Bayesian modelling of variance components makes analysis of resolvable incomplete block designs more efficient

Marcin Studnicki[1], Hans-Peter Piepho[2]

[1]Warsaw University of Life Sciences, Nowoursynowska 159, 02-776 Warsaw, Poland;
[2]University of Hohenheim, Fruwirthstraße 23, 70599 Stuttgart, Germany

WARSAW UNIVERSITY OF LIFE SCIENCES

Introduction

- Due to the development of statistical and computational methods, there is an are increasingly of interest in adopting Bayesian approaches to METs data analysis,
- One of the advantages of the Bayesian approach is the possibility to define a prior probability distribution for parameters of interest,
- This allows taking into account the knowledge about parameters available before a new experiment,

Introduction

- A prior distribution can be based on information available for the cultivars, its pedigree and trial locations

- The use of such additional information in cultivars evaluation and METs analysis is very rare

Introduction

In order to increase the precision of the analysis of individual trials laid out as alpha designs in MET, we make a proposal to create a prior distribution of variance components for replicates, blocks and plots, based on the results of previous (historical) trials. We propose different modelling approaches for the prior distributions. Additional, we evaluate the effectiveness of the Bayesian approach to the REML method classically used with MET.

Material and Methods

Historical data set

- come from real field trials with wheat of the Polish Post-Registration Variety Testing System
- grain yield from trials carried out between the 2009/2010 and 2019/2020 growing seasons
- each trials are carried out as alpha designs with two replicates, with the number of blocks depending on the number of cultivars

Material and Methods

Model

For individual trials laid out as alpha-designs we assume the linear model

$$y_{ijl}=\mu+\tau_i+\gamma_j+\rho_{l(j)}+\varepsilon_{ijl}$$

where, y_{ijl} is the yield response of the ith cultivar in the jth replicate and lth block within the jth replicate, μ is the overall mean, τ_i is the fixed effect of the ith cultivar, γ_j is the random effect of the jth replication, $\rho_{l(j)}$ is the random effect of the lth block nested in the jth replication, and εijl is the error effect.

Material and Methods

For each individual trial the sequential sums of squares (Type I SS in SAS) were obtain for replicates, SS_r, blocks, SS_b, and error, SS_e, as well as their expected mean squares, which are functions of the three variance components for replicates σ_r^2, blocks σ_b^2 and error σ_r^2 (Table 1).

.

Table 1: ANOVA table for random effects considerate in study linear model.

Source	Degrees of freedom (d.f.)	Sum of squares (SS)	Mean square (MS)	Expected mean squares E(MS)
Replicates	v_r	SS_r	$MS_r = SS_r / v_r$	$\sigma_e^2 + c_1\sigma_b^2 + c_2\sigma_r^2$
Blocks	v_b	SS_b	$MS_b = SS_b / v_b$	$\sigma_e^2 + c_3\sigma_b^2$
Error	v_e	SS_e	$MS_e = SS_e / v_e$	σ_e^2

WARSAW
UNIVERSITY
OF LIFE SCIENCES

Material and Methods

Conditionally on the variance components $\theta = (\sigma_r^2, \sigma_b^2, \sigma_e^2)^T$, the sums of squares have scaled central chi-squared distributions, i.e.

$$\frac{SS_i \mid \theta}{E(MS_i)} \sim \chi^2_{\nu_i} \qquad (i = r, b, e)$$

For the variance component vector θ, we assume four different prior distribution specifications.

Material and Methods

Prior dystrybution

The first is a trivariate log-normal distribution, given by

$$\begin{pmatrix} \log(\sigma_r^2) \\ \log(\sigma_b^2) \\ \log(\sigma_e^2) \end{pmatrix} \sim MVN\left[\begin{pmatrix} \theta_r \\ \theta_b \\ \theta_e \end{pmatrix}, \begin{pmatrix} \phi_r^2 & \phi_{rb} & \phi_{re} \\ & \phi_b^2 & \phi_{be} \\ & & \phi_e^2 \end{pmatrix}\right]$$

For this prior distribution we have two sets of hyperparameters,
1) the vector θ of log-variance component means
2) the variance-covariance matrix ϕ of the log-variance components

$$\theta = \begin{bmatrix} \theta_r \\ \theta_b \\ \theta_e \end{bmatrix} \qquad \phi = \begin{bmatrix} \phi_r^2 & \phi_{rb} & \phi_{re} \\ & \phi_b^2 & \phi_{be} \\ & & \phi_e^2 \end{bmatrix}$$

