Statistical methods for analysis of multienvironment trials in plant breeding: accuracy and precision

Dissertation to obtain the doctoral degree of Agricultural Sciences (Dr. sc. agr.)

Faculty of Agricultural Sciences

University of Hohenheim

Biostatistics Unit (340c), Institute of Crop Science

submitted by

Harimurti Buntaran

from *Pekalongan, Indonesia*

2021

Additional Statement

Some parts of this thesis were parts of the Licentiate thesis:

Buntaran, H. 2019. Assessment of statistical analysis of Swedish cultivar testing: a cross-validation study for model selection. Licentiate of Philosophy, Swedish University of Agricultural Sciences.

The licentiate is an integral part of this Doctoral study.

Die vorliegende Arbeit wurde am 09.03.2021 von der Fakultät Agrarwissenschaften der Universität Hohenheim als „Dissertation zur Erlangung des Grades eines Doktors der Agrarwissenschaften“ angenommen

Tag der mündlichen Prüfung: 27.07.2021

Leiter der Prüfung: Prof. Dr. Uwe Ludewig
Berichterstatter 1. Prüfer: Prof. Dr. Hans-Peter Piepho
Mitberichterstatter 2. Prüfer: Prof. Dr. Fred van Eeuwijk
Mitberichterstatter 3. Prüfer: Prof. Dr. Tobias Würschum

Table of contents

Nomenclature

Acronyms / Abbreviations

BLUE Best linear unbiased estimator

BLUP Best linear unbiased predictor

CS Compound symmetry

CV Cross-validation

EBLUE Empirical best linear unbiased estimator

EBLUP Empirical best linear unbiased predictor

FA Factor-analytic

GEI Genotype × environment interaction

GLSE generalised least squares estimation

LMM Linear mixed models

LOO Leave-one-out

MAR Missing at random

MET Multienvironment trials

MNAR Missing-not-at-random

MSE Mean squared error

MSEP Mean squared error of prediction

OLS Ordinary least squares

RC Random coefficient

REML Residual maximum likelihood

SEPD Standard errors of the predictions of pairwise differences of cultivar values

SEPV Standard errors of predictions of cultivar values

TPE Target population of environments

TPG Target population of genotypes

UN Unstructured

Chapter 1

General Introduction – Multienvironment trials: An epistemological framework for dissecting genotype × environment interaction

1.1 Introduction

The ultimate goal in a plant breeding programme is selecting cultivars that guarantee high yield and quality in varying environmental conditions because different cultivars perform differently in diverse environments, a phenomenon known as genotype×environment interaction (GEI) (Kang and Gorman, 1989). Therefore, multienvironment trials (MET) are carried out to assess cultivars' performance across diverse environmental conditions and thus provide cultivars' performance information via statistical analyses. In the MET, a large number of varieties are tested in several geographical regions. A reliable and robust statistical method is required to provide accurate predictions of the yield of tested cultivars. The MET results can help breeders select the best cultivars and recommend growers to select well-adapted cultivar to their regional conditions.

1.2 Plant breeding: a long history yet still has a long wish list

Plant breeding is the art and science of the genetic improvement of plants for the benefit of humankind, and it has been an integral part of agriculture since humans first selected one type of plant or seed in preference to another, instead of randomly taking what nature provided (Fehr, 1987; Sleper and Poehlman, 2006). Plant breeding is an art because it needs the breeder's skill in observing plants with favourable characteristics, i.e., economic value, environmental adaptation, nutritional, or aesthetic. Breeders depended solely on their eyes and intuition as their skills to judge or select novel plants at that time because the scientific knowledge,

e.g., genetics and statistics, was not as advanced as today. Thus, selection became the earliest form of plant breeding (Sleper and Poehlman, 2006).

Plant breeding then developed into science as knowledge progressed in genetics, statistics, and molecular biology (Sleper and Poehlman, 2006). Recently, the vast and rapid development of molecular genetics and computing power allows for more advanced breeding methods and accelerates cultivar developments and more accurate predictions of cultivar performance.

Increased yield has been the ultimate aim in many plant breeding programmes. Furthermore, one of the most important contributions of plant breeding has been developing better varieties for new agricultural areas (Allard, 1960). High crop productivity in a particular location year-in, year-out (Gepts and Pfeiffer, 2018), is still a high priority in breeding programmes and as a basis for cultivar recommendation. However, due to unpredictable climate change, breeders face a greater challenge to developing stable and high-yielding cultivars. Cultivars, which are drought-resistant, pathogens-pest-resistant, have higher biomass, are just a few to name from a long wish list of plant breeders.

1.3 Genotype × environment interaction (GEI)

Genotype × environment interaction (GEI) is differential genotypic expression across environments (Romagosa and Fox, 1993). In the GEI concept, the “genotype” term is interchangeably with “variety”, “crop”, and “cultivar” (Buntaran, 2019). The existence of GEI inhibits genetic analysis of performance reduces the efficiency of crop improvement in a plant breeding programme, mainly due to the confounding between tested genotypes comparison with the environment and complicate the breeding objectives (Cooper and Byth, 1996). The GEI issue has been discussed more than a half-century ago by Allard (1960) describing the biological complexity underlying GEI: “virtually all phenotypic effects are not related to the gene in any simple way. Rather they result from a chain of physico-chemical reactions and interactions initiated by genes but leading through complex chains of events, controlled or modified by other genes and the external environment, to the final phenotype”. Furthermore, Allard and Bradshaw (1964) also enunciated the complexity of GEI for plant breeders: "There is rather general agreement amongst plant breeders that interactions between genotype and environment have an important bearing on the breeding of better varieties. However, it is much more difficult to find agreement as to what we ought to know about genotype-environment interactions and what we should do about them". Thus, GEI complicates the selection of the best variety

because a cultivar can outperform in a one or some environments but underperform in others.

The issue of GEI received considerable attention in 1990 as the international symposium on 'Genotype-by-Environment Interaction and Plant Breeding' was held on 12 and 13 February at the Louisiana State University campus in Baton Rouge (Kang, 1990). Since that time, various GEI issues have come to the forefront in many breeding programmes throughout the world (Kang, 2020). The Crop Science Society of America even organised a symposium on the GEI issue and published papers in Crop Science volume 56. In this special issue from Crop Science, de Leon et al. (2016) defined GEI as the differential sensitivity of certain genotypes to different environments. Furthermore, van Eeuwijk et al. (2016) referred the GEI problem as the building of predictive models for genotype-specific reaction norms. The GEI concept can be illustrated as the slope of the line when genotype performance is plotted against an environmental gradient, which is also known as the reaction norm: the genotype-specific functional relationship between phenotype and environmental gradients (DeWitt and Scheiner, 2004; van Eeuwijk et al., 2016) as shown in Figure 1.1.

In Figure 1.1, five scenarios of reaction norm are shown. Figure 1.1a shows no GEI and no plasticity since there is no different mean of genotype performance across the environments, and the ranking of genotypes are the same across environments. Figure 1.1b also shows no GEI but plasticity because of the phenotype expression, in this case, yield, changes across the different environments. In Figure 1.1b, there is no GEI because the genotype and the environment behave additively, and the reaction norms are parallel (no difference ranking and changing mean differences among genotypes). The remaining plots show various situations in which GEI occurs: divergence (Figure 1.1c), convergence (Figure 1.1d), and the most crucial one, crossover interaction (Figure 1.1e). In divergence and convergence situation, the genotype ranking does not change across environments, but the mean difference between the three genotypes does. In the case of crossover interaction, the mean difference and the ranking between genotypes are shifted. In crop breeding, the crossover interaction is more important and problematic than the non-crossover interaction (Baker, 1990). McKeand et al. (1990) emphasized that in the crossover situation, breeders are faced with developing separate populations for each site type where genotypic rankings drastically change. Singh et al. (1999) pointed out that it is important to assess the frequency of crossover interactions because this pattern has substantial implications for specific-adaptation breeding.

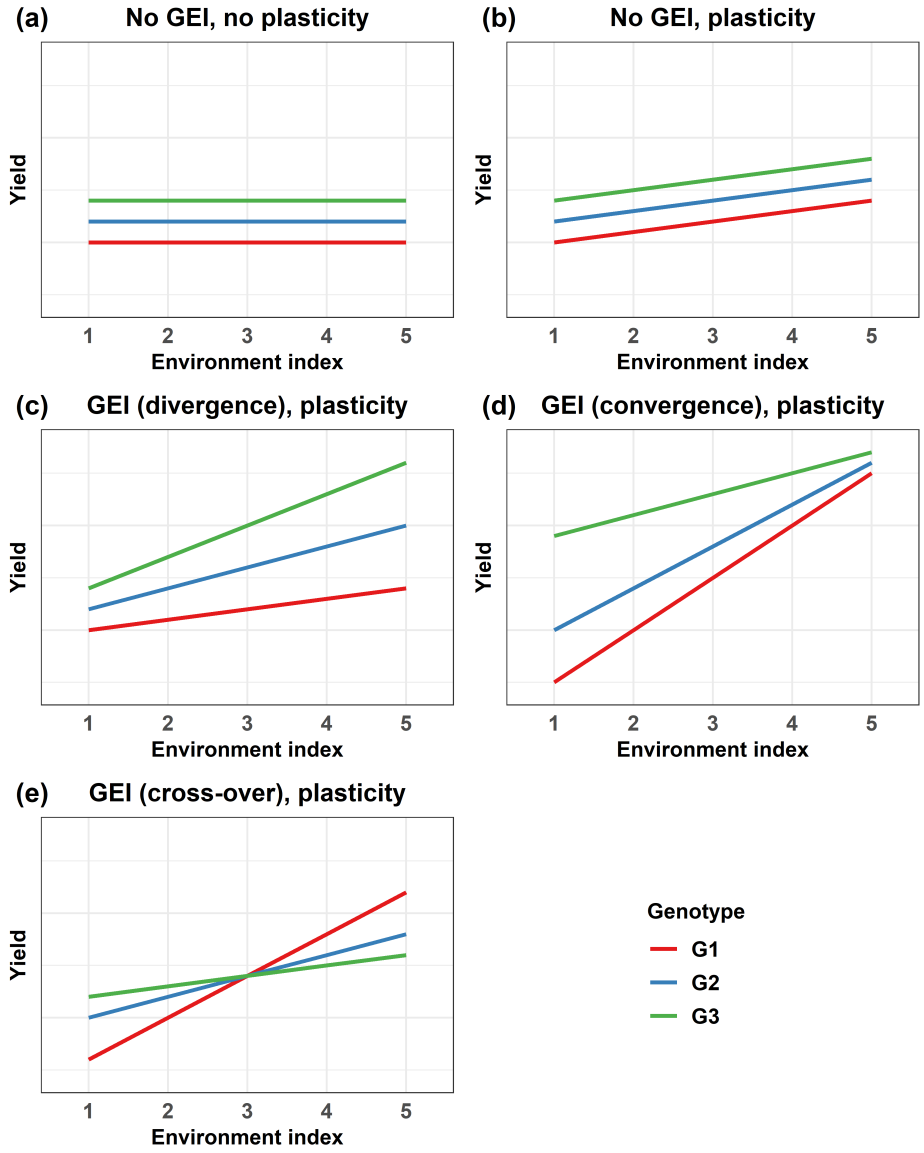

Figure 1.1 Illustration of GEI for three genotypes in five different environment conditions. No GEI in (a) and (b) versus GEI in (b) until (c). No plasticity in (a) versus plasticity in (b) until (e). The environment index shows the unfavourable environment conditions (1) to favourable environment conditions (5).

1.4 Multienvironment trials (MET)

The concepts of the target population of genotypes (TPG) and the target population of environments (TPE) are salient to understand the breeding concepts associated with GEI. The TPG and TPE help breeders define the set of genotypes and environments to obtain valid and precise inference and predictions (van Eeuwijk

et al., 2016). The TPG comprises all candidate genotypes to grow in the coming years (van Eeuwijk et al., 2016). The TPE contains a group of environments concerning the genotypic performance where new cultivars will be adopted. In other words, TPE describes the future growing conditions of the cultivars in the TPG (Comstock, 1977; Cooper and Byth, 1996; Cooper et al., 2014). The TPE is also indispensable to predict GEI since the identification of repeatable GEI is a challenge due to the unpredictable weather (de Leon et al., 2016).

In a breeding programme, developed or improved varieties are assessed in multiple environments, which constitute the potential representatives of the TPE (Cooper and DeLacy, 1994; DeLacy et al., 1996). The term "environment" may refer to a year-location combination. A multienvironment trial's (MET) objective is to determine which varieties are matched to a TPE, based on the reaction norm/expression of the varieties per se to the environments. Thus, METs aid breeders to determine the similarity of environments and grouping similar environments in MET. The information derived from MET is crucial not only for selection purposes in a breeding programme but also to provide advice or recommendation to growers in deciding which cultivar is the most suitable and performs the best in their growing conditions. Thus, robust statistical methods are necessary to obtain accurate predictions of genotype performance such as yield and obtain a reliable stability measure of each cultivar across environments.

