The Journal of Agricultural Science

cambridge.org/ags

Yielding stability of early maturing potato varieties: Bayesian analysis

M. Przystalski and T. Lenartowicz

Research Centre for Cultivar Testing, 63-000 Słupia Wielka, Poland

Crops and Soils Research Paper

Cite this article: Przystalski M, Lenartowicz T (2020). Yielding stability of early maturing potato varieties: Bayesian analysis. *The Journal of Agricultural Science* **158**, 564–573. https://doi.org/10.1017/S0021859620000945

Received: 25 March 2020
Revised: 30 September 2020
Accepted: 5 October 2020
First published online: 11 November 2020

Key words:
Environmental variance model; hierarchical model; series of field trials

Author for correspondence:
M. Przystalski, E-mail: marprzyst@gmail.com

Abstract

Field trials conducted in multiple years across several locations play an essential role in plant breeding and variety testing. Usually, the analysis of the series of field trials is performed using a two-stage approach, where each combination of year and site is treated as environment. In variety testing based on the results from the analysis, the best varieties are recommended for cultivation. Under a Bayesian approach, the variety recommendation process can be treated as a formal decision theoretic problem. In the present study, we describe Bayesian counterparts of two stability measures and compare the varieties in terms of the posterior expected utility. Using the described methodology, we identify the most stable and highest tuber yielding varieties in the Polish potato series of field trials conducted from 2016 to 2018. It is shown that variety Arielle was the highest yielding, the third most stable variety and was the second best variety in terms of the posterior expected utility. In the present work, application of the Bayesian approach allowed us to incorporate the prior knowledge about the tested varieties and offered a possibility of treating the variety recommendation process as a formal decision process.

Introduction

Field trials conducted in multiple years across several locations play an essential role in plant breeding and variety testing. In plant breeding, the lines are evaluated and the most promising lines are selected for further breeding. In variety testing, based on the results from the analysis of a series of field trials, the best varieties are recommended for cultivation. Usually, the analysis of the series of trials is performed using a two-stage approach, where each combination of year and site is treated as environment. In literature, several stability measures have been described, e.g. Shukla's stability variance (Shukla, 1972), regression on the environmental mean approach to assessments of stability (Finlay and Wilkinson, 1963; Eberhart and Russell, 1966; Digby, 1979), additive main effects and multiplicative interaction (AMMI) model (Gauch, 1988) or genotype main effects and genotype-by-environment interaction effects (GGE) model (Yan and Kang, 2003). However, with the increase of computing power of computers and the development of statistical methods, the assessment of stability based on linear mixed models has become more popular (see e.g. Piepho, 1999; Piepho and van Eeuwijk, 2002; Smith and Cullis, 2018; Studnicki *et al.*, 2017, 2019). In Piepho (1999), the above mentioned models were defined in terms of variance components.

In Poland, new varieties of important species are assessed before the registration in value-for-cultivation-and use (VCU) trials, and further in post-registration trials. Based on the results of post-registration trials, a recommendation to the farmers is given. Only stable and high-yielding varieties are recommended for cultivation. Looking at the variety recommendation process from the decision-theory perspective, the problem of recommending varieties for cultivation can be treated as a decision-making problem in a situation of uncertainty. In literature, similar problems were solved by using Bayesian decision theory (see e.g. Theobald and Talbot, 2002; Theobald *et al.*, 2006). For this reason, in the present work, a Bayesian approach was used. As in Theobald *et al.* (2002, 2006), we first fitted a Bayesian linear mixed model, and next the varieties were compared in terms of the posterior expected utility.

The Bayesian analysis has some benefits over the frequently used models in stability analysis. First of all, most of the frequently used stability measures are likelihood-based approaches. In Searle *et al.* (2006) (see also Harville, 1977), it was pointed out that the problem with likelihood-based estimation with random effects is that, except the simplest situations of balanced and orthogonal data, the sampling distributions of variances are unknown. In consequence, since the sampling distributions do not have analytical closed-form solutions, it is very difficult to properly account for estimation error associated with variances when they are used to compute weights for estimation of means. In contrast to likelihood-based estimation, Bayesian estimation produces posterior distributions that properly account for uncertainty in the estimation of all parameters. Moreover, in Theobald *et al.* (2002), it was

© The Author(s), 2020. Published by Cambridge University Press

CAMBRIDGE UNIVERSITY PRESS

Table 1. Sites used in the 3-year variety trials conducted from 2016 to 2018

	Geographic coordinates			Year		
Site	Latitude	Longitude	m a.s.l.	2016	2017	2018
Białogard (3)	54°00′N	16°00′E	24	X	x	x
Bobrowniki (18)	52°49′N	17°18′E	90	X	x	x
Cicibór Duży (48)	52°05′N	23°07′E	114	X	x	x
Chrząstowo (14)	53°11′N	17°35′E	10	X	–	–
Głębokie (15)	52°39′N	18°27′E	85	–	x	x
Karzniczka (1)	54°29′N	17°17′E	80	X	x	x
Kościelna Wieś (21)	51°48′N	18°01 E	120	X	x	x
Krzyżewo (11)	53°01′N	22°46′E	135	X	x	x
Masłowice (25)	51°15′N	18°38′E	174	X	x	x
Nowy Lubliniec (45)	50°16′N	23°06′E	274	X	x	x
Pawłowice (41)	50°28′N	18°29′E	240	x	x	x
Rarwino (4)	53°56′N	14°50′E	10	x	x	–
Rychliki (10)	53°59′N	19°32′E	80	x	x	x
Seroczyn (26)	52°00′N	21°56′E	180	x	x	x
Słupia (43)	50°38′N	19°58′E	290	x	x	x
Słupia Wielka (19)	52°13′N	17°14′E	85	x	x	–
Sulejów	51°21′N	19°52′E	188	x	x	x
Szczecin-Dąbie (5)	53°23′N	14°40′E	9	x	x	x
Świebodzin (23)	52°14′N	15°35′E	90	x	x	x
Tarnów (48)	50°35′N	16°47′E	300	x	x	x
Uhnin (49)	51°34′N	23°02′E	157	x	x	x
Węgrzce (39)	50°07′N	19°59′E	285	x	x	x
Wrócikowo (8)	53°49′N	20°40′E	142	x	x	x
Zybiszów (29)	51°04′N	16°55′E	130	x	x	x

pointed out that a Bayesian approach can alleviate problems associated with estimating complex models, such as zero estimates of variance components.

