes in each location
- $l = 1, \dots, L$, L - number of replications per genotype & location
- $J = \sum_{i=1}^P J_i$ - total number of locations

Linear mixed model

$$Y_{ijkl} = \mu_i + \alpha_{ik} + \lambda_{ij} + \gamma_{ijk} + b_{ijl} + \varepsilon_{ijkl}$$

- μ_i - mean (fixed) effect i -th sub-region
- λ_{ij} - effect of j -th location within i -th sub-region, $\text{var}(\lambda_{ij}) = \sigma_\lambda^2$
- γ_{ijk} - effect of genotype k in location j within sub-region i ,
 $\text{var}(\gamma_{ijk}) = \sigma_\gamma^2$
- b_{ijl} - effect of replication l in location j in sub-region i ,
 $\text{var}(b_{ijl}) = \sigma_b^2$

Linear mixed model

$$Y_{ijkl} = \mu_i + \alpha_{ik} + \lambda_{ij} + \gamma_{ijk} + b_{ijl} + \varepsilon_{ijkl}$$

- ε_{ijkl} - observational error, $\text{var}(\varepsilon_{ijkl}) = \sigma^2$
- α_{ik} - interaction effect of genotype k in sub-region i ,
 $\text{Cov}(\boldsymbol{\alpha}_k) = \sigma^2 \mathbf{D}$, \mathbf{D} positive definite
 $\boldsymbol{\alpha}_k = (\alpha_{1k}, \dots, \alpha_{pk})^\top$
- All random effects and observational errors have zero mean
- All random effects and observational errors are uncorrelated

Linear mixed model

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

- prediction of **genotype effects**

$$\alpha = (\alpha_1^\top, \dots, \alpha_K^\top)^\top$$

- prediction of **pairwise linear contrasts**

$$\theta^{k,k'} = \alpha_k - \alpha_{k'}, \quad k \neq k'$$

for given total number of locations J

BLUP for genotype effects

Best linear unbiased predictor for genotype effects $\hat{\alpha}_k$:

$$\hat{\alpha}_k = \left(\text{diag}(J_1, \dots, J_P) + \frac{Lv_2 + 1}{L} \mathbf{D}^{-1} \right)^{-1} \cdot \begin{pmatrix} J_1 (\bar{Y}_{1 \cdot k} - \bar{Y}_{1 \dots}) \\ \dots \\ J_P (\bar{Y}_{P \cdot k} - \bar{Y}_{P \dots}) \end{pmatrix}$$

$$\bar{Y}_{i \cdot k} = \frac{1}{J_i L} \sum_{j=1}^{J_i} \sum_{l=1}^L Y_{ijkl} \quad \& \quad \bar{Y}_{i \dots} = \frac{1}{J_i K L} \sum_{j=1}^{J_i} \sum_{k=1}^K \sum_{l=1}^L Y_{ijkl}$$

$$v_2 = \sigma_\gamma^2 / \sigma^2$$

BLUP for pairwise linear contrasts

Best linear unbiased predictor for pairwise linear contrasts $\theta^{k,k'}$:

$$\hat{\theta}^{k,k'} = \left(\text{diag}(J_1, \dots, J_P) + \frac{Lv_2 + 1}{L} \mathbf{D}^{-1} \right)^{-1} \cdot \begin{pmatrix} J_1 (\bar{Y}_{1 \cdot k \cdot} - \bar{Y}_{1 \cdot k' \cdot}) \\ \dots \\ J_P (\bar{Y}_{P \cdot k \cdot} - \bar{Y}_{P \cdot k' \cdot}) \end{pmatrix}$$

$$\bar{Y}_{i \cdot k \cdot} = \frac{1}{J_i L} \sum_{j=1}^{J_i} \sum_{l=1}^L Y_{ijkl} \quad \& \quad v_2 = \sigma_\gamma^2 / \sigma^2$$

Mean Squared Error (MSE) matrices

MSE matrix for predictor $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_K)^\top$ of genotype effects:

$$\text{Cov}(\hat{\alpha} - \alpha) = \sigma^2 \left[\frac{1}{K} \mathbf{1}_K \mathbf{1}_K^\top \otimes \mathbf{D} + (\mathbb{I}_K - \frac{1}{K} \mathbf{1}_K \mathbf{1}_K^\top) \otimes \left(\frac{1}{Lv_2 + 1} \mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1} \right)^{-1} \right]$$

$$\mathbf{F} = \text{block-diag}(\mathbf{1}_{LJ_1}, \dots, \mathbf{1}_{LJ_P})$$

MSE matrix for predictor $\hat{\theta}^{k,k'}$ of pairwise linear contrasts:

$$\text{Cov}(\hat{\theta}^{k,k'} - \theta^{k,k'}) = 2\sigma^2 \left(\frac{1}{Lv_2 + 1} \mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1} \right)^{-1}$$