1.5 Statistical modelling for GEI analysis

Statistical modelling is a pivotal basis for dissecting GEI. Medina (1992) stated, "Genotype-environment interaction is of major importance to the plant breeder in developing improved varieties because the relative rankings of varieties grown over a series of environments may differ statistically, causing problems in plant selection." Meredith Jr. (1984) mentioned: "Genotype × environment interactions are important to geneticists and breeders because the magnitude of the interaction component provides information concerning the likely area of adaption of a given cultivar. The relative magnitudes of the interaction, error, and genotypic components are useful in determining efficient methods of using time and resources in a breeding program." The following subsections introduce the statistical modelling in the context of GEI covered in this thesis.

1.5.1 The aspects and goal of statistical modelling

Box et al. (2005, p. 208) stated that "*all models are wrong and some are useful*, it is equally true that *no model is universally useful*". How useful? The most that can be expected from any model is that it can provide a useful approximation to reality (Box et al., 2005, p. 440). There are three essential aspects of modelling that can be taken from Box (1976) article titled "Science and Statistics," i.e., flexibility, parsimony, and worrying selectively. Flexibility means one should not fall in love with his/her model because there will always be discrepancies between theory and practice.

Moreover, there is no universal model for all kind of data, and data *per se* is very dynamic. Thus, in modelling, we need to have flexibility and courage to seek out, recognise, and exploit such discrepancies (Box, 1976). The next aspect is parsimony, which in the same spirit as William of Occam means that one should seek an economical description of natural phenomena. A model should not result in an overestimation or underestimation. Thus, in modelling, one should not aim for the most complex model to obtain the estimate values but the parsimony to obtain sufficient estimate values. The last aspect is worrying selectively. This aspect refers to the assertion that since all models are wrong, one should be alert to what is importantly wrong (Box, 1976).

The ultimate goal of statistical modelling is to obtain accurate and precise estimates. Accuracy measures how close an estimate $\hat{\theta}$ of a parameter θ is to the "true value" (Kotz et al., 2006) By contrast, the precision of an estimator $\hat{\theta}$, measures how narrow the distribution of $\hat{\theta}$ clusters about its expected value (Kotz et al., 2006). The precision of $\hat{\theta}$ is the reciprocal of the variance of $\hat{\theta}$. The model selection can be done by evaluating the accuracy of the produced estimates. Accuracy can be evaluated via a cross-validation (CV) study by estimating prediction error (Hastie et al., 2009). From a CV study, accuracy can be measured in terms of mean squared error (MSE), which consists of variance and squared bias of $\hat{\theta}$.

1.5.2 Linear mixed models (LMM): BLUE and BLUP in one model

A model is defined as a mathematical notation of the processes that give rise to the observations in a set of data (Stroup, 2012). A purely mathematical model is a deterministic device in that for a given set of inputs, it predicts the output with absolute certainty, and it leaves nothing to stochastic part (Schabenberger and Pierce, 2001). A model is considered as a statistical model when it includes a deterministic/systematic part and a stochastic/random part. Therefore, a statistical model describes the presumed impact of explanatory variables and the probability distributions associated with aspects of the process that are assumed to be characterised

by random variation (Stroup, 2012). In short, a statistical model comprises three components, i.e., systematic part, which consists of quantitative and/or qualitative explanatory variables, a random part (refers to residual error term), and an assumed distribution.

A linear model (LM) usually refers to a classical linear model with Gaussian error. In matrix notation, this LM is written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{1.1}$$

where $\mathbf{Y}$ is an $n \times 1$ vector of observations, $\mathbf{X}$ is an $n \times k$ incidence matrix for fixed effects factors, $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown fixed effect parameters to estimate, and $\mathbf{e}$ is a vector of residual errors and is assumed homoscedastic, uncorrelated, and following $N(0, \sigma^2\mathbf{I})$. In this case, the parameter estimates of $\boldsymbol{\beta}$ are solved using ordinary least squares (OLS), and the solutions are called best linear unbiased estimators (BLUE). Thus, in the classical LM, there is only one type of effect in the systematic part that is considered, i.e., fixed effect. The matrix structure of variance for the classical LM is $\mathbf{V} = \sigma^2\mathbf{I}$.

Linear mixed models (LMM) extend the classical LM to allow both fixed and random effects factors in one model (Eisenhart, 1947; Harville, 1976; Laird and Ware, 1982). A matrix formulation of LMM is as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{1.2}$$

where $\mathbf{Y}$ is a vector $n \times 1$ of observations, $\mathbf{X}$ is the incidence matrix for fixed effects with $n \times k$ matrix, $\boldsymbol{\beta}$ is a vector of unknown fixed effect parameters to estimate with $k \times 1$ matrix, $\mathbf{Z}$ is the incidence matrix for random effects with $n \times p$ matrix, $\mathbf{u}$ is a vector of unknown random effect parameters to estimate with $p \times 1$. Since $\mathbf{u}$ consists of random effect parameters, $\mathbf{u}$ is assumed to be $N(0, \mathbf{G})$ where $\mathbf{G}$ is the variance-covariance matrix of all random effects. The vector $\mathbf{e}$ consists of residual errors. The assumption of residual errors are more relaxed in the linear mixed models than in the classical linear models since it allows non-independence and heterogeneity, $N(0, \mathbf{R})$, where $\mathbf{R}$ is the variance-covariance matrix for the residuals.

Henderson (1950, 1963, 1975, 1984) developed mixed model equations (MME) to obtain the solutions of fixed effects and random effects for animal breeding purposes. The Henderson's MME is as follows:

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}. \tag{1.3}$$

The solutions to MME are the BLUE for $\boldsymbol{\beta}$ and the best linear unbiased predictors (BLUP) for $\mathbf{u}$. Unlike the classical LM, LMM has variances for random effects and the residual terms. Thus, the covariance matrix of $\mathbf{y}$ in the LMM is written as $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$. In practice, the actual variance components of $\mathbf{G}$ and $\mathbf{R}$ in Equation 1.3 are not known and therefore are estimated from the data. Thus, the appropriate acronyms become EBLUE and EBLUP, where the E refers to empirical.

A fixed effect is **estimated** with BLUE. The term "Best" means that the sampling variance is minimised. "Linear" indicates that the **estimates** are linear functions of the observed values. "Unbiased" implies that the expected values of the **estimates** are equal to their true values $E[BLUE(\boldsymbol{\beta})] = \boldsymbol{\beta}$ of a factor is considered as "fixed" if we are just interested in its particular value or, in general, if a factor only has a few levels and not coming from or representing from a probability distribution (McCulloch et al., 2008), and the conclusions apply only to the particular factor levels (Lynch and Walsh, 1998). For example, the effects of different soil types or effects of different fertilizers. The estimation of fixed effects in the LMM is slightly different from the estimation in the classical linear model. In the LMM, the fixed effects estimates are solved using generalised least squares estimation (GLSE), not OLS.

A random effect is **predicted** with BLUP. The expansion of "B" is the same as "B" in BLUE. "L" indicates the **predictions** are linear functions of the observed values. "Unbiased" implies that the expected values of the **predictions** are equal to their true values $E[BLUP(\mathbf{u})] = E(\mathbf{u}) = 0$. The term "prediction" is chosen by (Henderson, 1984) since in animal breeding, the interest is to evaluate the potential of breeding value of a mating between two potential parents and to predict the future records. The term "estimation" is more appropriate to estimate the value if an animal already born. Thus, it has become a common term in practice to "estimate" fixed effects and to "predict" random effects (Robinson, 1991).

In the opposite of the fixed effect, one of the assumptions underlying random effects is that when the experiment is repeated, the true value will change/not be constant. This is also the reason that the term "prediction" is used for a random effect. The other assumptions are: the levels of a factor are of no particular interest and represent or come from a probability distribution. Thus, in general, the random effect factor levels will have many levels to represent the whole population. While in the fixed effect, the parameter to estimate is the mean of individual levels, in the random effect, a variance (dispersion parameter) is the parameter to estimate. Therefore, the inference applies to a population (Searle et al., 1992, p. 17).

1.5.3 LMM in the context of MET

In plant breeding, an individual location or a trial is considered random, but the set of locations/trials is considered as a fixed effect (Bernardo, 1996). In other words, if a zone or region consists of several locations, then the effect of a zone is considered as fixed, and the effect of locations are considered as random because the condition of locations may change from year to year. Furthermore, the mean differences among different sets of environments are considered as nuisance factors that should be taken into account for genotypes comparisons (Bernardo, 2020). In the fixed effect, when the experiment is repeated, the effect of a factor will be the same, which means the true value of the fixed effect does not change in each repetition of the experiment (Blasco, 2017).

The model for an MET laid out as a generalized lattice design, e.g., an alpha-lattice design can be expressed as:

$$y_{ijkl} = \mu + r_{jk} + b_{jkl} + g_i + e_j + (ge)_{ij} + \epsilon_{ijkl} \tag{1.4}$$

where μ is the intercept, r_{jk} is the effect of the k-th replicate in the j-th environment, b_{jkl} is the effect of the l-th incomplete block nested in the j-th replicate in the j-th environment, g_i and e_j are the main effects of genotype and environment, $(ge)_{ij}$ is the interaction effect of the i-th genotype and j-th environment, ϵ_{ijkl} is the residual term.

In connection to Equation 1.2, the intercept is classified in the fixed-effects part, $\mathbf{X}\boldsymbol{\beta}$, while r_{jk} and b_{jkl} are classified in the random-effects part, $\mathbf{Zu}$. The ϵ_{ijkl} is classified in the residual part, $\mathbf{e}$, of Equation 1.2. The effect of $(ge)_{ij}$ can be assumed as fixed or random depends on the objective of the analysis. When predictions of genotypes for broad environment is preferred, then the predictions can be obtained using genotype means $\mu + g_i$, where the effect g_i is assumed to be random, $N(0, \mathbf{G}_g)$. Consequently, in this case, the effect of $(ge)_{ij}$ is also random, $N(0, \mathbf{G}_{ge})$. When predictions of genotype for environment-specific is preferred, then the predictions can be obtained using genotype-environment specific means $\mu + g_i + e_j + (ge)_{ij}$. Thus, for this scenario, the g_i and $(ge)_{ij}$ are classified in the random-effects part, $\mathbf{Zu}$, and the e_j is classified in the fixed-effects part, $\mathbf{X}\boldsymbol{\beta}$, of Equation 1.2.

1.5.4 Variance-covariance structures in LMM

In the MET analysis, the assumption of homogeneity variance is hardly ever fulfilled because genotypic variances tend to change across environments. Furthermore, genotypic correlations for pairs of these environments are not homogeneous

(Bustos-Korts et al., 2016). In the LMM framework, applying variance-covariance structures can be applied to the random effect of GEI and the residual terms to take into account this heterogeneity and achieve higher prediction accuracy. The variance-covariance for GEI effect is modelled in the $\mathbf{G}_{ge}$ matrix, and the variance-covariance of the residual term is modelled in the $\mathbf{R}$ matrix.

Some variance-covariance structures used to exploit the heterogeneity of the GEI term are compound symmetry (CS), unstructured (UN), and factor-analytic (FA). The CS structure implies that both variance and covariance are homogeneous. Thus, each environment has the same variance, and the genetic correlation is the same between all pairs of environments. The UN structure allows both heterogeneous covariance and variance. Thus, each environment has a unique genotype variance, and each pair of environments has a unique covariance. The number of parameters needed for this structure is $p(p+1)/2$, where p is the number of environments. The FA structures are often more useful than the UN structure for taking into account heterogeneity in complex genotype × environment models. These structures have fewer parameters than the UN structure (Isik et al., 2017). We here describe the FA structure with a single multiplicative term (FA1). In this structure, the $\mathbf{G}_{ge}$ is defined as $\mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$, where $\mathbf{\Lambda}$ is a vector of dimension $1 \times p$ that consists of loading factors λ_1 to λ_p and $\mathbf{\Psi}$ is a $p \times p$ diagonal matrix that consists of environment-specific genotype variances (ψ_j^2), $j = 1,2,...,p$. The residual variance structure can be modelled as a diagonal heterogeneous-environment specific, σ_j^2.

1.5.5 BLUP or BLUE for the cultivar effect

Bernardo (2020) mentioned several benefits of BLUP in the plant breeding framework. First, in the MET, the better genotypes will be tested in several years while the less superior genotypes will be discarded, which results in unbalanced data. BLUP allows analysing such unbalanced data while accounting for differences in the amount of data available for each genotype. Second, BLUP uses the information for all relatives measured to improve the prediction of breeding values. For example, when a breeder wants to compare two individuals, A and B, the comparison can be made solely on the basis performance of A and B alone. Using BLUP, the comparison will be more precise by including the information of relatives of A and relatives of B. In the MET, this feature is beneficial that using BLUP, we can borrow or recovery information of the same genotype in other environments, and so exploit the genetic correlation between environments (Atlin et al., 2000; Kleinknecht et al., 2013; Piepho et al., 2016), which improves the prediction accuracy of genotype performance compared to BLUE. In fact, DeLacy et al. (1996) asseverated that the advantage of using

BLUPs for prediction is that the predicted range is close to the 'actual' range, i.e., the range of cultivar performance in the target environments.