The use of the Bayesian approach in the context of agricultural field experiments is rare (Theobald and Talbot, 2002; Theobald *et al.*, 2002; Edwards and Jannink, 2006; Theobald *et al.*, 2006; Crossa *et al.*, 2011; Josse *et al.*, 2014; Orellana *et al.*, 2014; de Oliveira *et al.*, 2016; Edwards and Orellana, 2015; Bernardo *et al.*, 2018) and is mainly focused on AMMI models (Crossa *et al.*, 2011; Josse *et al.*, 2014, Bernardo *et al.*, 2018) and GGE biplots (de Oliveira *et al.*, 2016). The Bayesian counterpart of the Finlay–Wilkinson model has been described in Lian and de los Campos (2016). In the present study, Bayesian counterparts of two stability measures described in Piepho (1999) are given. In contrast to Edwards and Jannink (2006), the stability measures are defined in terms of Kronecker products. Moreover, by adopting the approaches described in Theobald *et al.* (2002, 2006) to our variety recommendation problem, we compare varieties in terms of the posterior expected utility. Using the described methodology, we identify the most stable and highest yielding potato varieties for a Polish series of field trials conducted in years 2016–2018. As in Edwards and Jannink (2006) and Orellana *et al.* (2014), the stability analysis was performed under the heterogeneity of error variances assumption.

Materials and methods

Data

The data set consists of potato field trials performed in years 2016–2018. The trials used in the study were conducted in experimental stations (sites) belonging to the Research Centre for Cultivar Testing (COBORU) (Table 1).

Most of the trials were organized in a randomized complete block design with three replicates. Only in 2018, at six sites, the trials were conducted according to a 1-resolvable design with three replicates and two blocks. On each plot, there were 60 plants planted. Plots that experienced unusual damage, such as boar damage, were excluded from further analyses. During the 3 years of study, in total, 22 varieties were tested. Since we were interested in assessing yielding stability, we choose to analyse varieties which were tested for 3 years. The list of varieties used in the present study with their country of origin is given in Table 2. During the 3-year period, most of the varieties were tested in all sites, only variety Everest was tested at five sites in years 2016 and 2017.

https://doi.org/10.1017/S0021859620000945 Published online by Cambridge University Press

Table 2. Varieties used in the 3-year variety trials conducted from 2016 to 2018 and their country of origin

		Year		
Variety	Country	2016	2017	2018
Arielle	The Netherlands	X	X	x
Denar	Poland	x	X	x
Everest	The Netherlands	x	x	x
Impala	The Netherlands	x	x	x
Lord	Poland	x	x	x
Miłek	Poland	x	x	x
Riviera	The Netherlands	x	x	x
Viviana	Germany	x	x	x

In Polish potato VCU and post-registration trials, one of the analysed characteristics is tuber yield, which is observed on plots. The observed tuber yields are expressed in decitonnes per hectare (dt/ha).

Statistical analysis

The observations were modelled according to a linear mixed model. For the sake of clarity, throughout the paper by environment we will mean a combination of year and location.

In the analysed data set, 1-resolvable design was used only in six out of 67 environments. Speed *et al.* (1985) have shown that dropping the block effect for this design provides still a valid analysis. For this reason, we omitted in our model the effect of blocks nested within replicates and environments.

Let y_{ikr} be the potato tuber yield for the i-th variety ($i = 1, \ldots, I$) at the k-th environment ($k = 1, \ldots, K$) and in the r-th replicate ($r = 1, \ldots, R$). Then the model can be written as:

$$y_{ikr} = \mu_i + u_k + v_{ik} + w_{kr} + e_{ikr} \quad (1)$$

where μ_i denotes the mean of the i-th variety. By u_k, v_{ik}, w_{kr} and e_{ikr} we denote in model (1) the random effects of environments, variety × environment interaction, of replicates nested within environment and of errors, respectively. Using vector notation, model (1) can be written as:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{u} + \mathbf{Z}_2\boldsymbol{v} + \mathbf{Z}_3\boldsymbol{w} + \boldsymbol{e} \quad (2)$$

where $\boldsymbol{y}$ is a vector of potato tuber yields and $\boldsymbol{\beta}$ is the vector of fixed variety means. By $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$, we denote in model (2) the random vectors of environment effects, of environment × variety interaction effects, and of replicates nested within environment effects, respectively. Further in model (2), by $\boldsymbol{e} = [\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_K]'$, where $\boldsymbol{e}_k = [e_{1,k,1}, \ldots, e_{I,k,R}]'$ ($k = 1, \ldots, K$), we denote the random vector of errors. The matrices $\mathbf{X}$, $\mathbf{Z}_1$, $\mathbf{Z}_2$ and $\mathbf{Z}_3$ in model (2) are design matrices for $\boldsymbol{\beta}$, $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$, respectively.

To fit the model, a hierarchical Bayesian linear mixed model was used. The model has three hierarchies or stages. In the first, it was assumed that observations are exchangeable samples from normal distribution:

$$\boldsymbol{y}|\boldsymbol{\beta},\ \boldsymbol{u},\ \boldsymbol{v},\ \boldsymbol{w}, \mathbf{R}_0 \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\boldsymbol{u} + \mathbf{Z}_2\boldsymbol{v} + \mathbf{Z}_3\boldsymbol{w},\ \oplus \sigma_k^2\mathbf{I}_{n_k})$$

Table 3. Summary statistics for tuber yield: mean, standard deviation (s.d.), minimum (min), median (med), maximum (max) of observations for tuber yield

	Tuber yield [t/ha]				
Year	mean	s.d.	min	med	max
2016	57.21	11.15	14.86	54.42	78.19
2017	48.13	11.15	20.34	47.21	85.99
2018	42.51	15.89	8.95	41.71	85.17

where $\mathbf{R}_0 = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2)$ is a matrix of residual variances, n_k is the number of plots in the k-th environment, and $\oplus$ denotes the direct sum of matrices. In case when the number of plots is equal in all trials, the above covariance matrix can be written as $\mathbf{R}_0 \otimes \mathbf{I}$, where $\otimes$ denotes the Kronecker product.

In the second, the prior distributions are assigned to $\boldsymbol{\beta}$, $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$; these correspond to the assumptions made on fixed and random effects in a Bayesian linear model. For the vector of variety means $\boldsymbol{\beta}$ a vague, large-variance Gaussian prior was assigned, i.e.

$$\boldsymbol{\beta} \sim N(0, 10^8\mathbf{I})$$

The vector of environment effects $\boldsymbol{u}$ was assigned a normal distribution with mean vector of zeros as:

$$\boldsymbol{u}|\sigma_u^2 \sim N(0, \sigma_u^2\mathbf{I})$$

whereas the vector of replicates nested within environments $\boldsymbol{w}$ was assigned a normal distribution with mean vector of zeros as:

$$\boldsymbol{w}|\sigma_w^2 \sim N(0, \sigma_w^2\mathbf{I})$$

For each trial ($k = 1, \ldots, K$), we assumed that $\boldsymbol{e}_k|\sigma_k^2 \sim N(0, \sigma_k^2\mathbf{I}_{n_k})$. This implies that conditionally on $\sigma_1^2, \ldots, \sigma_K^2$, the vector $\boldsymbol{e}$ follows

$$\boldsymbol{e}|\mathbf{R}_0 \sim N(0,\ \oplus \sigma_k^2\mathbf{I}_{n_k})$$

The vector of environment × variety interaction $\boldsymbol{v}$ was assigned a normal distribution with mean vector of zeros and covariance matrix $\mathbf{G}_0 \otimes \mathbf{I}$, i.e.