Material and Methods

The second prior, to ensure that the parameters of the variance-covariance matrix for the log-variances are properly constrained to suitable values, we used the following parameterizations of trivariate log-normal distribution:

$$\phi_r^2 = \exp(\lambda_{rr}) \qquad \pi_{rb} = \frac{\exp(2\lambda_{rb}) - 1}{\exp(2\lambda_{rb}) + 1}, \quad \phi_{rb} = \pi_{rb}\phi_r\phi_b$$

$$\phi_b^2 = \exp(\lambda_{bb}) \qquad \pi_{re} = \frac{\exp(2\lambda_{re}) - 1}{\exp(2\lambda_{re}) + 1}, \quad \phi_{re} = \pi_{re}\phi_r\phi_e$$

$$\phi_e^2 = \exp(\lambda_{ee}) \qquad \pi_{be} = \frac{\exp(2\lambda_{be}) - 1}{\exp(2\lambda_{be}) + 1}, \quad \phi_{be} = \pi_{be}\phi_b\phi_e$$

The exponential specification for the variances makes sure these are positive, whereas the use of the inverse of Fisher's z transformation ensures that a correlation π obeys the constraint $|\pi| < 1$.

Material and Methods

Prior distribution

Considered other prior distributions for the variances, we use the gamma and inverse gamma distributions. For the these prior distributions, the shape parameter α, the scale parameter β and variance-covariance matrix of the log-variance components ϕ constitute hyper-parameters.

Material and Methods

Computational methods for Bayesian approach:

- empirical Bayes approach - PROC NLMIXED in SAS, used maximum likelihood estimation by adaptive Gauss-Hermite quadrature

- full Bayes approach – PROC MCMC in SAS, used the Markov Chain Monte Carlo (MCMC) method with Gibbs sampling

Material and Methods

Prior distribution

When we want implementing gamma and inverse gamma in the NLMIXED procedure of SAS, we faced the challenge that these only allow normally distributed random effects.

To fit these two typ of prior distribution, we use the inverse cumulative distribution function (c.d.f.) method.

WARSAW
UNIVERSITY
OF LIFE SCIENCES

Material and Methods

Simulations study

- Based on the historical data set we simulated new individual trials (n = 1000).
- We use two schemes of simulations new trials
  - with a small number of blocks (2 blocks) and small size of blocks (3 plot per block)
  - with a relatively large number (10 blocks) and size of blocks (10 plot per block).