The motivation to use BLUP has been formulated by Smith et al. (2001) who asserted a "deficiency in the traditional fixed cultivar-effects approach in terms of obtaining reliable predictions of future yield performance." This deficiency has been discussed by Patterson and Silvey (1980), who stated that "differences between trials means for newly recommended cultivars are, on the average, about 27% too large." Thus, in practice, the estimation of cultivar's yield may be too optimistic, and the ranking of cultivars may be not accurate since the cultivar effect is fixed. In Chapter 2, the performance of EBLUE and EBLUP for zone-based prediction, including the complex variance-covariance structures, was assessed.

1.5.6 Stagewise analysis

The MET data can be analysed by a single-stage analysis or stage-wise analysis (two stages or more). A single-stage analysis is considered as the gold standard (Gogel et al., 2018). A single-stage analysis has an advantage from theoretical consideration since the estimation of fixed and random effects are done in a single model from plot-level data (Piepho et al., 2012a). Nevertheless, the most common disadvantage is the computational burden, especially when the numbers of cultivars and environments are large and a complex variance-covariance structure for the cultivar×environment interaction effects is assumed (Möhring and Piepho, 2009; Welham et al., 2010).

The computational burden in the single-stage analysis motivates a stage-wise analysis that splits the analysis into two (or more) stages. Damesa et al. (2017) and Piepho et al. (2012a) reported that the stage-wise analysis was able to reduce the computational burden substantially. In the stagewise analysis, each trial is analysed separately using BLUE, in the first stage, to obtain adjusted cultivar means per trial. Thus, the cultivar effects are modelled as fixed. In the second stage, the adjusted cultivar means from the first stage are analysed jointly, using an appropriate mixed model to compute marginal means for cultivars across environments. In this stage, the cultivar effects may be modelled as fixed or random. Piepho and Eckl (2014) mentioned another advantage of stage-wise analyses for practical analyses: it facilitates a combined analysis of different trials with different experimental designs in the first stage and subsequently allows modelling structures for heterogeneity of variance between trials easily.

The two-stage analysis had similar results to the single-stage analysis when the fully efficient (FE) method was used, where the full variance-covariance matrix

of the estimated cultivar means from the first stage is forwarded to the second stage reported by Damesa et al. (2017). The results are not equivalent because the variance parameter values used between the single-stage and FE are different due to the fact that residual maximum likelihood estimates (REML) will differ slightly between the two analyses (Damesa et al., 2017). They are mathematically equivalent only if identical variance parameter values are used. The primary issue of stagewise analysis is that storing full variance-covariance is hard to do, so a diagonal approximation is often used. Moreover, the most critical part is choosing the method to forward the information on precision (standard errors, variance-covariance matrix of the adjusted means) between stages to account for heteroscedasticity as well as for covariances among the adjusted means (Damesa et al., 2017; Möhring and Piepho, 2009).

Möhring and Piepho (2009) showed, via simulation, that weighting can improve efficiency, but the unweighted method was acceptable if the assumptions of the model were correct, i.e., when error variances are independent of the genotype × environment interaction structure. They also mentioned that the weighting method's performance did not depend on the evaluation criterion but the dataset. Welham et al. (2010) conducted a simulation study and showed that the two-stage unweighted method performed poorly due to the loss of information in estimating the estimates of cultivar performance, both overall and within environments. Similar to Gogel et al. (2018), Welham et al. (2010) focused on prediction for individual sites.

Moreover, Gogel et al. (2018) advocated a move away from two-stage analysis asserting that the computing power needed to analyze large and complex MET datasets is already available. Their study of wheat MET data confirmed the equivalence of a two-stage factor-analytic (FA) analysis with a known variance-covariance matrix from Stage 1 to a single-stage analysis. An essential distinction between the studies of Damesa et al. (2017) and Gogel et al. (2018) is that Damesa et al. (2017) focus on predicting means across zones, whereas the study of Gogel et al. (2018) focused on predictions for individual locations. Chapter 3 determines the best statistical analysis strategy for zone-based prediction cultivar testing combined with the complex variance-covariance structure to exploit the heterogeneity of cultivar × zone interaction.

1.5.7 Random coefficient models for predictions of the untested site

While MET are usually designed to cover the whole TPE, none of the trials in an MET coincides exactly with a grower's field or location. Thus, grower's fields, the real target of breeding, must be seen as new locations in the TPE. In the same vein, it

may be said that the MET analysis is usually used to produce predictions of tested genotypes for a new location, making use of the information from tested locations. Predicting genotype performance in a new location is akin to predicting values that have no records at all, as reported by Henderson (1977), who showed that BLUP could be used to predict breeding values for the animals that had no records. This approach is in the same spirit to obtain genotype predictions in a new location with no records.

Reporting the precision or precision measures quantified by standard errors and prediction intervals is highly desirable due to growers' fields hardly ever coinciding with the trials' location. Furthermore, in practice, cultivar yield will never reach the same value as the predicted mean values from the MET. When there are no precision measures reported, growers are left with having no information regarding the precision in the predictions that are reported. The critical challenge is that the standard errors of predictions of cultivar means obtained from the routine analysis of MET are only valid for the locations where the trials were carried out, but not for the untested locations or growers' fields. However, the precision of the predictions for the untested locations is crucial to assist growers in selecting a cultivar for their farm or field.

Incorporating environmental covariates can be worthwhile to improve the precision measures of the predicted mean values of the cultivars. Heslot et al. (2014) reported that the environmental covariates that are responsible for GEI are useful to enhance the predictive capability of MET analyses and evaluate the adaptability of the genotypes to the new target environment. The most commonly used types of environmental covariates are soil and meteorological covariates (van Eeuwijk et al., 2016). The regression on environmental covariates is usually modelled by fixed effects. This type of modelling is appropriate when only studying the pattern of GEI at the tested locations. Such models are also appropriate for making predictions in an unstructured TPE. However, when the TPE is sub-divided into zones, it is necessary to model genotypic effects as random to borrow strength between zones (Kleinknecht et al., 2013; Piepho et al., 2016). If such modelling is coupled with factorial regression approaches, genotype-specific regression coefficients must be modelled as random effects and give rise to what are known as random coefficient (RC) models (Longford, 1993; Milliken and Johnson, 2002). In Chapter 4, the utilisation of RC models to improve the precision and accuracy of yield predictions in some new locations is explored.

1.6 Cross-validation

Cross-validation (CV) is a method to evaluate statistical methods' performance by estimating the test error rate (James et al., 2013). A CV is conducted to evaluate a model's performance, which is known as model assessment, and select a model with a proper level of flexibility, which is known as model selection (James et al., 2013). In the CV, a dataset is split into a training set and a validation set. The training set is used to train the model, while the validation set is used to validate the model's prediction from the training set.

The two most-used methods to conduct CV are leave-one-out (LOO) and *k*-fold CV. The LOO CV leaves one data point as the validation set. Thus, if there is a set of *n* data points, there will be *n* iterations of fitting. For example, with 10 data points, ten iterations are done because of each time, one data is left out as a validation set. A *k*-fold CV divides randomly a set of data points into *k* groups, or folds, in equal size. The first fold is kept for validation, and the model is trained on *k*-1 folds. The process is iterated *k* times, and each time a different fold or a different group of data points are used for validation. Thus, a *k*-fold CV may require fewer iterations than LOO CV. Thus, the assessment was measured based on the discrepancies between observed and predicted pairwise differences.

The difference between the predictions from the training set and the validation set will be measured using mean squared error (MSE). The smallest MSE of a model will be regarded as the best-performed model, and so may be selected because the best-performed model provides the highest prediction accuracy among the compared models.

In the MET setting, the model evaluation can be carried out using a criterion based on the pairwise differences of tested cultivars. The rationale of using pairwise differences is that the main interest in cultivar trials is to predict differences among cultivars rather than individual cultivars' performance (Piepho, 1998). Piepho (1998) proposed the mean squared error of prediction (MSEP) to assess the accuracy of estimates of differences between cultivars in various environments. Chapter 2, Chapter 3, and Chapter 4 used a measure similar to Piepho's MSEP based on differences for measuring the prediction accuracy of the models.

1.7 Objective of this study

The main objective of this work is to determine the most appropriate approach for zone-based cultivar prediction. Chapter 2 focuses on the performance of EBLUE and EBLUP for zone-based prediction. Chapter 3 deals with the stagewise analysis and

weighting methods for zone-based prediction, and Chapter 4 focuses on the random coefficient (RC) models for projecting genotype performance in some untested locations.

In Chapter 2, the CV study assesses the performance of EBLUE and EBLUP for zone-based prediction in cultivar testing, including complex variance-covariance structures in Swedish cultivar trials on two fungicide levels. This chapter deals with model selection for single-year and multi-year models. Additionally, the necessity of the division of agricultural zones/zonation is evaluated. For single-year models evaluation, a 2-fold CV was used for model evaluation. The reason for conducting this type of CV was the decreasing number of trials in recent years. Thus, the aim was to train the model with a small number of trials. For multi-year models assessment, a modified leave-one-out (LOO) CV was carried out to mimic the current Swedish practice of predicting cultivar performance based on results from five years. A set of data from five consecutive years was used as a training set. Then, the following sixth year was used as validation. Besides mimicking the current-practice, the set of cultivars in the early years and recent years differ a lot. For example, when the training set consists of recent years, and the validation set consists of early years, only very few cultivars are shared between both sets. Consequently, most of the cultivars that are predicted in the training set would not be available in the validation set because the validation set comprises early years.

The best statistical analysis strategy for zone-based prediction cultivar testing, i.e., single-stage or two-stage analyses combined with complex variance-covariance structure focused on the fungicide-treated subsets of datasets were assessed in Chapter 3. Most other studies comparing single-stage and stage-wise analyses used Pearson's moment-product correlation or Spearman's rank correlation between the cultivar estimates between those two analyses (Damesa et al., 2017; Gogel et al., 2018; Möhring and Piepho, 2009; Piepho et al., 2012a). The consequence of using these correlations was the correlation coefficient estimates often are around 0.90, implying that the single-stage and stage-wise analyses provide similar results. In comparison to the Pearson correlation, a CV study can measure the prediction errors of the model using MSEP, which is more desirable for choosing the model to predict cultivar performance in MET analysis. In this study, a LOO CV was performed for comparison and selection. One location was left out as a validation set, and used the remaining locations as a training set.

The accuracy and precision of yield predictions in some new locations using the RC models were assessed in Chapter 4. The locations represent growers' fields. The prediction accuracy was evaluated via a CV study. Again, a LOO CV was performed for comparison and selection. For the models with covariates such as

the RC models, the covariates in the validation set were used for predictions. The precision of predictions is assessed with standard errors of predictions of genotypic values (SEPV) and standard errors of the predictions of pairwise differences of genotypic values (SEPD).

Chapter 2

A cross-validation of statistical models for zoned-based prediction in cultivar testing

Harimurti Buntaran[1,2], Hans-Peter Piepho[3], Jannie Hagman [2], Johannes Forkman[1,2]

1. Department of Energy and Technology, Swedish University of Agricultural Sciences, Box 7032, 750 07 Uppsala, Sweden
2. Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Box 7043, 750 07 Uppsala, Sweden
3. Biostatistics Unit, Institute of Crop Science, University of Hohenheim, 70599 Stuttgart, Germany

This chapter is published as:

Buntaran, H., Piepho, H.-P., Hagman, J., and Forkman, J. 2019. A cross-validation of statistical models for zoned-based prediction in cultivar testing. Crop Science 59:1544–1553. doi.org/10.2135/cropsci2018.10.0642

This chapter was part of:

Buntaran, H. 2019. Assessment of statistical analysis of Swedish cultivar testing: a cross-validation study for model selection. Licentiate of Philosophy, Sveriges lantbruksuniversitet.

RESEARCH

A Cross-Validation of Statistical Models for Zoned-Based Prediction in Cultivar Testing

Harimurti Buntaran,* Hans-Peter Piepho, Jannie Hagman, and Johannes Forkman

ABSTRACT

The principal goals of a plant breeding program are to provide breeders with cultivar information for selection purposes and to provide farmers with high-yielding and stable cultivars. For that reason, multi-environment trials need to be done to predict future cultivar yield, and a robust statistical procedure is needed to provide reliable information on the tested cultivars. In Sweden, the statistical procedure follows the tradition of modeling cultivar effects as fixed. Moreover, the analysis is performed separately by zone and level of fungicide treatment, and so the factorial information regarding cultivar × zone × fungicide combinations is not explored. Thus, the question arose whether the statistical method could be improved to increase accuracy in zone-based cultivar prediction, since the cultivar recommendation is zone based. In this paper, the performance of empirical best linear unbiased estimation (E-BLUE) and empirical best linear unbiased prediction (E-BLUP) are compared using cross-validation for winter wheat (*Triticum aestivum* L.) and spring barley (*Hordeum vulgare* L.), in single-year and multiyear series of trials. Data were obtained from three agricultural zones of Sweden. Several linear mixed models were compared, and model performance was evaluated using the mean squared error of prediction criterion. The E-BLUP method outperformed the E-BLUE method in both crops and series. The prediction accuracy for zone-based yield was improved by using E-BLUP because the random-effects assumption for cultivar × zone interaction allows information to be borrowed across zones. We conclude that E-BLUP should replace the currently used E-BLUE approach to predict zone-based cultivar yield.