$$\boldsymbol{v}|\mathbf{G}_0 \sim N(0, \mathbf{G}_0 \otimes \mathbf{I}).$$

Since we are interested in identifying the most stable varieties, by assuming different covariance matrices and modifying the list of random vectors in (2), we obtained different models. Our base model is Shukla model (SH) for which $\mathbf{G}_0 = \mathrm{diag}(\sigma_{11}^2, \ldots, \sigma_{II}^2)$. Next, by removing from the model (2) the random vector of environment effects and assuming that $\mathbf{G}_0$ is completely unrestricted covariance matrix (i.e. the elements in the symmetric variance-covariance matrix $\mathbf{G}_0$ may take any value, as long $\mathbf{G}_0$), we obtained the environmental variance model (ENVIR; Piepho, 1999). Since in ENVIR model, the random vector of environment effects $\boldsymbol{u}$ is excluded, the random vector $\boldsymbol{v}$ is no longer vector of interaction effects rather a random vector of environment-variety effects. For both models, the diagonal elements of covariance matrix $\mathbf{G}_0$ are variances corresponding to the I varieties, which are interpreted as stability measures.

https://doi.org/10.1017/S0021859620000945 Published online by Cambridge University Press

Table 4. List of models, covariance matrices for 3-year series, and DIC values for different models

Model	Random terms[a]	Covariance matrix for the vector of variety × environment effects (***v***)	DIC				
			Chain 1	Chain 2	Chain 3	Chain 4	Chain 5
SH	Env + Env.Var + Env.Rep	$\text{diag}(\sigma_1^2, \ldots, \sigma_{II}^2) \otimes \mathbf{I}$	6434.633	6434.706	6434.679	6434.660	6434.648
ENVIR	Env.Var + Env.Rep	$\boldsymbol{\Sigma} \otimes \mathbf{I}$	6422.215	6422.239	6422.419	6422.094	6422.354

DIC, deviance information criterion.
[a]Env, environment; Var, variety; Rep, replicate; Env.Rep, effect of replicates nested within environments. Depending on the model Env.Var denotes either environment × variety interaction (SH) or environment-variety effect (ENVIR).

In the final or last stage, prior distributions were assigned to hyperparameters σ_u^2, σ_w^2, $\mathbf{R}_0$ and $\mathbf{G}_0$. We assumed that both σ_u^2 and σ_w^2 follow an inverse-Wishart distribution with 2.002 degrees of freedom and one-dimensional scale matrix equal to $(0.002/2.002)\mathbf{I}$. For covariance matrix $\mathbf{G}_0$, as in Gelman *et al.* (2014), we assigned inverse-Wishart distribution prior distribution with $\dim\mathbf{G}_0 + 1$ degrees of belief. In our stability analysis, for covariance matrix $\mathbf{G}_0$ as the scale matrix, we used diagonal matrix of order $\dim\mathbf{G}_0$ with elements equal to empirical variances of each variety divided by $\dim\mathbf{G}_0$. Finally, for matrix $\mathbf{R}_0$, we assigned inverse-Wishart prior distribution with 2.002 degrees of belief. For matrix $\mathbf{R}_0$ as the scale matrix, we used the diagonal matrix of order K, with elements equal to (0.002/2.002). In the supplement of Longdon *et al.* (2011), it was pointed out that inverse-Wishart prior distribution defined in this way induces a marginal prior distribution on the variances that is equivalent to an inverse-γ distribution with shape and scale equal to 0.001.

Both SH and ENVIR models were fitted using R-package MCMCglmm (Hadfield, 2012). Each model was run five times with different starting values. The length of each chain was set to 2 000 000 iterations with a burn-in period of 1 000 000 iterations and a thinning interval equal to 50. The convergence of chains was examined by visual inspection of trace plots and using the Gelman and Rubin (1992) convergence diagnostic (see also Cowles and Carlin, 1996; Gelman *et al.*, 2014) implemented in the coda R-package (Plummer *et al.*, 2006). For each model and chain, the deviance information criterion (Spiegelhalter *et al.*, 2002; DIC; in smaller-is-better form; see also Sorensen and Gianola, 2002) was calculated. For clear presentation, only the results for the best model will be reported. For the best model, the estimates from all five chains were summarized using the summary function from coda R-package (Plummer *et al.*, 2006).

Further, based on the obtained posterior mean yields and stability measure associated with the best model, all varieties were ranked. To combine the rankings, we applied Kang's (1988) nonparametric rank-sum stability procedure, where both posterior mean yield and posterior stability measure associated with the best model were used as a selection criterion. As in Kang (1988), the varieties with the lowest rank sum are the most desirable.

Further, for the best model, based on the variance components and the posterior variety means, coefficients of variation (CV) were calculated using the formula:

$$\text{CV}_i = \frac{\hat{\tau}_i}{\hat{\mu}_i} \times 100\%, \quad i = 1, \ldots, I \tag{3}$$

where $\hat{\tau}_i$ is equal to $\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_{ii}^2}$ for the SH model, $\sqrt{\hat{\sigma}_{ii}^2}$ for the ENVIR model, and $\hat{\sigma}_{ii}^2$ is the estimate of the i-th diagonal element of $\mathbf{G}_0$ either in the SH model or the ENVIR model.

Since choosing which of the number of varieties for future sowing is a decision theoretic problem, optimum Bayesian choice of the best variety requires a utility function to be specified. In the present work, the critical-level approach to stability was used and for each variety, the utility function was defined in the following way: when the yield y is below the critical level γ, the utility is zero; above γ, the utility is equal to a positive constant c times y. Given the model parameters and assuming normality, the expected utility for variety i is equal to

$$E(y_i|D, \gamma) = c\left[\mu_i + \tau_i \times \varphi\left(\frac{\gamma - \mu_i}{\tau_i}\right)\right] \tag{4}$$

where τ_i is equal to $\sqrt{\sigma_u^2 + \sigma_{ii}^2}$ for the SH model, and $\sqrt{\sigma_{ii}^2}$ for the ENVIR model, σ_{ii}^2 is the i-th diagonal element of $\mathbf{G}_0$ either in the SH model or the ENVIR model, $\varphi(\cdot)$ is the density function of the standard normal distribution, and D, depending on the model, is a set of parameters of either the SH model or the ENVIR model. Given γ, the best variety is the one maximizing the posterior expected value of this expression. Since the parameters in the preceding equation are usually unknown, one has to use their estimates instead. From our observations, it follows that varieties with mean tuber yields above 42 t/ha are being recommended. For this reason, in the present work, we set the critical level γ equal to 42 t/ha. Moreover, from the discussions with farmers, it follows that the farmers' tuber yield is equal to 80% of the tuber yield obtained in the trials. For this reason, we set in (4) c equal to 0.8.