Further θ instead of $\theta^{k,k'}$

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:

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

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

Experimental design

Moment matrix:

$$\mathbf{M}(\xi) = \text{diag}(w_1, \dots, w_P)$$

For exact designs

$$\mathbf{M}(\xi) = \frac{1}{LJ} \mathbf{F}^\top \mathbf{F}$$

Search for optimal weights w_i^ to minimize*

- a) MSE matrix of $\hat{\alpha}$*
- b) MSE matrix of $\hat{\theta}$*

A-criterion

A-criterion for prediction of **pairwise linear contrasts**

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

$$\Phi_A(\xi) = \text{tr} \left(\mathbf{M}(\xi) + \mathbf{C}^{-1} \right)^{-1}, \quad \mathbf{C} = \frac{LJ}{Lv_2 + 1} \mathbf{D}$$

Bayesian A-criterion

A-criterion for prediction of **genotype effects**

differs by a constant term \Rightarrow **same optimal designs**

Further focus on prediction of **pairwise linear contrasts**

Weighted A-criterion

Weighted A-criterion for prediction of **pairwise linear contrasts**

$$\Phi_{A_w} = \sum_{i=1}^P \ell_i \text{var}(\hat{\theta}_i - \theta_i)$$

$\ell_1, \dots, \ell_P > 0$ - sub-regions' coefficients

$$\Phi_{A_w}(\xi) = \text{tr} \left(\mathbf{L} (\mathbf{M}(\xi) + \mathbf{C}^{-1})^{-1} \right)$$

$$\mathbf{L} = \text{diag}(\ell_1, \dots, \ell_P)$$

Bayesian linear criterion

Example: Compound Symmetry

Compound Symmetry (CS) covariance structure:

$$\mathbf{D} = a \mathbf{1}_P \mathbf{1}_P^\top + b \mathbb{I}_P, \quad b > 0 \quad \& \quad a > -b/P$$

Remark (1)

In the CS model the (balanced) design ξ_P with $w_i = \frac{1}{P}$, $i = 1, \dots, P$, is A-optimal for the prediction of the genotype effects and their pairwise linear contrasts

Remark (2)

The balanced design ξ_P and the weighted design ξ_ℓ with $w_i = \ell_i/\ell$, $\ell = \sum_{i=1}^P \ell_i$, $i = 1, \dots, P$ are in general not optimal with respect to the weighted A-criterion for the prediction of the pairwise linear contrasts

Counter-example

$P = 2$ sub-regions, $L = 1$ replication, $J \geq 5$ locations

Variance parameters: $v_2 = 1$ & $\mathbf{D} = \text{diag}(1, 1)$

Sub-regions' coefficients: $l_1 = 4$ & $l_2 = 1$

Optimal designs:

$$\xi := \begin{pmatrix} x_1 & x_2 \\ w & 1 - w \end{pmatrix}, \quad w = J_1/J$$

Moment matrix:

$$\mathbf{M} = \text{diag}(w, 1 - w)$$

Counter-example

Weighted A-criterion:

$$\Phi_{A_w}(\xi) = \frac{4}{w + 2/J} + \frac{1}{1 - w + 2/J}$$

Optimal weight:

$$w^* = \frac{2J + 2}{3J}$$

Numbers of locations:

J	15	30	45	60
J_1^*	11	21	31	41
J_1^ℓ	12	24	36	48

$$J_1^* = w^* J \quad \& \quad J_1^\ell = 0.8J$$

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

$P = 5$ sub-regions, $L = 2$ replication

Variance parameters:

$$\sigma^2(v_2 + 1) = 493 \quad \& \quad \sigma^2 \mathbf{D} = \mathbf{V} + 31 \mathbf{1}_5 \mathbf{1}_5^T + 18 \mathbb{I}_5$$

$\Rightarrow v_2$ & \mathbf{D} depend on σ^2

\mathbf{V} in factor-analytic (FA) model

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

\mathbf{V} in compound symmetry model

$$\begin{pmatrix} 308 & 270 & 270 & 270 & 270 \\ 270 & 308 & 270 & 270 & 270 \\ 270 & 270 & 308 & 270 & 270 \\ 270 & 270 & 270 & 308 & 270 \\ 270 & 270 & 270 & 270 & 308 \end{pmatrix}$$