H. Buntaran and J. Forkman, Dep. of Energy and Technology, Swedish Univ. of Agricultural Sciences, Box 7032, 750 07 Uppsala, Sweden; H. Buntaran, J. Hagman, and J. Forkman, Dep. of Crop Production Ecology, Swedish Univ. of Agricultural Sciences, Box 7043, 750 07 Uppsala, Sweden; H.-P. Piepho, Biostatistics Unit, Institute of Crop Science, Univ. of Hohenheim, 70599 Stuttgart, Germany. Received 23 Oct. 2018. Accepted 25 Mar. 2019. *Corresponding author (harimurti.buntaran@slu.se). Assigned to Associate Editor Lucia Gutierrez.

Abbreviations: BLUE, best linear unbiased estimation; BLUP, best linear unbiased prediction; CV, cross-validation; DMY, dry matter yield; E-BLUE, empirical best linear unbiased estimation; E-BLUP, empirical best linear unbiased prediction; FA, factor analytic; MET, multi-environment trials; MF, multiyear and fixed effects for cultivar; MR, multiyear and random effects for cultivar; MSEP, mean squared error of prediction differences; SF, single-year and fixed effects for cultivar; SR, single-year and random effects for cultivar.

The aim of multi-environment trials (METs) is to evaluate and test the performance of cultivars in various environmental conditions. The MET results not only provide cultivar information to breeders for selection purposes but also are the basis for advice to farmers in deciding which cultivar is the best or the most suitable concerning their local field conditions. Thus, reliable statistical methods are necessary to give both breeders and farmers accurate information.

In Swedish cultivar trials, the statistical method used for analyzing MET data has not been changed for many years. Moreover, the number of trials has been decreasing in recent years. Hence, there is a demand for improvement in statistical analysis to provide better accuracy for zoned-based cultivar performance assessment and ranking in different environments based on a reduced number of trials. Currently, the analyses are done with an unweighted two-stage analysis (Möhring and Piepho, 2009). At the first stage, each experiment is analyzed using a linear mixed

Published in Crop Sci. 59:1544–1553 (2019).
doi: 10.2135/cropsci2018.10.0642

Chapter 3

Cross-validation of stagewise mixed-model analysis of Swedish variety trials with winter wheat and spring barley

Harimurti Buntaran[1,2,3], Hans-Peter Piepho[3], Paul Schmidt[3], Jesper Rydén[1], Magnus Halling [2], Johannes Forkman[1,2]

1. Department of Energy and Technology, Swedish University of Agricultural Sciences, Box 7032, 750 07 Uppsala, Sweden
2. Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Box 7043, 750 07 Uppsala, Sweden
3. Biostatistics Unit, Institute of Crop Science, University of Hohenheim, 70599 Stuttgart, Germany

This chapter is published as:

Buntaran, H., Piepho, H.-P., Schmidt, P., Rydén, J., Halling, M., and Forkman, J. 2020. Cross-validation of stagewise mixed-model analysis of swedish variety trials with winter wheat and spring barley. Crop Science 60:2221–2240. doi.org/10.1002/csc2.20177

This chapter was part of:

Buntaran, H. 2019. Assessment of statistical analysis of Swedish cultivar testing: a cross-validation study for model selection. Licentiate of Philosophy, Sveriges lantbruksuniversitet.

Received: 5 November 2019 | Accepted: 16 April 2020 | Published online: 7 July 2020

DOI: 10.1002/csc2.20177

Crop Science

ORIGINAL RESEARCH ARTICLE

Crop Breeding & Genetics

Cross-validation of stagewise mixed-model analysis of Swedish variety trials with winter wheat and spring barley

Harimurti Buntaran[1,2,3] | **Hans-Peter Piepho**[1] | **Paul Schmidt**[1] | **Jesper Rydén**[2] | **Magnus Halling**[3] | **Johannes Forkman**[2,3]

[1] Biostatistics Unit, Institute of Crop Science, Univ. of Hohenheim, Fruwirthstraße 23, Stuttgart 70599, Germany

[2] Department of Energy and Technology, Swedish Univ. of Agricultural Sciences, Box 7032, Uppsala 750 07, Sweden

[3] Department of Crop Production Ecology, Swedish Univ. of Agricultural Sciences, Box 7043, Uppsala 750 07, Sweden

Correspondence
Harimurti Buntaran, Biostatistics Unit, Institute of Crop Science, University of Hohenheim, Fruwirthstraße 23, 70599 Stuttgart, Germany.
Email: harimurti.buntaran@uni-hohenheim.de

Assigned to Associate Editor Alexander Lipka.

Funding information
Stiftelsen lantbruksforskning - Swedish farmers' foundation for agricultural research, Grant/Award Number: O-17-20-963

Abstract

In cultivar testing, linear mixed models have been used routinely to analyze multienvironment trials. A single-stage analysis is considered as the gold standard, whereas two-stage analysis produces similar results when a fully efficient weighting method is used, namely when the full variance–covariance matrix of the estimated means from Stage 1 is forwarded to Stage 2. However, in practice, this may be hard to do and a diagonal approximation is often used. We conducted a cross-validation with data from Swedish cultivar trials on winter wheat (*Triticum aestivum* L.) and spring barley (*Hordeum vulgare* L.) to assess the performance of single-stage and two-stage analyses. The fully efficient method and two diagonal approximation methods were used for weighting in the two-stage analyses. In Sweden, cultivar recommendation is delineated by zones (regions), not individual locations. We demonstrate the use of best linear unbiased prediction (BLUP) for cultivar effects per zone, which exploits correlations between zones and thus allows information to be borrowed across zones. Complex variance–covariance structures were applied to allow for heterogeneity of cultivar × zone variance. The single-stage analysis and the three weighted two-stage analyses all performed similarly. Loss of information caused by a diagonal approximation of the variance–covariance matrix of adjusted means from Stage 1 was negligible. As expected, BLUP outperformed best linear unbiased estimation. Complex

Abbreviations: σ^2_C, variance component estimate of the cultivar; σ^2_{CZ}, variance component estimate of cultivar × zone; σ^2_{ZCL}, variance component estimate of cultivar × location; σ^2_{ZL}, variance component estimate of location; 1S, single-stage; 2S, two-stage; AVVAR, average variance of a difference; BLUE, best linear unbiased estimation; BLUP, best linear unbiased prediction; C × Z, cultivar × zone; CS, compound symmetry; EBLUE, empirical BLUE; EBLUP, empirical BLUP; F, fixed effects for cultivars; FA, factor-analytic; FA1, factor-analytic order 1; FE, fully efficient; ID, identity; LR, location-specific residual variances; MET, multienvironment trial; MSEP, mean squared error of prediction differences; SW, Smith's weighting; TPE, target population of environments; U, unweighted; UN, unstructured; W, weighted; ZR, zone-specific residual variances.

Chapter 4

Projecting results of zoned multi-environment trials to new locations using environmental covariates with random coefficient models: accuracy and precision

Harimurti Buntaran[1], Johannes Forkman[2], Hans-Peter Piepho[1]

1. Biostatistics Unit, Institute of Crop Science, University of Hohenheim, 70599 Stuttgart, Germany

2. Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Box 7043, 750 07 Uppsala, Sweden

This chapter is published as:

Buntaran, H., Forkman, J., and Piepho, H.-P. 2021. Projecting results of zoned multi-environment trials to new locations using environmental covariates with random coefficient models: accuracy and precision. Theoretical and Applied Genetics 134:1513–1530. doi.org/10.1007/s00122-021-03786-2

Theoretical and Applied Genetics (2021) 134:1513–1530
https://doi.org/10.1007/s00122-021-03786-2

ORIGINAL ARTICLE

Projecting results of zoned multi-environment trials to new locations using environmental covariates with random coefficient models: accuracy and precision

Harimurti Buntaran[1] · Johannes Forkman[2] · Hans-Peter Piepho[1]

Received: 27 September 2020 / Accepted: 29 January 2021 / Published online: 8 April 2021

Abstract

Key message **We propose the utilisation of environmental covariates in random coefficient models to predict the genotype performances in new locations.**

Abstract Multi-environment trials (MET) are conducted to assess the performance of a set of genotypes in a target population of environments. From a grower's perspective, MET results must provide high accuracy and precision for predictions of genotype performance in new locations, i.e. the grower's locations, which hardly ever coincide with the locations at which the trials were conducted. Linear mixed modelling can provide predictions for new locations. Moreover, the precision of the predictions is of primary concern and should be assessed. Besides, the precision can be improved when auxiliary information is available to characterize the targeted locations. Thus, in this study, we demonstrate the benefit of using environmental information (covariates) for predicting genotype performance in some new locations for Swedish winter wheat official trials. Swedish MET locations can be stratified into zones, allowing borrowing information between zones when best linear unbiased prediction (BLUP) is used. To account for correlations between zones, as well as for intercepts and slopes for the regression on covariates, we fitted random coefficient (RC) models. The results showed that the RC model with appropriate covariate scaling and model for covariate terms improved the precision of predictions of genotypic performance for new locations. The prediction accuracy of the RC model was competitive compared to the model without covariates. The RC model reduced the standard errors of predictions for individual genotypes and standard errors of predictions of genotype differences in new locations by 30–38% and 12–40%, respectively.

Introduction

The main goal of a plant breeding programme is to develop well-adapted genotypes in a target population of environments (TPE). Multi-environment trials (MET) are conducted to evaluate candidate genotypes in the TPE, and to understand and exploit the pattern of genotype × environment interactions (GEI) in the TPE. GEI is the differential response of genotypes across different environments (Kang and Gorman 1989). GEI in a TPE can be exploited to make more targeted predictions and recommendations on cultivars. This is of particular interest when there is crossover interaction, which poses a challenge when selecting genotypes for broad adaptation.

Identification of environmental covariates that are responsible for GEI is useful to enhance the predictive capability of MET analyses (Heslot et al. 2014) and evaluate the adaptability of the genotypes to the new target environment. The most commonly used types of environmental covariates are soil and meteorological covariates (van Eeuwijk et al. 2016). Incorporating environmental covariates in the GEI analysis has been done by factorial regression (Denis 1988; Piepho et al. 1998; van Eeuwijk and Elgersma 1993). Furthermore, environmental covariates have been used in a linear mixed model framework, such as in quantitative trait loci (QTL)

Communicated by Martin Boer.

✉ Hans-Peter Piepho
hans-peter.piepho@uni-hohenheim.de

Harimurti Buntaran
harimurti.buntaran@uni-hohenheim.de

Johannes Forkman
johannes.forkman@slu.se

[1] Biostatistics Unit, Institute of Crop Science, University of Hohenheim, Fruwirthstraße 23, 70599 Stuttgart, Germany

[2] Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Box 7043, 750 07 Uppsala, Sweden

Chapter 5

General Discussion

MET is crucial to evaluate cultivars' performance and provide a report for breeders and a recommendation for growers. Due to the heterogeneity of environmental conditions, it is important to use a statistical model that captures this heterogeneity, and so accurate and precise estimates for the future performance of the tested cultivars can be obtained. This thesis demonstrates the use of EBLUP as a projection of cultivars' performance and the eminence of weighting methods to account for trials' heterogeneity in MET. Furthermore, the utilisation of environmental covariates to project cultivars' performance in new locations and to improve the precision of the predictions *per se* is also presented.

5.1 Evaluation of EBLUE and EBLUP performances for zone-based predictions

In Chapter 2, it was shown that random-cultivar-effects (EBLUP) model is preferable for routine zone-based yield prediction compared to fixed-cultivar-effects (EBLUE) models. The EBLUP models achieved lower MSEPs than the EBLUE models for all datasets, i.e., for single-year and multi-year in the winter wheat and spring barley datasets. The EBLUE model performed poorly and was the least performing since it had the largest MSEP. In Chapter 3, where the EBLUE and EBLUP were compared in the stagewise analysis with various weighting methods, the EBLUP also outperformed the EBLUE. Thus, the CV study in the Chapter 2 and Chapter 3 confirmed that EBLUP had better accuracy for zone-based predictions.

Regarding the necessity of zonation, Chapter 2 showed the results were considerably different between the two crops. The model without zonation performed modestly for winter wheat datasets than for spring barley datasets. A plausible biological reason is that winter wheat is grown in winter weather conditions with large local variation, as compared to spring barley, which is sown in the springtime.