Next, for a risk analysis, the probabilities that the yields of varieties fall below a certain critical level in the environments were calculated. Since the probability of falling below a certain level depends on the variety mean and variance, it can be treated as a different stability measure which combines mean and variance in a meaningful way (Eskrigde, 1990; Eskridge and Mumm, 1992). We assumed that the tuber yield follows a normal distribution. The probability that the tuber yield of the i-th variety falls below a certain critical level in a randomly chosen environment is equal to (Eskrigde, 1990)

$$p(i) = P(y_{ik} < \delta) = \Phi\left(\frac{\delta - \mu_i}{\sqrt{\sigma_{ii}^2}}\right), \quad i = 1, \ldots, I \tag{5}$$

where μ_i is the mean for the i-th variety, σ_{ii}^2 is the variance of i-th variety across environments, $\Phi(\cdot)$ is the standard normal c.d.f.

https://doi.org/10.1017/S0021859620000945 Published online by Cambridge University Press

Table 5. Posterior variety means, environmental variances for tuber yield, coefficients of variation and posterior expected utilities

Variety	Mean [t/ha]		Environmental variance		Kang's rank-sum	CV [%]		Posterior expected utility
	Estimate	CI	Estimate	CI		Estimate	CI	
Arielle	**52.99 [1]**	(50.10, 55.87)	142.11 [3]	(100.62, 200.09)	**4**	**22.43 [1]**	(18.81, 26.96)	44.85 [2]
Denar	51.50 [3]	(48.25, 54.76)	181.94 [6]	(129.00, 255.69)	9	26.12 [5]	(21.88, 31.46)	44.52 [3]
Everest	52.26 [2]	(48.48, 56.08)	196.76 [7]	(134.39, 284.51)	9	26.75 [6]	(22.08, 32.65)	**45.19 [1]**
Impala	51.39 [4]	(48.23, 54.55)	171.65 [5]	(121.45, 241.66)	9	25.42 [4]	(21.26, 30.61)	44.31 [4]
Lord	48.65 [5]	(45.03, 52.27)	*224.79 [8]*	(159.34, 316.29)	*13*	*30.73 [8]*	*(25.64, 37.20)*	42.17 [5]
Mitek	*41.83 [8]*	(39.29, 44.38)	**110.75 [1]**	(78.05, 155.93)	9	25.08 [3]	(20.96, 30.20)	*36.78 [8]*
Riviera	43.35 [7]	(40.77, 45.93)	112.78 [2]	(79.72, 158.73)	9	24.42 [2]	(20.43, 29.43)	38.00 [7]
Viviana	43.80 [6]	(40.70, 46.91)	166.14 [4]	(117.26,233.75)	10	29.35 [7]	(24.46, 35.49)	39.07 [6]

The numbers in square brackets denote the ranking of variety. Most favourable scores are highlighted in bold while least favourable scores are shown in italics.

Finally, by expressing $p(i_2)$ as a function of $p(i_1)$ (Piepho, 1996)

$$p\,(i_2) = \Phi\left(\frac{\Phi^{-1}[\,p(i_1)]\sqrt{\sigma^2_{i_1 i_1}} + \mu_{i_1} - \mu_{i_2}}{\sqrt{\sigma^2_{i_2 i_2}}}\right), \quad 0 < p(i_1) < 1 \tag{6}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal c.d.f., the relative risks for the best varieties from each country can be plotted. Since the parameters μ_{i_1}, μ_{i_2}, $\sigma^2_{i_1 i_1}$, $\sigma^2_{i_2 i_2}$ in the preceding equations are usually unknown, one has to use their estimates instead.

Results

Firstly, to improve convergence and mixing of the chains, the observed tuber yields were divided by 10. Mean, standard deviation (s.d.), minimum (min) median (med) and maximum (max) of tuber yield expressed in tonnes per hectare (t/ha) for the three studied years are given in Table 3.

From Table 3, it can be seen that tuber yield varied from 8.95 to 85.99 t/ha. Furthermore, the lowest mean tuber yield was observed in 2018, while the highest mean tuber yield was observed in 2016.

The data set was analysed five times using model (2) (Table 4) with different starting values (see Supplements S1 and S2). For each model and chain, the values of DIC were calculated (Table 4).

It can be seen that, for each chain, the smallest value of DIC was obtained for the ENVIR model. This means that the ENVIR model was the best among the two fitted models, and that this model should be preferred. For clarity, the results for the ENVIR model are presented. Before looking at the estimates from the ENVIR model, we first examined the convergence of the random and fixed effects by visually inspecting the trace plots (Supplements S3 and S4). From the trace plots, one can observe that all chains show good convergence. This was confirmed by the Gelman and Rubin tests. The values of point estimates of potential scale reduction factor (Point est.) for all fixed and random effects were equal to one (see Supplement S7). The evolution of the Gelman and Rubin's shrink factor for the parameters of interests is shown in Supplements S5 and S6.

For the series of field trials, the analysis of the ENVIR model provided several estimated parameters and statistics (Table 5). The variance component for replicates nested within environments was equal to 2.728 (Supplement S7). The mean error variance from the series of field trials was equal to 5.804. A detailed inspection of the residual posterior variances revealed that the biggest error variance occurred in environments 201610, 201611, 201623, 201629 and 201710, i.e. in 2016 and sites Rychliki, Krzyżewo, Świebodzin, Zybiszów, and in 2017 and in site Rychliki (see Supplement S5).

Table 5 reports the estimates of posterior variety means, environmental stability variances, Kang's rank sums, coefficients of variations and posterior expected utilities.

The estimated posterior variety means with 95% credible intervals (CI) are reported in columns two and three. The posterior distributions of variety means are shown in Supplement S4. Among the tested varieties, variety Arielle had the highest tuber yield. This variety was also the best among the Dutch varieties tested in the series of field trials. For the Polish varieties, variety

https://doi.org/10.1017/S0021859620000945 Published online by Cambridge University Press