Material and Methods

Simulations study

Evaluation of prior distribution and computational methods

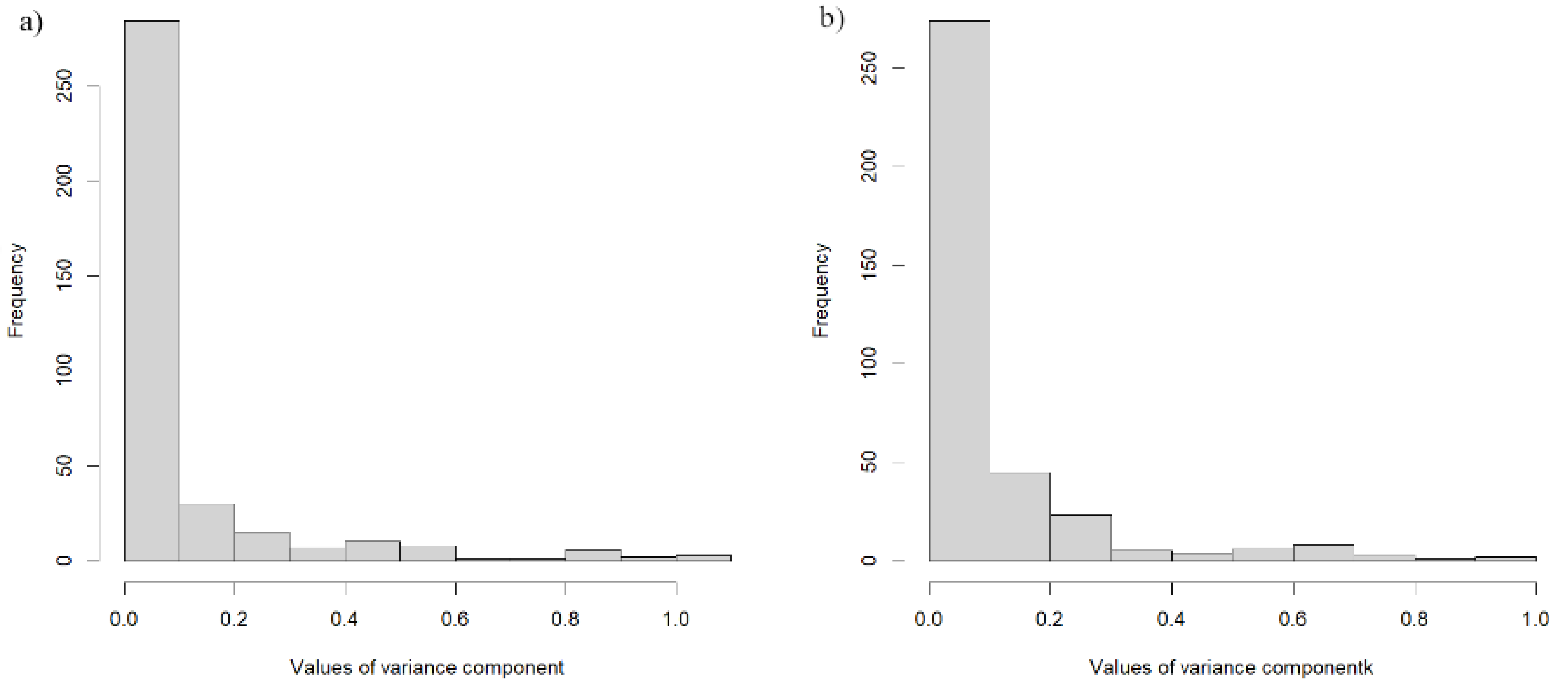
- the mean squared error of estimated treatment differences (MSED)
- the mean squared error (MSE)

WARSAW
UNIVERSITY
OF LIFE SCIENCES

Results

Historical data set – results of REML analysis

Figure 1. Values of variance components for replicate effects (a) and block effects (b) across individual trials from historical data set estimated by REML model.



WARSAW
UNIVERSITY
OF LIFE SCIENCES

Results

Symulation study - MSED

Table 2. The mean squared error of estimated treatment differences (MSED) for study models in two simulated data sets.

Model		3 plots per block	10 plots per block
REML		2.6126	2.0755
NLMIXED	3-variate normal (1)	1.9195	1.9965
	3-variate normal (2)	1.9181	1.9443
	Gamma	1.9099	1.9221
	Invers gamma	1.8919	1.9081
Full Bayesian	3-variate normal (1)	1.8543	1.9906
	3-variate normal (2)	1.8291	1.8996
	Gamma	1.8011	1.9034
	Invers gamma	1.8102	1.8855

Results

Symulation study - MSE

Table 3. The mean squared error (MSE) of variance components for study models in two simulated data sets.

Model		Variance components for replication		Variance components for block		Variance components for error	
		3 plots per block	10 plots per block	3 plots per block	10 plots per block	3 plots per block	10 plots per block
REML		0.4726	0.2545	0.3953	0.1229	0.0369	0.0131
NLMIXED	3-variate normal (1)	0.3696	0.2298	0.3133	0.1109	0.0321	0.0123
	3-variate normal (2)	0.3687	0.2292	0.3128	0.1099	0.0321	0.0121
	Gamma	0.3483	0.2291	0.3051	0.1091	0.0320	0.0119
	Invers gamma	0.3466	0.2284	0.3048	0.1084	0.0320	0.0121
Full Bayesian (MCMC)	3-variate normal (1)	0.3511	0.2265	0.3088	0.1017	0.0317	0.0109
	3-variate normal (2)	0.3502	0.2261	0.3081	0.1012	0.0323	0.0111
	Gamma	0.3398	0.226	0.3029	0.1019	0.0321	0.0109
	Invers gamma	0.3387	0.2244	0.3022	0.1006	0.0318	0.0103

Conclusions

- **The Bayesian approach with prior distribution built on historical data set demonstrated non-negligible gains in precision and recovery inter-block information in compere to REML methods**

- **This benefit of the Bayesian approach is more strongly observed for simulated data in the 3 plots per block scheme**

- **The difference between the two computational methods was small, but in favour of full Bayes approach**

WARSAW
UNIVERSITY
OF LIFE SCIENCES

Thank you for your attention