Thus, spring barley is grown under less diverse local conditions. In the winter, environmental conditions vary locally, from mild and humid to cold and dry, causing different stress factors to predominate (Bergjord Olsen et al., 2018). Since the model with zonation generally performed better than the model without zonation, it was essential to incorporate the zone in the analysis.

5.1.1 Go for EBLUP?

Since it becomes a routine procedure to use LMM for analysing MET data, the frequent question is whether to model cultivar effects as fixed or random. Modelling cultivar effects as random is advised when the primary goal is to select the best cultivars from the population under study and when the effects and residuals presumably follow a normal distribution. From a statistical perspective, using EBLUP to predict cultivars' performance is better than EBLUE because the rankings of the estimated cultivars are expected to be close to the rankings of the cultivar effects and provide more accurate predictions (McCulloch et al., 2008; Searle et al., 1992; Smith et al., 2005). From a biological perspective, the cultivars can be considered as a random sample of the current genetic variability (Curti et al., 2014). Furthermore, EBLUP is necessary to account for heterogeneity since modelling variance-covariance structure is not applicable in EBLUE.

The shrinkage feature in EBLUP can mitigate the over-optimistic predictions of the top-performing cultivars and over-pessimistic predictions of poorly performing cultivars. The magnitude of the shrinkage depends on the "shrinkage factor", and, in a simple model, the shrinkage factor is a function of heritability as described in Galwey (2014, p. 169). Shrinkage thus reduces the spread of the predictions compared to fixed effects estimation (Robinson, 1991). Means higher than the overall mean are shrunk downwards to the overall mean, while the means that are lower than the overall mean are slightly increased (shrunken upwards towards the overall mean). This shrinkage is shown by the different ranking of the cultivars between the EBLUE and EBLUP methods, as shown in Chapter 2 and Chapter 3. The best cultivar, according to the EBLUE method, was different. The ranking of the other cultivars was also different between the two models. For cultivar recommendation, where a correct ranking of cultivars is essential, the EBLUP method should be preferred due to its smaller MSEP.

Furthermore, with random cultivar$\times$zone effects, the accuracy of predictions within zones is improved due to the information is borrowed across zones by exploiting the cultivar correlations between zones (Atlin et al., 2000; Kleinknecht et al., 2013; Piepho et al., 2016). Lee et al. (2017, p. 144) pointed out the benefits of

using BLUP as follows: "With a random effect specification, we gain significant parsimony. In such situations, even if the true model is the fixed effect model, i.e., there is no random sampling involved, the use of random effect estimation has been advocated as shrinkage estimation (James and Stein, 1992)." However, it is essential to note that when the cultivar correlations between or among zones are small, then the information that can be borrowed across zone will be very little. In this case, EBLUP will not be more beneficial than BLUE. However, when the cultivar correlations between or among zones are high, then EBLUP will be favorable to EBLUE.

The empirical datasets were not perfectly normally distributed as shown in the Supplemental Figures of Chapter 2. However, BLUP *per se* does not require normality (Searle et al., 1992, p. 270 and 273). In fact, Henderson (1963) showed the derivation of BLUP in the MME without assuming a normal distribution. In practice, the variance components are unknown and must be estimated. REML estimates may be imprecise in small datasets, which makes the benefits of using EBLUP is uncertain. The simulation study from Forkman and Piepho (2013) reported, however, that imprecise variance component estimates were not a severe problem for the application of EBLUP in small randomised complete block experiments.

5.1.2 Dealing missing data with EBLUP

The merit of EBLUP will only be valid when data are missing at random (MAR). However, it has often been a common practice to decide which cultivars should be tested in particular zones, depending on their expected performance in those zones. Specifically, cultivars might not be tested in a zone if they are expected to perform less well in that zone. In this case, the cultivars are not MAR. Forkman (2013) showed that analyses of incomplete datasets using GLSE based on mixed models with random environmental effects can give unexpected estimates. If there is a doubt that cultivars are missing at random, it might be better to use a model with EBLUE of trials because comparisons among cultivars are then based exclusively on within-trial information, and between-trial information is not recovered (Piepho et al., 2012b). Thus, it is advised to strive for complete datasets for the single-year analysis.

In the multi-year series, it is often to exclude from the analysis all cultivars that have not been tested in the latest year and at least two years. To gain benefit of using EBLUP, it is recommended that all cultivars should be retained in the analysis (Piepho and Möhring, 2006). The reason is that all cultivars involved in selection decisions should be included in the analysis to avoid selection bias, as pointed out

by (Piepho and Möhring, 2006). Piepho and Möhring (2006) also mentioned that the removal of data leads to a missing-not-at-random (MNAR) pattern that causes invalid variance component estimates. Besides, if the missing data pattern is MNAR, then EBLUP will systematically be associated with varying degrees of shrinkage, which leads to bias. For example, if a cultivar is very little tested, then the shrinkage of all its predicted effects will be large, and so the prediction will be less accurate. A good cultivar candidate still needs to be tested across many trials to obtain reliable information on its performance.

5.2 Stagewise analysis strategy

5.2.1 Weighting is crucial in stagewise analysis

Three weighting methods for the stagewise analysis were assessed in Chapter 3, i.e., fully efficient, the average variance of a difference (AVVAR) weights (Möhring and Piepho, 2009), and Smith's diagonal weights (Smith et al., 2001). Based on the MSEP from the CV study, the stagewise analysis using these three weighting methods was very similar to the single-stage analysis for both crop datasets. Furthermore, the results confirmed that the loss of information resulting from a two-stage analysis with diagonal weights instead of the single-stage analysis is acceptable (Möhring and Piepho, 2009).

Similar results of the single-stage and two-stage analysis were attained due to the weighting application in the two-stage analysis. The diagonal approximate weighting such as Smith's weighting (Smith et al., 2001) uses the diagonal part of the inverse of the variance-covariance matrix. This inverse of the variance-covariance matrix had small, and hence negligible, off-diagonal elements, whereas the diagonal elements used for weighting in Stage 2 were large by comparison. Thus, the use of two-stage weighting is reasonable. The fully efficient method was the closest one to the single-stage analysis because all information from Stage 1 was carried forward to Stage 2.

When there was no weighting applied, the two-stage unweighted strategy performed better than the single-stage method that assumed no heterogeneity variance structure in its model. Thus, it was shown that the simple two-stage unweighted strategy produced better predictions than the far too simple single-stage strategy. Therefore, the use of adjusted means from Stage 1 was more accurate than a single-stage approach that neglects the heterogeneity of variance in replicates and incomplete blocks across locations.

The idea of two-stage analysis can be viewed as being similar to Bayesian Updating, which utilise the prior distribution from the previous posterior distribution (Sorensen and Gianola, 2002). Thus, the Bayes theorem has "memory," and the inferences can be updated sequentially. Compared to the two-stage analysis, the result of Stage 1 can be regarded as a posterior distribution that will be used as the prior distribution for Stage 2. In fact, BLUP is empirical Bayesian when the distribution of random effects is Gaussian as mentioned by Robinson (1991). Hence, Bayesian Updating might be comparable with the "frequentist" BLUP of two-stage analysis. Therefore, a further study comparing the "frequentist" two-stage analysis with the Bayesian Updating framework would be worthwhile.

5.2.2 MSEP vs. correlation coefficients

It is more difficult to detect that EBLUP performed better than EBLUE when Pearson's and Spearman's correlation coefficients were used exclusively as shown in Chapter 3 than when using MSEP. Besides, it is also difficult to see any difference in performance between the single-stage and two-stage approaches. The MSEP provides a more apparent distinction between the EBLUP methods and the EBLUE method and clearer discrimination between the single-stage and the two-stage approach. According to Kobayashi and Salam (2000), correlation is not the best measure for model evaluation since the mean squared deviation is easier to interpret and more useful for direct comparisons between model output and measurement. Thus the MSEP from the CV study in Chapter 3 was used as additional evidence.

5.2.3 Single-stage or stagewise analyses?

This question is inevitable since the results between the single-stage analysis and stagewise analysis with weighting were similar. The choice depends on the two aspects, the computational resources and the practicality of data handling. When the computational resources allow, the single-stage can be preferred. When the computation resources are limited, then the two-stage weighting strategy is not worse than the single-stage analysis since the results will not differ much from the single-stage analysis. Furthermore, different software packages affected the required time and memory allocation for the analysis. The software used in Chapter 3 was optimised for the single-stage strategy. Thus, when the fully efficient analysis was carried out, the memory allocation had to be increased and more time consuming than the single-stage analysis. Therefore, the diagonal approximation weighting is a reasonable option due to a more practical procedure to store the weighting, less-intensive computation, and the loss of information is negligible.

The practicality of data handling can determine the choice of strategy when, in some cases, the data from all the trials may not be available at once. In this case, the two-stage weighting analysis will be more practical in two aspects. First, since the readily available trial data can be analysed instantly and provide individual trial information. Second, time-saving by storing the adjusted means of each trial and the accompanying precision measures, which later can be used for the next stage analyses while data from other trials are still being collected. Another reason to prefer the two-stage weighting strategy is especially when each trial has a different experimental design. In this case, fitting a model for single-stage analysis may not be straightforward and easy. The two-stage weighting method is also preferable if one wants to assess each trial thoroughly because, with a vast number of trials, it will be cumbersome to check each trial thoroughly with a single-stage analysis because of the vast variance component estimates produced by the single-stage analysis.

5.2.4 In which stage should EBLUP be used?

The merit of using EBLUP for obtaining the prediction accuracy and ranking has been discussed in the Subsections 5.1.1 and 5.1.2. The use of BLUP in Stage 1 or other stages before the final stage is discouraged. The use of BLUP in any stage before the final stage will cause double shrinkage since BLUP is also used in the final stage. If BLUP were to be used in Stage 1 or other stages before the final stage, predictions would need to be unshrunk before proceeding to the next stage or the final stage (Smith et al., 2001), but it is not obvious how this should be done or how to perform the weighting in other stages and the final stage. Some progress could potentially be made by taking recourse to the so-called "deregressed proofs" as used in animal breeding (Calus et al., 2016).

5.3 Complex variance-covariance structures may not be necessary

The results of Chapter 2 and Chapter 3 indicated that the models with FA variance-covariance structure did not improve the accuracy compared to the simpler variance-covariance structures. The investigated FA covariance structure allows heterogeneous variances and unique pairwise correlations between zones. The FA structure is useful because it allows heterogeneous variance and covariance using fewer parameters than the unstructured covariance structure. However, in Chapter 2, the REML estimation for the FA structure and the model with many interaction terms combined with the heterogeneous residual structure were computationally very

demanding. Thus, in Chapter 2, the combinations of FA structures for interaction effects and heterogeneous structures for residual effects were not explored.

In Chapter 3, the FA and the UN structures were far less performant than the CS structure for single-stage analyses as well as two-stage analyses. In fact, no model with FA structure was among the top-five performing models. In fact, with merely three zones, using the FA structure has the same number of parameters as the UN structure (i.e., six parameters). The FA structure may be more useful when the number of zones is larger than three. Chapter 3 results also showed that the variance component estimates for the cultivar×zone effects were relatively small compared with the other components. Hence, there may be no need for complex variance-covariance structures for this dataset. Nevertheless, when the variance component estimates are large, then more parameters with complex variance-covariance might be needed to account for the heterogeneity of variance for cultivar×zone effects. Recently, Prus and Piepho (2021) investigated the optimal designs for trials allocation to sub-regions MET. With CS structure, the optimal designs were less sensitive to the number of locations and its variance compared to the FA structure (Prus and Piepho, 2021). Thus, it may well occur that using the FA structure was not worthwhile due to the variance estimates of cultivar×zone being relatively small. Another reason may be the relatively small number of trials in each zones.

5.4 Zone-based prediction is preferable to individual locations

Obtaining predictions per agro-ecological zone (larger TPE) is more informative than predictions for individual locations (Damesa et al., 2017). The reason is that growers are interested in the cultivar that performs well on average across broad environmental conditions and the next growing season (the next growing season can be considered as a new environment that no trial has previously been conducted in). Another reason is that when a prediction is made for an individual location, the predictions of closest trial location can be applied for the grower's field. However, in this case, the valid standard errors for the predictions cannot be achieved since the pattern of interaction between a grower's field, which is the target site, and the nearest trial location and the corresponding with years are unknown. Nonetheless, if predictions made for zones or a whole TPE, the valid inferences can be obtained due to the availability of random sample of trial locations and years for that TPE *per se* (Damesa et al., 2017).

Furthermore, from a breeder's perspective, prediction of cultivar performance in a specific location is rarely of interest. In fact, Swedish official cultivar trials has the same objective, i.e., to recommend well-performing cultivars for each zone, not

for individual trial locations. Thus, accurate information regarding which cultivars perform well within zones or perform above average across TPE is essential for growers and breeders.