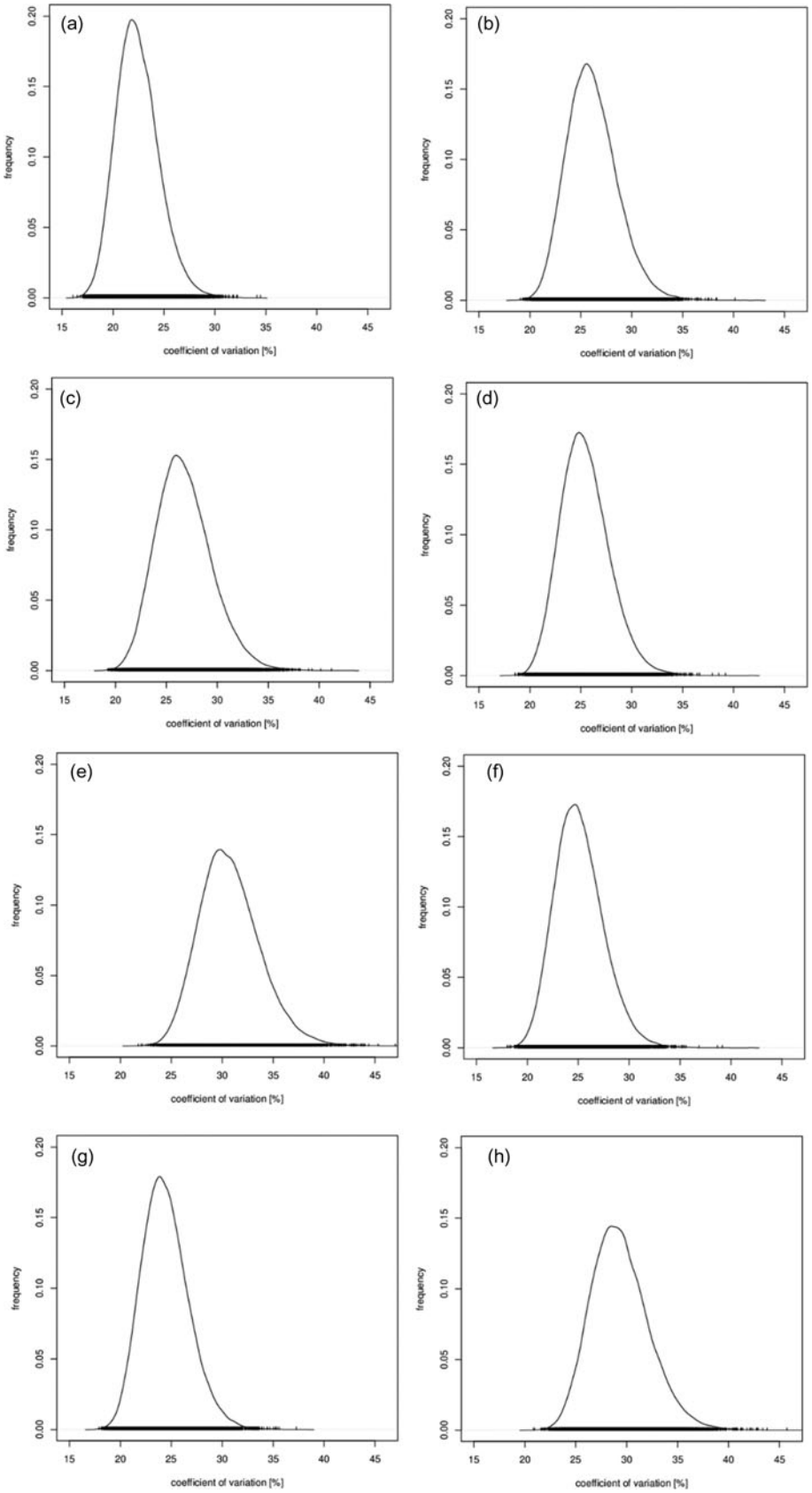


Fig. 1. Posterior distributions of coefficients of variation estimated in the Bayesian environmental variance model for varieties: Arielle (*a*), Denar (*b*), Everest (*c*), Impala (*d*), Lord (*e*), Miłek (*f*), Riviera (*g*) and Viviana (*h*).

https://doi.org/10.1017/S0021859620000945 Published online by Cambridge University Press

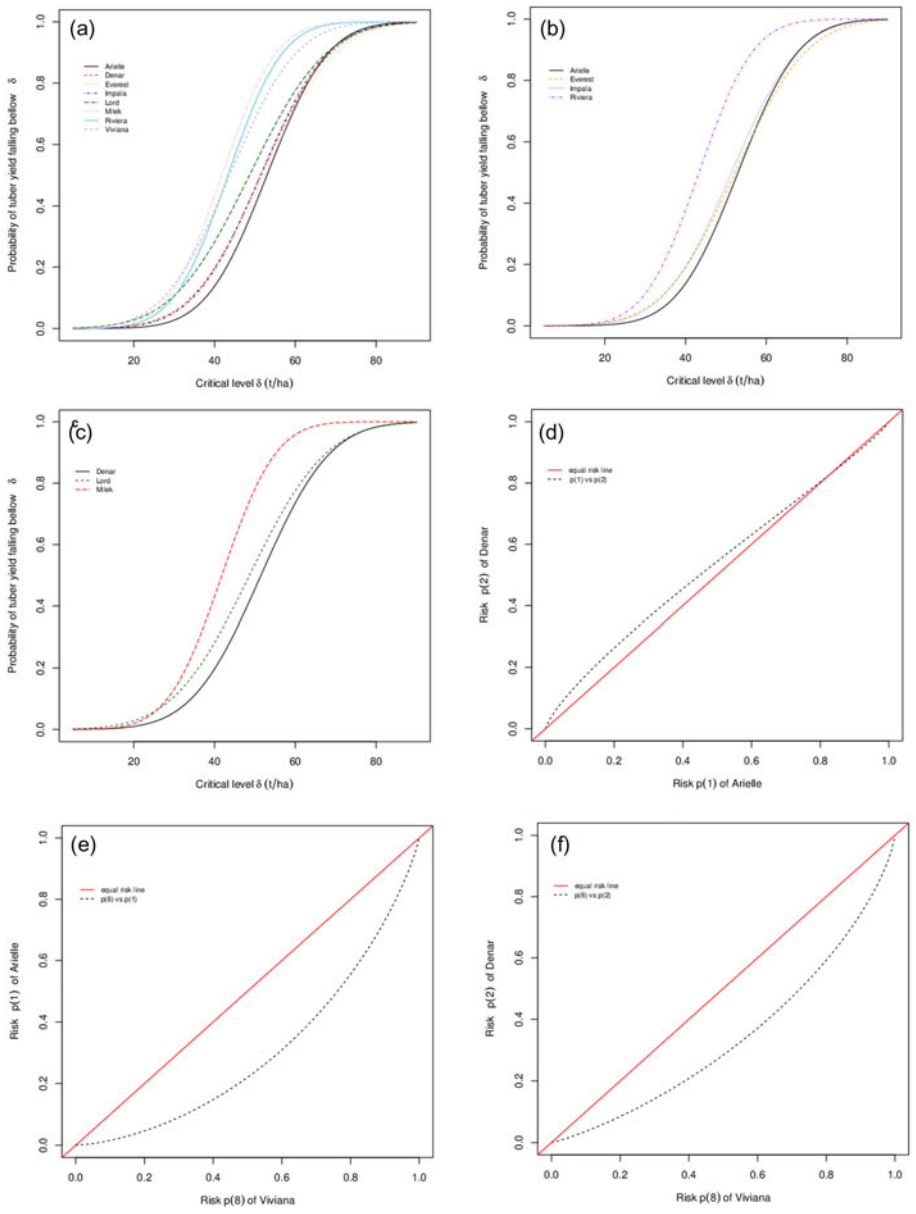


Fig. 2. Colour online. Risk functions for varieties in the series of field trials (top panel) and relative risk functions for the best Dutch, Polish and German varieties (bottom panel).