Optimal designs: A-criterion

Optimal numbers of locations per sub-region in FA model

J	σ^2	Approximate design ξ_a^*					Eff _a	Exact design ξ_e^*					Eff _e
		w_1	w_2	w_3	w_4	w_5		J_1	J_2	J_3	J_4	J_5	
10	50	0.37	0.10	0.10	0.33	0.10	0.91	4	1	1	3	1	0.91
	200	0.37	0.10	0.10	0.33	0.10	0.92	4	1	1	3	1	0.92
	400	0.34	0.10	0.15	0.31	0.10	0.94	3	1	2	3	1	0.94
20	50	0.33	0.13	0.18	0.31	0.04	0.95	7	3	3	6	1	0.95
	200	0.31	0.15	0.19	0.29	0.06	0.96	6	3	4	6	1	0.96
	400	0.29	0.16	0.20	0.27	0.09	0.97	6	3	4	5	2	0.97
40	50	0.27	0.17	0.20	0.25	0.10	0.97	11	7	8	10	4	0.98
	200	0.26	0.18	0.20	0.24	0.12	0.98	10	7	8	10	5	0.98
	400	0.25	0.19	0.21	0.23	0.13	0.98	10	8	8	9	5	0.99

$$\text{Eff}_a = \Phi_A(\xi_a^*) / \Phi_A(\xi_P) \quad \& \quad \text{Eff}_e = \Phi_A(\xi_e^*) / \Phi_A(\xi_P)$$

In CS model balanced design ξ_P is optimal

Optimal designs: weighted A-criterion

Sub-regions' coefficients:

l_1	l_2	l_3	l_4	l_5
813685	432716	477365	995298	1174818

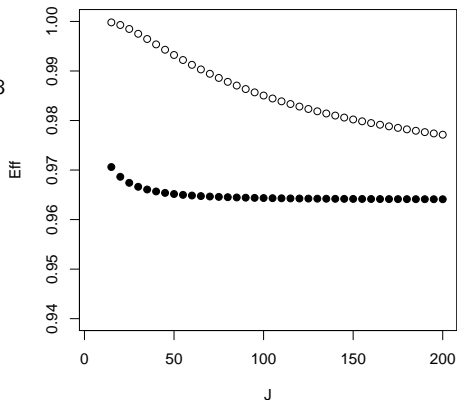
Weighted design:

$$\xi_\ell = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0.21 & 0.11 & 0.12 & 0.26 & 0.30 \end{pmatrix}$$

On graphic:

- - efficiency of ξ_ℓ
- - efficiency of ξ_P

in CS model



Optimal designs: weighted A-criterion

Optimal numbers of locations per sub-region in FA model

J	σ^2	Approximate design ξ_a^*					$\text{Eff}_{a,P}$	$\text{Eff}_{a,\ell}$	Exact design ξ_e^*				
		w_1	w_2	w_3	w_4	w_5			J_1	J_2	J_3	J_4	J_5
10	50	0.35	0.10	0.10	0.35	0.10	0.88	0.92	3	1	1	4	1
	200	0.34	0.10	0.10	0.36	0.10	0.89	0.93	3	1	1	4	1
	400	0.34	0.10	0.10	0.36	0.10	0.90	0.94	3	1	1	4	1
20	50	0.34	0.05	0.09	0.37	0.15	0.91	0.95	7	1	2	7	3
	200	0.33	0.05	0.11	0.35	0.16	0.92	0.96	7	1	2	7	3
	400	0.30	0.08	0.12	0.32	0.18	0.94	0.97	6	2	2	6	4
40	50	0.28	0.09	0.13	0.30	0.19	0.94	0.97	11	4	5	12	8
	200	0.27	0.10	0.14	0.29	0.20	0.95	0.98	11	4	6	11	8
	400	0.27	0.10	0.14	0.29	0.20	0.96	0.98	10	5	6	11	8

$$\text{Eff}_{a,P} = \frac{\Phi_{A_w}(\xi_a^*)}{\Phi_{A_w}(\xi_P)} \quad \& \quad \text{Eff}_{a,\ell} = \frac{\Phi_{A_w}(\xi_a^*)}{\Phi_{A_w}(\xi_\ell)}$$

Literature

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

Harman, R. and Prus, M. (2018). Computing optimal experimental designs with respect to a compound Bayes Risk criterion. *Statistics and Probability Letters*, 137, 135-141.

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. 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*.

Thank you for your attention!