5.5 Accuracy and precision in new locations

As stated before, the trials' location hardly ever coincide with the growers' field. Thus, the cultivar yield will never be equivalent to the predicted mean values from the MET analysis. The growers' field can be viewed as a new location or an untested location. While the zone-based predictions are valid for the new locations within a particular zone, the precision measures, i.e., standard errors of the predictions obtained from the MET analysis, are only valid for the trials' locations and not valid for the growers' field. Hence, the valid standard errors of the predictions need to be computed and reported to provide how precise or reliable the prediction of the cultivar that is selected by the growers. So, the location effect has to be random since the variance component estimate of location as the uncertainty is needed to compute the standard errors of the predictions for the new locations. When the location effect is fixed, the variance component estimate cannot be obtained, and the standard errors of the predictions in the new location cannot be computed.

Poorter et al. (2010) demonstrated that the quantitative environmental factors in the response curve framework could be used to form a reference for results of future experiments. In the same vein, the environmental factors or covariates that represent the growers' field can improve the accuracy and precision of the cultivar predictions. The environmental covariates can be integrated into the random cultivar effect in the random coefficient (RC) models (Longford, 1993; Milliken and Johnson, 2002).

5.5.1 Precision improvement in the RC models

In Chapter 4, a total of 14 models, including the RC models and fixed effects of cultivar was assessed via a CV study, and their standard errors of prediction values (SEPV) and standard errors of the predictions of pairwise differences of genotypic values (SEPD) for the cultivars in the new locations. The selected covariate was clay and used in the quadratic term. The RC models reduced the SEPV averages for all new locations by 30–38%, and the SEPD averages by 12–40%. Thus, Chapter 4 demonstrated the RC models could improve the precision of predictions of genotypes performance and the precision of genotypes comparisons. Using the covariates in the random coefficients term, the SEPV is evaluated at specific values

of the covariates (Milliken and Johnson, 2002), which can substantially decrease the SEPV of the RC models compared to the models without any random coefficients.

The RC model that utilised the linear and quadratic term in the cultivar×zone interaction effects, but not in the cultivar main effects, performed better than the most complex RC model. The justification that the complex model was not better than the reduced one was due to the quadratic regression per cultivar showing that the variation between cultivars was not that large, as the regression curves of cultivars were close to each other. Hence, the inclusion of random coefficients in the cultivar term may not be worthwhile. The random coefficients in the cultivar×zone term were more beneficial than in the cultivar main-effect term since the SEPD for the RC model that only had the random coefficients in the cultivar main-effect was higher that for the RC model with random coefficients in the cultivar×zone term. Moreover, the quadratic regression cultivar×zone term showed that the variation between cultivar×zone effects is large, as the curves of cultivar×zone are more widely spread out than the quadratic regression per cultivar.

5.5.2 Accuracy of the RC models

Although the precision was certainly improved via the RC model, its MSEP was not the smallest. The models with the smallest MSEP were the models without random coefficient terms but utilise EBLUP for the cultivar effect. Nevertheless, the difference in the MSEP between the best RC model and the models that had the smallest MSEP was negligible. Compared to the accuracy, the models that had the smallest MSEP had huge prediction intervals and uncertainty. In this case, although the models provided more accurate predictions, it will be very uncertain that the yield of the cultivar will be close to the predictions. The model can be selected by jointly considering SEPV, SEPD, and the MSEP. In this case, the RC model is preferred since the uncertainty was far lower, although its MSEP was not the smallest.

It should be noted that the model with fixed effects for cultivar×zone and interactions with the covariate was not preferable, although the SEPV and the SEPD were comparable to the RC models. The reason is because the MSEP of this model was the larger. Thus, this model was the least performant. The predictions of this model were indeed far less accurate or more biased with narrow prediction intervals.

The utilisation of the RC models for improving prediction accuracy was also demonstrated by Baba et al. (2020) and Jarquín et al. (2014).Baba et al. (2020) showed that using RC models for multitrait analysis, and modelling the covariates with fixed Legendre regression coefficients, improved the prediction accuracy for a trait with a limited number of records or low heritability. Compared to Baba et al. (2020),

Chapter 4 did not use Legendre but only quadratic terms of covariate. The study of Jarquín et al. (2014) also used RC models to improve the prediction accuracy of prediction of the performance of newly released lines and predictions in incomplete field Trials. The major difference between Chapter 4 and Jarquín et al. (2014) was that Jarquín et al. (2014) used an extensive number of environmental covariates with marker data and utilised this information by computing an environmental kinship matrix, for which a single variance component was fitted. Thus, implicitly, the model used by Jarquín et al. (2014) assumes that the slopes for the different covariates have the same variance and no correlation between them. On the other hand, in Chapter 4, a vast number of covariates and marker data were not available but allowed for heterogeneity in variance between slopes, and for covariance between slopes and intercepts to maintain the invariance feature of RC models (Longford, 1993; Piepho and Ogutu, 2002; Wolfinger, 1996).

5.6 Variance-covariance structure for RC models

As stated before, the utilisation of unstructured covariance or allowing for heterogeneity in variance between slopes, and for covariance between slopes and intercepts is crucial to maintain the invariance feature of RC models. However, when the number of covariates increases, fitting such RC models becomes more challenging, and it is not apparent how this can best be done. One option to circumvent numerical problems is to fit a low-rank approximation to the unstructured variance-covariance matrix for intercepts and slopes, i.e. an FA model (Jennrich and Schluchter, 1986). Fitting an FA model guarantees that the variance-covariance matrix is positive definite. If the order of the FA model equals the number of slope terms plus the intercept, the model is equivalent to the unstructured model, whereas lower-rank approximations are obtained by reducing the order.

5.7 Handling the covariates

5.7.1 Covariate selection

The benefit of using a covariate in the RC models depends on choosing the appropriate covariate. Covariates can initially be selected based on the biological considerations. Still, it is necessary to check whether these covariate candidates improve the model fit. In fact, with a large number of covariates, covariate selection will be beneficial. R-square (R^2) for mixed models (Piepho, 2019) is an option for covariate selection, as demonstrated in Hadasch et al. (2020). Further research to

explore the best approach to accommodate a larger number of covariates is certainly worthwhile.

5.7.2 Covariate scale

The covariate scale is crucial when implementing RC models. In Chapter 4 the covariate scale issue occurred when the most complex RC model was fitted. This scaling issue occurred when the random coefficients were fitted to both cultivar and cultivar×zone effects, where the variance component of the linear term of the random coefficients of the cultivar was 0. This problem occurred due to the cultivar's main effect having a lower variance than the cultivar×zone interaction effect, and there was likely competition between cultivar main effect and cultivar×zone interaction effect absorbing the variance. Some scalings such as subtraction-of-the-minimum and covariate-centering were also attempted. When these scalings were used, the RC models did not converge. Thus, the covariate scaling is essential for model convergence and to obtain the appropriate variance parameter estimates. The clay covariate was scaled for all fitted models by $(clay - 40)/10$ because it yielded a positive definite variance-covariance matrix.

5.8 Prospect and outlook

The EBLUP model was the preferable method for zone-based cultivar prediction as demonstrated in Chapter 2 and Chapter 3. Moreover, the stagewise weighting strategy was comparable to the single-stage strategy. Thus, the EBLUP method with stagewise weighting strategy was recommended for the routine analysis. However, implementing this strategy comes with price that the datasets should be in the MAR condition, which required to retain all cultivars in the analysis.

When the environmental covariates are available, the EBLUP method can be extended to RC models as presented in Chapter 4. The benefit of utilising RC models is that the precision of predictions for new locations is improved by using the environmental covariates with respect to the new locations. Recently, a study by Neyhart et al. (2021) followed the same vein of predicting genotype performance in unobserved environments. The authors assessed genome-wide predictions in the unobserved environments for both between and within breeding generations. A recent paper by Resende et al. (2020) proposed the geospatial (geographic information system/GIS) genotype-environment interaction (GIS–GEI) method within an enviromics framework. The GIS-GEI involves the joint analysis of MET data, accounting for phenotypic, genotypic, and envirotypic sources of information. It

is anchored into a geoprocessing environment (a land area with nine pixels) that employs enviromic markers, e.g., time-trend climate data, landscape or management treatment information, obtained through modern envirotyping techniques. The proposed RC modelling in Chapter 4 approach is ideally suited for integration in an enviromics-driven GIS–GEI framework. Additionally, the RC modelling can be implemented in the Bayesian framework as proposed by Theobald et al. (2002), who utilised a Bayesian method for making predictions with incorporating environmental covariates.

References

A

Allard, R. W. (1960). *Principles of Plant Breeding*. John Wiley & Sons, Inc., New York.

Allard, R. W. and Bradshaw, A. D. (1964). Implications of genotype-environmental interactions in applied plant breeding[1]. *Crop Science*, 4:503–508. https://doi.org/10.2135/cropsci1964.0011183X000400050021x.

Atlin, G. N., Baker, R. J., McRae, K. B., and Lu, X. (2000). Selection response in subdivided target regions. *Crop Science*, 40:7–13. http://dx.doi.org/10.2135/cropsci2000.4017.

B

Baba, T., Momen, M., Campbell, M. T., Walia, H., and Morota, G. (2020). Multi-trait random regression models increase genomic prediction accuracy for a temporal physiological trait derived from high-throughput phenotyping. *PLOS ONE*, 15:e0228118. https://doi.org/10.1371/journal.pone.0228118.

Baker, R. (1990). Crossover genotype–environmental interaction in spring wheat. In Kang, M., editor, *Genotype-by-Environment Interaction and Plant Breeding*, page 42–51. Louisiana State University Agricultural Center, Baton Rouge.

Bergjord Olsen, A. K., Persson, T., de Wit, A., Nkurunziza, L., Sindhøj, E., and Eckersten, H. (2018). Estimating winter survival of winter wheat by simulations of plant frost tolerance. *Journal of Agronomy and Crop Science*, 204:62–73. https://doi.org/10.1111/jac.12238.

Bernardo, R. (1996). Best linear unbiased prediction of maize single-cross performance. *Crop Science*, 36:50–56. https://doi.org/10.2135/cropsci1996.0011183X003600010009x.

Bernardo, R. (2020). *Breeding for quantitative traits in plants*. Stemma Press, Woodbury, third edition.

Blasco, A. (2017). *Bayesian Data Analysis for Animal Scientists: The Basics*. Springer International Publishing, Cham.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, 71:791–799. https://doi.org/10.2307/2286841.

Box, G. E. P., Hunter, S., and Hunter, W. G. (2005). *Statistics for experimenters: design, innovation, and discovery*. John Wiley & Sons, Inc., Hoboken.

Buntaran, H. (2019). *Assessment of statistical analysis of Swedish cultivar testing: a cross-validation study for model selection*. Licentiate thesis, Sveriges lantbruksuniversitet.

Bustos-Korts, D., Malosetti, M., Chapman, S., and van Eeuwijk, F. (2016). Modelling of genotype by environment interaction and prediction of complex traits across multiple environments as a synthesis of crop growth modelling, genetics and statistics. In Yin, X. and Struik, C. P., editors, *Crop Systems Biology: Narrowing the gaps between crop modelling and genetics*, pages 55–82. Springer International Publishing, Cham. http://dx.doi.org/10.1007/978-3-319-20562-5_3.

C

Calus, M. P. L., Vandenplas, J., ten Napel, J., and Veerkamp, R. F. (2016). Validation of simultaneous deregression of cow and bull breeding values and derivation of appropriate weights. *Journal of Dairy Science*, 99:6403–6419. https://doi.org/10.3168/jds.2016-11028.

Comstock, R. (1977). Quantitative genetics and the design of breeding programme. In Pollak, E. J., Kempthorne, O., and Bailey, J. T. B., editors, *International Conference on Quantitative Genetics*, pages 705–718. Iowa State University Press.

Cooper, M. and Byth, D. E. (1996). Understanding plant adaptation to achieve systematic applied crop improvement - a fundamental challenge. In Cooper, M. and Hammer, G. L., editors, *Plant Adaptation and Crop Improvement*. CAB International, Wallingford.

Cooper, M. and DeLacy, I. H. (1994). Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theoretical and Applied Genetics*, 88:561–572. https://doi.org/10.1007/BF01240919.

Cooper, M., Messina, C. D., Podlich, D., Totir, L. R., Baumgarten, A., Hausmann, N. J., Wright, D., and Graham, G. (2014). Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. *Crop and Pasture Science*, 65:311–336. https://doi.org/10.1071/CP14007.

Curti, R. N., de la Vega, A. J., Andrade, A. J., Bramardi, S. J., and Bertero, H. D. (2014). Multi-environmental evaluation for grain yield and its physiological determinants of quinoa genotypes across Northwest Argentina. *Field Crops Research*, 166:46–57. https://doi.org/10.1016/j.fcr.2014.06.011.

D

Damesa, T. M., Möhring, J., Worku, M., and Piepho, H.-P. (2017). One step at a time: Stage-wise analysis of a series of experiments. *Agronomy Journal*, 109:845–857. http://dx.doi.org/10.2134/agronj2016.07.0395.

de Leon, N., Jannink, J.-L., Edwards, J. W., and Kaeppler, S. M. (2016). Introduction to a special issue on genotype by environment interaction. *Crop Science*, 56:2081–2089. https://doi.org/10.2135/cropsci2016.07.0002in.