Denar had the highest tuber yield. The German variety Viviana had one of the lowest tuber yields in the series of field trials.

In columns four and five of Table 5, the estimates of the posterior environmental variances with their 95% CI are reported. One can observe that variety Miłek was the most stable variety, i.e. this variety had the smallest value of the environmental variance. This variety was also the most stable among the Polish varieties. The second best variety in terms of environmental variance was variety Riviera, which was also the most stable variety among the Dutch varieties. The German variety Viviana was the fourth best variety. On the other hand, variety Lord was the most unstable variety in the series of field trials. The posterior distributions of environmental variances are shown in Supplement S3.

The Kang's rank sums are reported in column six of Table 5. It can be noticed that variety Arielle had the smallest rank sum, which was equal to 4 and means that this variety is the most desirable. A different pattern was observed for the Polish varieties where the smallest rank sum was obtained for varieties Denar and Miłek and was equal to 9. The rank sum for variety Viviana was equal to 10.

The estimated posterior coefficients of variation (CV) with their 95% CI are given in columns six and seven of Table 5. One can observe that the smallest value of CV was obtained for variety Arielle and was equal to 22.42%. The second smallest value of CV was obtained for variety Riviera. Among the Polish varieties, the smallest value of CV was obtained for variety Miłek. On the other hand, variety Lord was the worst in terms of CV. Figure 1 shows the posterior distributions of coefficients of variations, from which the convergence of the Gibbs sampler can be inferred.

In the last column of Table 5, the expected posterior utilities are reported. The highest value of the expected utility was obtained for variety Everest and was equal to 45.19. The second best variety in terms of the posterior expected utility was variety Arielle (44.85). On the other hand, the smallest value of the posterior expected utility was obtained for variety Miłek.

https://doi.org/10.1017/S0021859620000945 Published online by Cambridge University Press

Next, the estimates from Table 5 (posterior variety means and environmental variances) were plugged in to (5) and the risks of all varieties as well as for each group of varieties were plotted against the critical level in Figs 2(*a*)–(*c*), respectively.

From Fig. 2(*a*), one cannot characterize any variety as the most stable across all critical levels. It can be seen that variety Arielle up to a critical level around 60 t/ha had the lowest risk. Above this level, variety Everest displayed the lowest risk probability and variety Arielle was more risky (Fig. 2(*b*)). A different pattern can be observed for the Polish varieties (Fig. 2(*c*)). It can be noticed that up to a critical level around 70 t/ha, variety Denar had the lowest risk. Above this level, both varieties Denar and Lord had equal risk. Finally, using (6), in Figs 2(*d*)–(*f*), the relative risk functions for varieties Arielle, Denar and Viviana are plotted. Comparing varieties Arielle and Denar, it can be noticed that up to a risk about 0.8, variety Denar was slightly more risky than variety Arielle, whereas above this level, both varieties Arielle and Denar had equal risk (Fig. 2(*d*)). Now, comparing variety Viviana with varieties Arielle and Denar, one can observe that Viviana was always more risky than Arielle and Denar (Figs 2(*e*) and (*f*)).

Discussion

Usually, the stability of plant traits is assessed in multi-environment trials, which are analysed using either a two-stage or one-stage approach (see e.g. Piepho *et al*., 2012; Caliński *et al*., 2017; Damesa *et al*., 2017; Studnicki *et al*., 2019). In the two-stage approach, one first analyses each trial separately and next, variety means are analysed using AMMI models or linear mixed models. In the one-stage approach, the analysis is performed on plot data. In the present work, we followed a one-stage approach. In a Bayesian framework, a two-stage approach is commonly used in a network meta-analysis.

In the present study, the stability analysis was performed on plot data using a Bayesian linear mixed model. This approach was preferred for several reasons. First of all, the Bayesian approach allows to incorporate the agronomist's knowledge about the likely values of average yields and variance components into the assessment of variety performance in a systematic way by using proper prior distribution. In the present work, as a prior knowledge about the tested varieties, the values of empirical variances of each variety were used. A similar approach was used in Mathew *et al*. (2016), where the authors applied the multi-trait animal model in the analysis of plant breeding trials. However, instead of using empirical variances, one can use the knowledge about varieties from the previous studies. Further, Bayesian analysis offers a possibility of calculating posterior distributions of new quantities, which are functions of model parameters. In the present study, the posterior distribution for coefficients of variation of each variety was calculated. In a similar manner, plant breeders can calculate the posterior distribution of heritability of a given trait. An exemplary code can be found in Villemereuil (2012). Moreover, by considering a multi-trait model and by using the Bayesian approach, plant breeders will be able to calculate the posterior distribution of heritability of each trait as well as the posterior distribution of the genetic correlation between the traits. Finally, under the Bayesian approach, the variety recommendation process can be treated as a formal decision problem.

In the present work, by assuming different covariance matrices for the random vector of environment × variety interaction and modifying the list of random effects in our model, we obtained different models, which were compared in terms of DIC. A similar strategy has been described in Piepho (1999). In that paper, the author also analysed two-way data and his goal was also to find the most stable variety. He assumed different covariance matrices for the environment × variety interaction and modified the list of random effects, this way he obtained several models and using different goodness of fit statistics chose the best one. In the same paper, he pointed out that within unifying mixed modelling framework, the problem of choosing an appropriate stability measure can be regarded as the problem of selecting the most appropriate variance–covariance structure and concluded that the choice of stability measure is data-dependent. In a different study, Studnicki *et al*. (2017) used a similar approach to find a model with the highest predictability for means of cultivar × location and of cultivar × region.

In our Bayesian model, we assumed the heterogeneity of random errors. However, in the literature, many authors in the analyses of two-way or three-way data assume homogeneity of error variances (see e.g. Moore and Dixon, 2015; Caliński *et al*., 2017; Studnicki *et al*., 2019). It should be stressed that the model under the residual error variances homogeneity assumption can be justified from the randomization theory (see e.g. Caliński *et al*., 2017 and the references therein). From the practical point of view, this assumption does not fully reflect the reality and may affect stability analysis. In Hu *et al*. (2013, 2014), it was shown that residual error variances homogeneity assumption may limit the accuracy of genotype evaluations or reliability of varietal recommendations. Moreover, Edwards and Jannink (2006) in their study have found that for all sources of variability in their model, the variance components in homogeneous variance model were numerically larger than the marginal variance estimates in the heterogeneous variance model. The heterogeneity of error variances was assumed by Edwards and Jannink (2006) in Bayesian modelling of multi-environment trials. A similar assumption was used by Smith and Cullis (2018) in construction of their plant breeding selection tools. The same assumption was also used by Malosetti *et al*. (2008) in mixed-model analyses for multi-trait multi-environment data. From the computational point of view, the heterogeneity of error variances assumption can be easily implemented in statistical packages like Genstat, SAS, ASREML or some R-packages like MCMCglmm. An exemplary Genstat code can be found in Malosetti *et al*. (2008).

The two models described in the present work can be easily programmed in MCMCglmm R-package. In this package, the syntax used to specify the model is similar to that by an R interface to ASREML (Gilmour *et al*., 2002; Butler *et al*., 2007), so scientist familiar with ASREML will not have problems with specifying their own model. Moreover, this package allows to define the covariance matrices for random effects in terms of Kronecker product as in Genstat, SAS or ASREML, and which is impossible in programs such as WinBUGS or JAGS. However, in the MCMCglmm package, by default, only a single chain is generated, while in WinBUGS or JAGS, one can use more than one chain to estimate the parameters. To fit the model with different starting values using MCMCglmm package, one has to run the model several times. This may be time consuming, especially for large data sets. To overcome this problem, one can use parallel computing and run the model with different starting values on separate cores.

In the MCMCglmm package, by default, an inverse-Wishart distribution is used as prior for variance components and

https://doi.org/10.1017/S0021859620000945 Published online by Cambridge University Press

covariance matrices. As in Gelman *et al.* (2014), in the present study to estimate the covariance matrix $\mathbf{G}_0$, an inverse-Wishart distribution with dim $\mathbf{G}_0 + 1$ degrees of belief was used. On the other hand, in Lunn *et al.* (2013), it was pointed out that the least informative proper inverse-Wishart is obtained by setting degrees of belief equal to dim $\mathbf{G}_0$. As in Lunn *et al.* (2013), a similar assumption regarding the degrees of belief in the inverse-Wishart distribution was used in Mathew *et al.* (2016). However, in literature concerning Bayesian statistics, it was argued against using inverse-Wishart priors for covariance matrices because they impose a degree of informativity and the posterior inferences are sensitive to the choice of hyper-parameters. Instead, it was proposed to use half-Cauchy distribution or inverse-*gamma* distribution as a prior for variance components (Gelman, 2006) and scaled inverse Wishart distribution (Gelman *et al.*, 2014) as prior for covariance matrices (Gelman *et al.*, 2014). To overcome this problem, one can use the guidelines found in the supplement of Longdon *et al.* (2012) or in Villemereuil (2012). In both of these papers, the authors have shown how to parameterize the inverse-Wishart distribution so that it corresponds to several commonly used prior distributions. Finally, it should be stressed that, at least for now, the list of covariance matrices in the MCMCglmm package, which can be used in modelling the environment × variety, is short. For example, one cannot fit the mixed-model counterparts of Finlay–Wilkinson model and of AMMI model (see Piepho, 1999), since the factor-analytic covariance matrix has not been implemented. However, by combining the available list of covariance matrices with the grouping of random effects as described in Piepho and van Eeuwijk (2002), one could try to extend the methodology described in the present study to the analysis of three-way data. We plan to explore this problem in future work.

In the literature, several Bayesian counterparts of the classical stability measures and models have been described (Crossa *et al.*, 2011; Josse *et al.*, 2014; de Oliveira *et al.*, 2016; Lian and de los Campos, 2016; Bernardo *et al.*, 2018). It would be interesting to compare the results of stability analyses obtained by different Bayesian counterparts of the classical stability measures and how similar the rankings of high-yielding and stable varieties obtained by different methods would be.

In the present work, we have shown that recommendation of varieties for future sowing can be treated as a formal Bayesian decision theoretic problem. For this purpose, we first fitted a Bayesian model, and next using a very simple utility function, we calculated the posterior expected utility of each variety. A similar strategy was described by Theobald *et al.* (2002, 2006). In these papers, the authors used Bayesian decision theory with more complicated utility functions to choose a crop variety and fertilizer dose, and to estimate economically optimum seed rates for winter wheat. For example, using the approach described in Theobald *et al.* (2002, 2006) or described in the present study, agronomists can choose variety and planting date or find the optimal plant density and variety combination, e.g. in soybean or maize.

In the present work, for each variety, the probabilities of falling below a certain level were calculated. These probabilities can be treated as a different stability measure which combines mean and variance in a meaningful way (Eskrigde, 1990; Eskridge and Mumm, 1992). Eskrigde (1990), and Eskridge and Mumm (1992) used this concept to find high-yielding and stable varieties. Recently, Marcholdt *et al.* (2019) used this concept to evaluate the yield performance and stability of wheat depending on the fertilization level and combination. For example, combining the approach presented in this paper with the methodology described in Piepho (1996,1998) could gain additional information for recommending varieties. In agronomy studies, by using the concept of falling below a certain threshold, agronomists can calculate the probability of a new planting data or plant density outperforming the standard one. Moreover, one can calculate the probability of a new agronomic system outperforming the old one. We plan to explore this problem in a future work.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0021859620000945

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflict of interest. None.

Ethical standards. Not applicable.

References

Bernardo Júnior LAY, Silva CP, de Oliveira LA, Nuvunga JJ, Pires LPM, Von Pinho RG and Balestre M (2018) AMMI Bayesian models to study stability and adaptability in maize. *Agronomy Journal* **110**, 1–12.

Butler D, Cullis BR, Gilmour AR and Gogel BJ (2007) *Analysis of Mixed Models for S language Environments: ASRemlR Reference Manual.* Brisbane, Australia: Queensland DPI. Available at http://www.vsni.co.uk/resources/doc/asreml-R.pdf.

Caliński T, Czajka S, Kaczmarek Z, Krajewski P, Pilarczyk W, Siatkowski I and Siatkowski M (2017) On mixed model analysis of multi-environment variety trials: a reconsideration of the one-stage and the two-stage models and analyses. *Statistical Papers* **58**, 433–465.

Cowles MK and Carlin BP (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* **91**, 833–904.

Crossa J, Perez-Elizalde S, Jarquin D, Cotes JM, Viele K, Liu G and Cornelius PL (2011) Bayesian Estimation of additive main effects and multiplicative interaction model. *Crop Science* **51**, 1458–1469.

Damesa TM, Möhring J, Worku M and Piepho HP (2017) One step at a time: stage-wise analysis of series of experiments. *Agronomy Journal* **109**, 845–857.

de Oliveira LA, Silva CP, Nuvunga JJ, Silva AQ and Balestre M (2016) Bayesian GGE biplot models applied to maize multi-environment trials. *Genetics and Molecular Research* **15**, 1–21.

Digby PGN (1979) Modified joint regression analysis for incomplete variety x environment data. *Journal of Agricultural Sciences* **93**, 81–86.

Eberhart SA and Russell WA (1966) Stability parameters for comparing varieties. *Crop Science* **6**, 36–40.

Edwards JW and Jannink JL (2006) Bayesian modeling of heterogeneous error and genotype × environment interaction variances. *Crop Science* **46**, 820–833.