DeLacy, I. H., Basford, K. E., Cooper, M., Bull, J. K., and McLaren, C. G. (1996). Analysis of multi-environment trials - an historical perspective. In Cooper, M. and Hammer, G. L., editors, *Plant Adaptation and Crop Improvement*. CAB International, Wallingford.

DeWitt, T. J. and Scheiner, S. M. (2004). *Phenotypic Plasticity: Functional and conceptual approaches*. Oxford University Press, New York.

E

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics*, 3:1–21. https://doi.org/10.2307/3001534.

F

Fehr, W. R. (1987). *Principles of Cultivar Development*. Macmillan Publishing Company, New York.

Forkman, J. (2013). The use of a reference variety for comparisons in incomplete series of crop variety trials. *Journal of Applied Statistics*, 40:2681–2698. https://doi.org/10.1080/02664763.2013.825703.

Forkman, J. and Piepho, H. P. (2013). Performance of empirical BLUP and Bayesian prediction in small randomized complete block experiments. *The Journal of Agricultural Science*, 151:381–395. https://doi.org/10.1017/S0021859612000445.

G

Galwey, N. W. (2014). *Introduction to Mixed Modelling: Beyond Regression and Analysis of Variance*. John Wiley & Sons, West Sussex, second edition.

Gepts, P. and Pfeiffer, T. (2018). From classical plant breeding to molecular crop improvement. In Chrispeels, M. J. and Gepts, P., editors, *Plants, genes & agriculture: sustainability through biotechnology*. Oxford University Press, New York.

Gogel, B., Smith, A., and Cullis, B. (2018). Comparison of a one- and two-stage mixed model analysis of Australia's National Variety Trial Southern Region wheat data. *Euphytica*, 214:44. https://doi.org/10.1007/s10681-018-2116-4.

H

Hadasch, S., Laidig, F., Macholdt, J., Bönecke, E., and Piepho, H. P. (2020). Trends in mean performance and stability of winter wheat and winter rye yields in a long-term series of variety trials. *Field Crops Research*, 252:107792. https://doi.org/10.1016/j.fcr.2020.107792.

Harville, D. (1976). Extension of the Gauss-Markov Theorem to include the estimation of random effects. *The Annals of Statistics*, 4:384–395. http://www.jstor.org/stable/2958209.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning. Springer Series in Statistics*. Springer, New York.

Henderson, C. R. (1950). Estimation of genetic parameters. *Annals of Mathematical Statistics*, 21:309–310.

Henderson, C. R. (1963). Selection index and expected genetic advance. In Hanson, W. D. and Robinson, H., editors, *Statistical Genetics and Plant Breeding*, volume 982, pages 141–163, Washington DC. National Academy of Sciences, National Research Council.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31:423–447. https://doi.org/10.2307/2529430.

Henderson, C. R. (1977). Best linear unbiased prediction of breeding values not in the model for records. *Journal of Dairy Science*, 60:783–787. http://www.sciencedirect.com/science/article/pii/S0022030277839350.

Henderson, C. R. (1984). *Applications of Linear Models in Animal Breeding*. University of Guelph, Guelph.

Heslot, N., Akdemir, D., Sorrells, M. E., and Jannink, J.-L. (2014). Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theoretical and Applied Genetics*, 127:463–480. https://doi.org/10.1007/s00122-013-2231-5.

I

Isik, F., Holland, J., and Maltecca, C. (2017). Multi environmental trials. In *Genetic Data Analysis for Plant and Animal Breeding*, pages 227–262. Springer International Publishing, Cham.

J

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer, New York.

James, W. and Stein, C. (1992). Estimation with quadratic loss. In Kotz, S. and Johnson, N. L., editors, *Breakthroughs in Statistics: Foundations and Basic Theory*, pages 443–460. Springer, New York.

Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., Piraux, F., Guerreiro, L., Pérez, P., Calus, M., Burgueño, J., and de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics*, 127:595–607. https://doi.org/10.1007/s00122-013-2243-1.

Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 42:805–820. www.jstor.org/stable/2530695.

K

Kang, M. S. (1990). *Genotype-by-Environment Interaction and Plant Breeding*. Louisiana State University Agricultural Center, Baton Rouge.

Kang, M. S. (2020). Genotype–environment interaction and stability analyses: An update. In Kang, M. S., editor, *Quantitative Genetics, Genomics and Plant Breeding*. CAB International, Wallingford, second edition.

Kang, M. S. and Gorman, D. P. (1989). Genotype × environment interaction in maize. *Agronomy Journal*, 81:662–664. https://doi.org/10.2134/agronj1989.00021962008100040020x.

Kleinknecht, K., Möhring, J., Singh, K., Zaidi, P., Atlin, G., and Piepho, H. (2013). Comparison of the performance of best linear unbiased estimation and best linear unbiased prediction of genotype effects from zoned Indian maize data. *Crop Science*, 53:1384–1391. http://dx.doi.org/10.2135/cropsci2013.02.0073.

Kobayashi, K. and Salam, M. U. (2000). Comparing simulated and measured values using mean squared deviation and its components. *Agronomy Journal*, 92:345–352. https://doi.org/10.2134/agronj2000.922345x.

Kotz, S., Read, C., Balakrishnan, N., Vidakovic, B., and Johnson, N. L. (2006). Accuracy and precision. In *Encyclopedia of Statistical Sciences*, volume 1. John Wiley & Sons, Inc., New Jersey.

L

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38:963–974. https://doi.org/10.2307/2529876.

Lee, Y., Nelder, J. A., and Pawitan, Y. (2017). *Generalized linear models with random effects: Unified analysis via H-likelihood*. CRC Press, Boca Raton.

Longford, N. T. (1993). *Random Coefficient Models*. Oxford University Press, New York.

Lynch, M. and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland.

M

McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008). *Generalized, linear, and mixed models*. John Wiley & Sons, Inc., Hoboken, second edition.

McKeand, S. E., Li, B., Hatcher, A. V., and Weir, R. J. (1990). Stability parameter estimates for stem volume for loblolly pine families growing in different regions in the Southeastern United States. *Forest Science*, 36:10–17. https://doi.org/10.1093/forestscience/36.1.10.

Medina, R. C. (1992). Some exact conditional tests for the multiplicative model to explain genotype-environment interaction. *Heredity*, (2):128–132.

Meredith Jr., W. R. (1984). Quantitative genetics. In Kohel, R. J. and Lewis, C. F., editors, *Cotton and cotton improvement*, pages 131–150. Am. Soc. Agron. and Crop Sci. Soc. Am., Madison.

Milliken, G. A. and Johnson, D. E. (2002). *Analysis of Messy Data, Volume III: Analysis of Covariance*. CRC Press, Boca Raton.

Möhring, J. and Piepho, H.-P. (2009). Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Science*, 49:1977–1988. http://dx.doi.org/10.2135/cropsci2009.02.0083.

N

Neyhart, J. L., Gutiérrez, L., and Smith, K. P. (2021). Using environmental similarities to design training sets for genomewide selection. *Crop Science*, 61:396–409. https://doi.org/10.1002/csc2.20303.

P

Patterson, H. D. and Silvey, V. (1980). Statutory and recommended list trials of crop varieties in the United Kingdom. *Journal of the Royal Statistical Society: Series A (General)*, 143:219–240. https://rss.onlinelibrary.wiley.com/doi/abs/10.2307/2982128.

Piepho, H.-P. (1998). Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. *Theoretical and Applied Genetics*, 97:195–201.

Piepho, H.-P. (2019). A coefficient of determination (R^2) for generalized linear mixed models. *Biometrical Journal*, 61:860–872. https://doi.org/10.1002/bimj.201800270.

Piepho, H.-P. and Eckl, T. (2014). Analysis of series of variety trials with perennial crops. *Grass and Forage Science*, 69:431–440. https://onlinelibrary.wiley.com/doi/abs/10.1111/gfs.12054.

Piepho, H.-P. and Möhring, J. (2006). Selection in cultivar trials—is it ignorable? *Crop Science*, 46:192–201. https://doi.org/10.2135/cropsci2005.04-0038.

Piepho, H.-P., Möhring, J., Schulz-Streeck, T., and Ogutu, J. O. (2012a). A stage-wise approach for the analysis of multi-environment trials. *Biometrical Journal*, 54:844–860. https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201100219.

Piepho, H.-P., Nazir, M. F., Qamar, M., Rattu, A.-u.-R., ud Din, R., Hussain, M., Ahmad, G., e Subhan, F., Ahmad, J., Abdullah, Laghari, K. B., Vistro, I. A., Kakar, M. S., Sial, M. A., and Imtiaz, M. (2016). Stability analysis for a countrywide series of wheat trials in Pakistan. *Crop Science*, 56:2465–2475. http://dx.doi.org/10.2135/cropsci2015.12.0743.

Piepho, H.-P. and Ogutu, J. O. (2002). A simple mixed model for trend analysis in wildlife populations. *Journal of Agricultural, Biological, and Environmental Statistics*, 7:350. https://doi.org/10.1198/108571102366.

Piepho, H. P., Williams, E. R., and Madden, L. V. (2012b). The use of two-way linear mixed models in multitreatment meta-analysis. *Biometrics*, 68:1269–1277. https://doi.org/10.1111/j.1541-0420.2012.01786.x.

Poorter, H., Niinemets, Ü., Walter, A., Fiorani, F., and Schurr, U. (2010). A method to construct dose–response curves for a wide range of environmental factors and plant traits by means of a meta-analysis of phenotypic data. *Journal of Experimental Botany*, 61:2043–2055. https://doi.org/10.1093/jxb/erp358.

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*. https://doi.org/10.1007/s13253-020-00426-y.

R

Resende, R. T., Piepho, H.-P., Rosa, G. J. M., Silva-Junior, O. B., e Silva, F. F., de Resende, M. D. V., and Grattapaglia, D. (2020). Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. *Theoretical and Applied Genetics*. https://doi.org/10.1007/s00122-020-03684-z.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statist. Sci.*, 6:15–32. https://projecteuclid.org:443/euclid.ss/1177011926.

Romagosa, I. and Fox, P. N. (1993). Genotype × environment interaction and adaptation. In Hayward, M. D., Bosemark, N. O., and Romagosa, I., editors, *Plant breeding: principles and prospects*. Springer Science+Business Media, Dordrecht.

S

Schabenberger, O. and Pierce, F. J. (2001). *Contemporary statistical models for the plant and soil sciences*. CRC Press, Boca Raton.

Searle, S., Casella, G., and McCulloch, C. (1992). *Variance Components*. John Wiley & Sons, New York.

Singh, M., Ceccarelli, S., and Grando, S. (1999). Genotype x environment interaction of crossover type: detecting its presence and estimating the crossover point. *Theoretical and Applied Genetics*, 99:988–995. https://doi.org/10.1007/s001220051406.

Sleper, D. A. and Poehlman, J. M. (2006). *Breeding Field Crops*. Blackwell Publishing, Ames, fifth edition.

Smith, A., Cullis, B., and Gilmour, A. (2001). Applications: The analysis of crop variety evaluation data in Australia. *Australian & New Zealand Journal of Statistics*, 43:129–145. https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-842X.00163.

Smith, A. B., Cullis, B. R., and Thompson, R. (2005). The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *The Journal of Agricultural Science*, 143:449–462. https://doi.org/10.1017/S0021859605005587.

Sorensen, D. and Gianola, D. (2002). Bayesian updating. In *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*, pages 249–257. Springer, New York. https://doi.org/10.1007/0-387-22764-4$_5$.

Stroup, W. W. (2012). *Generalized linear mixed models: modern concepts, methods and applications*. CRC Press, Boca Raton.

T

Theobald, C. M., Talbot, M., and Nabugoomu, F. (2002). A Bayesian approach to regional and local-area prediction from crop variety trials. *Journal of Agricultural, Biological, and Environmental Statistics*, 7:403–419. https://doi.org/10.1198/108571102230.

V

van Eeuwijk, F. A., Bustos-Korts, D. V., and Malosetti, M. (2016). What should students in plant breeding know about the statistical aspects of genotype × environment interactions? *Crop Science*, 56:2119–2140. http://dx.doi.org/10.2135/cropsci2015.06.0375.

W

Welham, S. J., Gogel, B. J., Smith, A. B., Thompson, R., and Cullis, B. R. (2010). A comparison of analysis methods for late-stage variety evaluation trials. *Australian & New Zealand Journal of Statistics*, 52:125–149. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-842X.2010.00570.x.

Wolfinger, R. D. (1996). Heterogeneous variance–covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics*, 1:205–230. http://www.jstor.org/stable/1400366.