Edwards JW and Orellana M (2015) Increasing selection response by Bayesian modeling of heterogeneous environmental variances. *Crop Science* **55**, 556–563.

Eskridge KM and Mumm RF (1992) Choosing plant cultivars based on the probability of outperforming a check. *Theoretical and Applied Genetics* **84**, 894–900.

Eskrigde KM (1990) Selection of stable cultivars using a safety first rule. *Crop Science* **30**, 369–374.

Finlay K and Wilkinson GN (1963) The analysis of adaptation in a plant breeding programme. *Australian Journal of Agricultural Research* **14**, 742–754.

Gauch HG (1988) Model selection and validation for trials with interaction. *Biometrics* **44**, 705–715.

Gelman A (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–533.

https://doi.org/10.1017/S0021859620000945 Published online by Cambridge University Press

Gelman A and Rubin DB (1992) Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* **7**, 457–511.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A and Rubin DB (2014) *Bayesian Data Analysis*, 3rd Edn. Boca Raton: Chapman & Hall/CRC.

Gilmour AR, Gogel BJ, Cullis BR, Welham SJ and Thompson R (2002) *ASReml User Guide Release 1.0*. Hemel Hempstead, UK: VSN International Ltd. Available at http://www.VSNIntl.com.

Hadfield JD (2012) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software* **33**, 1–22.

Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **80**, 132–138.

Hu X, Yan S and Shen K (2013) Heterogeneity of error variance and its influence on genotype comparison in multi-location trials. *Field Crops Research* **149**, 322–328.

Hu X, Yan S and Li S (2014) The influence of error variance variation on analysis of genotype stability in multi-environment trials. *Field Crops Research* **156**, 84–90.

Josse J, van Eeuwijk F, Piepho HP and Denis JB (2014) Another look at Bayesian analysis of AMMI models for genotype-environment data. *Journal of Agricultural, Biological and Environmental Statistics* **19**, 240–257.

Kang MS (1988) A rank sum method for selecting high-yielding, stable corn genotypes. *Cereal Research Communications* **16**, 113–115.

Lian L and de los Campos G (2016) FW: an R package for Finlay–Wilkinson regression that incorporates genomic/pedigree information and covariance structures between environments. *G3 Genes, Genomes, Genetics* **6**, 586–597.

Longdon B, Hadfield JD, Webster CL, Obbard DJ and Jiggins FM (2011) Host phylogeny determines viral persistence and replication in novel hosts. *PLoS Pathology* **7**, e1002260.

Longdon B, Hadfield JD, Webster CL, Obbard DJ and Jigins FM (2012) Host phylogeny determines viral persistence and replication in novel hosts. *PLoS Pathology* **7**(9), e1002260.

Lunn D, Jackson C, Best N, Thomas A and Spiegelhalter D (2013) *The BUGS Book. A Practical Introduction to Bayesian Analysis*. Boca Raton: Chapman& Hall/CRC.

Macholdt J, Piepho HP and Honermeier B (2019) Mineral NPK and manure fertilization affecting the yield stability of winter wheat: results from long-term field experiment. *European Journal of Agronomy* **102**, 14–22.

Malosetti M, Ribout JM, Vargas M, Crossa J and van Eeuwijk FA (2008) A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress (*Zea mays* L). *Euphytica* **161**, 241–257.

Mathew B, Holand AM, Koistinen P, Léon J and Sillanpää MJ (2016) Reparametrization-based estimation of genetic parameters in multi-trait animal model using integrated nested Laplace approximation. *Theoretical and Applied Genetics* **129**, 215–225.

Moore KJ and Dixon PM (2015) Analysis of combined experiments revisited. *Agronomy Journal* **107**, 763–771.

Orellana M, Edwards JW and Carriquiry A (2014) Heterogeneous variances in multienvironment yield trials for corn hybrids. *Crop Science* **54**, 1048–1056.

Piepho HP (1996) Simplified procedure for comparing the stability of cropping systems. *Biometrics* **52**, 315–320.

Piepho HP (1998) Methods for comparing the yield stability of cropping systems. *Journal of Agronomy and Crop Science* **180**, 193–213.

Piepho HP (1999) Stability analysis using the SAS system. *Agronomy Journal* **91**, 154–160.

Piepho HP and van Eeuwijk FA (2002) Stability analysis in crop performance evaluation. In Kang M (ed). *Crop Improvements: Challenges in the Twenty-First Century*. Binghampton, New York: Food Products Press, pp. 315–351.

Piepho HP, Möhring J, Schulz-Streeck T and Ogutu JO (2012) A stage-wise approach for analysis of multi-environment trials. *Biometrical Journal* **54**, 844–860.

Plummer M, Best N, Cowles K and Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. *RNews* **6**, 7–11.

Searle SR, Casella G and McCulloch CE (2006) *Variance Components*. Hoboken: Wiley.

Shukla GK (1972) Some statistical aspects of partitioning genotype-environment components of variability. *Heredity* **29**, 237–245.

Smith AB and Cullis BR (2018) Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. *Euphytica* **214**, 143.

Sorensen D and Gianola D (2002) *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics. Statistics for Biology and Health*. New York: Springer-Verlag.

Speed TP, Williams ER and Patterson HD (1985) A note on the analysis of resolvable block designs. *Journal of the Royal Statistical Society Series B* **47**, 357–361.

Spiegelhalter DJ, Best NG, Carlin BP and van der Linde A (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B* **64**, 583–689.

Studnicki M, Paderewski J, Piepho HP and Wójcik-Gront E (2017) Prediction accuracy and consistency in cultivar ranking for factor-analytic linear mixed models for winter wheat multienvironmental trials. *Crop Science* **57**, 2506–2516.

Studnicki M, Derejko A, Wójcik-Gont E and Kosma M (2019) Adaptation pattern of winter wheat in agro-ecological regions. *Scientia Agricola* **72**, 148–156.

Theobald CM and Talbot M (2002) The Bayesian choice of crop variety and fertilizer dose. *Journal of the Royal Statistical Society Series C – Applied Statistics* **51**, 23–36.

Theobald CM, Talbot M and Nabugoomu F (2002) Bayesian approach to regional and local-area prediction from crop variety trials. *Journal of Agricultural, Biological and Environmental Statistics* **7**, 403–419.

Theobald CM, Roberts AMI, Talbot M and Spink JM (2006) Estimation of economically optimum seed rates for winter wheat from series of trials. *Journal of Agricultural Science Cambridge* **144**, 303–316.

Villemereuil P (2012) Estimation of biological trait heritability using the animal model: how to use MCMCglmm R package Guide. https://devillemereuil.legtux.org/wpcontent/uploads/2012/tuto_eng.pdf.

Yan W and Kang MS (2003) *GGE Biplot Analysis: A Graphical Tool for Breeders, Genetists and Agronomists*. Boca Raton: CRC Press.

https://doi.org/10.1017/S0021859620000945 Published online by Cambridge University Press