Summary

Multienvironment trials (MET) are carried out every year in different environmental conditions to evaluate a vast number of cultivars, i.e., yield, because different cultivars perform differently in various environmental conditions, known as genotype×environment interactions (GEI). MET aim to provide accurate information on cultivar performance so that a recommendation of which cultivar performs the best in a growers' field condition can be available. MET data is often analysed via mixed models, which allow the cultivar effect to be random. The random effect of cultivar enables genetic correlation to be exploited across zones and taking into account the trials' heterogeneity. A zone can be viewed as a larger target of population environments (TPE). The accuracy and precision of the cultivar predictions are crucial to be evaluated. The prediction accuracy can be evaluated via a cross-validation (CV) study, and the model selection can be done based on the lowest mean squared error prediction (MSEP). Also, since the trials' locations hardly coincide with growers' field, the precision of predictions needs to be evaluated via standard errors of predictions of cultivar values (SEPV) and standard errors of the predictions of pairwise differences of cultivar values (SEPD). Thus, in this thesis, the central objective is to assess the model performance and conduct model selection via a CV study for zone-based cultivar predictions.

In the second chapter, the performance between empirical best linear unbiased estimations (EBLUE) and empirical best linear unbiased predictions (EBLUP) for zone-based prediction was assessed. Different CV schemes were implemented between the single-year and multi-year datasets to mimic the practice. A complex variance-covariance such as factor-analytic (FA) was imposed to account for the heterogeneity of cultivar×zone effect. The MSEP showed that the EBLUP models outperformed the EBLUE models. The agroecological zone division was necessary since it improved the accuracy and was preferable to make cultivar recommendations. The FA variance-covariance structure did not improve the accuracy compared to the more straightforward covariance structure. Hence, the EBLUP model with a simple covariance structure may be selected for the single and multi-year datasets.

A more comprehensive assessment was done in the third chapter, which assessed single-stage analysis and stagewise analysis. The unweighted and three weighting methods were compared in the stagewise analysis, i.e. two diagonal approximation methods and the fully efficient method. The assessment was based on the MSEP

instead of Pearson's and Spearman's correlation coefficients since the correlation coefficients are often very close between the compared models. The MSEP showed that single-stage EBLUP analysis and the stagewise weighting EBLUP strategy were very similar. Thus, the loss of information due to diagonal approximation is minor. In fact, the MSEP showed a more apparent distinction between the single-stage and the stagewise weighting analyses with the unweighted EBLUE compared to the correlation coefficients. The simple compound symmetry variance-covariance structure was sufficient for the cultivar$\times$zone effect than the complex unstructured or the FA. With only three zones, the complex structure may not be necessary. The choice between the single-stage and stagewise weighting analysis, therefore, depends on the computational resources and the practicality of data handling.

In the fourth chapter, the accuracy and precision of the predictions were assessed for the new locations. The environmental covariates were combined with the EBLUP in the random coefficient (RC) models since the covariates provide more information for the new locations. The MSEP showed that the RC models were not the model with the smallest MSEP, but the RC models had the lowest SEPV and SEPD. Thus, the model selection can be done by joint consideration of the MSEP, SEPV, and SEPD. The models with EBLUE and covariate interaction effects performed poorly regarding the MSEP. The EBLUP models without random coefficient performed best, but the SEPV and SEPD were large, considered unreliable. The covariate scale and selection are essential to obtain a positive definite covariance matrix. Allowing covariance between slopes and between slopes and intercepts is crucial to maintaining the RC models' invariance feature. The RC framework is suitable to be implemented with GIS data and, therefore, provide an accurate and precise projection of cultivar performance for the new locations or environments.

To conclude, the EBLUP model for zoned-based predictions should be preferred to obtain the predictions and rankings closer to the true values and rankings, which are also confirmed by the MSEP. The stagewise weighting analysis can be recommended due to its practicality and its computational efficiency. Furthermore, projecting cultivar performances to the new locations should be done to provide more targeted information for growers. The available environmental covariates can be utilised to improve the predictions' accuracy and precision in the new locations in the RC model framework. Such information is certainly more valuable for growers and breeders than just providing means across a whole target population of environments.

Zusammenfassung

Multi-Umwelt-Versuche (MET) werden jedes Jahr unter verschiedenen Umweltbedingungen durchgeführt, um eine große Anzahl von Sorten im Hinblick auf den Ertrag zu bewerten, da verschiedene Sorten unter verschiedenen Umweltbedingungen unterschiedlich abschneiden, was als Genotyp-Umwelt-Interaktionen (GEI) bekannt ist. MET zielen darauf ab, genaue Informationen über die Leistung von Sorten zu liefern, so dass eine Empfehlung ausgesprochen werden kann, welche Sorte unter den Bedingungen des Anbaubetriebes am besten abschneidet. MET-Daten werden häufig über gemischte Modelle analysiert, die einen Zufallseffekt der Sorte zulassen. Der zufällige Effekt der Sorte ermöglicht es, die genetische Korrelation über Zonen hinweg und unter Berücksichtigung der Heterogenität der Versuche zu nutzen. Eine Zone kann als eine größere Zielpopulation von Umwelten (TPE) betrachtet werden. Die Genauigkeit und Präzision der Sortenvorhersagen sind entscheidend für die Bewertung der Sorten. Die Vorhersagegenauigkeit kann über eine Kreuzvalidierungsstudie (CV) bewertet werden, und die Modellauswahl kann auf der Grundlage der Vorhersage mit dem niedrigsten mittleren quadratischen Fehler (MSEP) erfolgen. Da die Versuchsstandorte außerdem kaum mit den Feldern der Landwirte übereinstimmen, muss die Genauigkeit der Vorhersagen über Standardfehler der Vorhersagen der Sortenwerte (SEPV) und Standardfehler der Vorhersagen der paarweisen Unterschiede der Sortenwerte (SEPD) bewertet werden. Daher ist das zentrale Ziel dieser Arbeit, die Modellleistung zu bewerten und eine Modellauswahl über eine CV-Studie für zonenbasierte Sortenvorhersagen durchzuführen.

Im zweiten Kapitel wurde die Leistung zwischen empirischen besten linearen unverzerrten Schätzungen (EBLUE) und empirischen besten linearen unverzerrten Vorhersagen (EBLUP) für zonenbasierte Vorhersagen bewertet. Es wurden verschiedene CV-Schemata zwischen den Einjahres- und Mehrjahresdatensätzen implementiert, um die Praxis nachzuahmen. Eine komplexe Varianz-Kovarianz, wie z. B. die faktoranalytische (FA), wurde eingeführt, um die Heterogenität des Sorten-Zonen-Effekts zu berücksichtigen. Die CV zeigte, dass die EBLUP-Modelle besser abschnitten als die EBLUE-Modelle. Die agrarökologische Zoneneinteilung war notwendig, da sie die Genauigkeit verbesserte und für die Abgabe von Sortenempfehlungen vorzuziehen war. Die FA-Varianz-Kovarianz-Struktur verbesserte die Genauigkeit nicht im Vergleich zu der einfacheren Kovarianz-Struktur. Daher

kann das EBLUP-Modell mit einer einfachen Kovarianzstruktur für die ein- und mehrjährigen Datensätze gewählt werden.

Eine umfassendere Bewertung wurde im dritten Kapitel durchgeführt, in dem die einstufige Analyse und die stufenweise Analyse bewertet wurden. Bei der stufenweisen Analyse wurden die ungewichtete Analyse und drei Gewichtungsmethoden verglichen, d. h. zwei diagonale Näherungsmethoden und die vollständig effiziente Methode. Die Bewertung erfolgte anhand des MSEP anstelle der Korrelationskoeffizienten von Pearson und Spearman, da die Korrelationskoeffizienten zwischen den verglichenen Modellen oft sehr nahe beieinander lagen. Der MSEP zeigte, dass die einstufige EBLUP-Analyse und die stufenweise gewichtete EBLUP-Strategie sehr ähnlich waren. Somit ist der Informationsverlust durch die diagonale Approximation sehr gering. Tatsächlich zeigte das MSEP eine deutlichere Unterscheidung zwischen der einstufigen und der stufenweise gewichteten Analyse mit dem ungewichteten EBLUE in Bezug auf die Korrelationskoeffizienten. Die einfache Compound Symmetry-Varianz-Kovarianz-Struktur war für den Sorten-Zonen-Effekt ausreichend im Vergleich zu den komplexeren unstrukturierten oder FA Modellen. Mit nur drei Zonen ist eine komplexe Struktur möglicherweise nicht notwendig. Die Wahl zwischen der einstufigen und der stufenweisen Gewichtungsanalyse hängt daher von den Rechenressourcen und der Praktikabilität der Datenverarbeitung ab.

Im vierten Kapitel wurden die Genauigkeit und Präzision der Vorhersagen für die neuen Standorte bewertet. Die Umweltkovariaten wurden mit dem EBLUP in den Zufalls-Koeffizienten (RC)-Modellen kombiniert, da die Kovariablen mehr Informationen für die neuen Standorte liefern. Der MSEP zeigte, dass die RC-Modelle nicht das Modell mit dem kleinsten MSEP waren, aber die RC-Modelle hatten die niedrigsten SEPV und SEPD. Somit kann die Modellauswahl durch die gemeinsame Betrachtung des MSEP, SEPV und SEPD erfolgen. Die Modelle mit EBLUE und Kovariaten-Interaktionseffekten schnitten in Bezug auf den MSEP schlecht ab. Die EBLUP-Modelle ohne Zufallskoeffizient schnitten am besten ab, aber der SEPV und SEPD waren groß und wurden als unzuverlässig angesehen. Die Kovariatenskala und -auswahl sind wesentlich, um eine positiv definite Kovarianzmatrix zu erhalten. Das Zulassen der Kovarianz zwischen Steigungen und zwischen Steigungen und Achsenabschnitten ist entscheidend für die Beibehaltung der Invarianz der RC-Modelle. Der RC-Rahmen ist geeignet, um mit GIS-Daten implementiert zu werden und somit eine genaue und präzise Projektion der Kultivierungsleistung für neue Standorte oder Umgebungen zu liefern.

Zusammenfassend lässt sich sagen, dass das EBLUP-Modell für zonenbasierte Vorhersagen zu bevorzugen ist, um Vorhersagen und Rangfolgen zu erhalten, die näher an den wahren Werten und Rangfolgen liegen, was auch durch das MSEP

bestätigt wird. Die stufenweise Gewichtungsanalyse kann aufgrund ihrer Praktikabilität und ihrer Berechnungseffizienz empfohlen werden. Darüber hinaus sollte eine Projektion der Sortenleistungen auf die neuen Standorte durchgeführt werden, um Landwirten und Landwirtinnen gezieltere Informationen zu liefern. Die verfügbaren Umweltkovariablen können genutzt werden, um die Genauigkeit und Präzision der Vorhersagen an den neuen Standorten im Rahmen des RC-Modells zu verbessern. Solche Informationen sind für Landwirte und Züchter sicherlich wertvoller als die Bereitstellung von Mittelwerten über eine ganze Zielpopulation von Umwelten.

Acknowledgements

Around three years and five months ago, I began this fascinating and fruitful journey. It was started in SLU, Uppsala, Sweden, where it felt like I would enter a Swedish coniferous forest. The first step was ambivalent, and mist since this journey will be challenging and seemed beyond my skills and knowledge. I said to myself, there was no way back. The step had to be taken, and the trip was on. On this trip, I had many unexpected episodes, i.e., more complicated and complex obstacles in the study and daily life than I expected. Still, the hurdles also made it more thrilling, more rewarding, and more magnificent.

Furthermore, the journey became more exciting when I had to enter the superb experience in Hohenheim. I had the immense privilege of learning and developing myself to comprehend an idea about science at heart in terms of research. The opportunity to live in Sweden and Germany and travelling, where I met so many fascinating people and visited many incredible places, were the most astonishing part of this experience. I found the uniqueness of each place and person that develop my personality. I will keep all of you in my heart. I will try to remember all of you with the great moments as a beautiful story written in my life.

First of all, I want to mention two important persons, **Johannes** and **Hans-Peter**, for the invaluable opportunity to work in two unique places, i.e., SLU and Hohenheim. You have believed in me and always been supportive. **Johannes**, you were not afraid to hire me to work on the SLU project. I remember it was not easy for you to have me in the project due to my study background, but you got it through. You gave me an incredible opportunity to learn many things in the SLU project to meet between academic research and the stakeholders. I enjoyed the moment when we had brainstorming in statistics and agriculture fields. I appreciated the freedom you gave me to develop ideas and also sometimes to fail. I admired your decision to relocate me to Hohenheim to continue my journey. That was the unexpected episode from the first step of this journey when you picked me up in Arlanda and took me to my apartment. I hope that we will stay connected and collaborate on many exciting projects in the future.

Hans-Peter, I was tremendously surprised when Johannes sent me the study plan in SLU; your name was written there. I remember the first time we met in the XVIth Meeting of the EUCARPIA Section Biometrics in Plant Breeding, in Wageningen, September 2015. I was thrilled when I met you at that time since I wanted to study

under your supervision. Your fast, thorough, and efficient working style influenced me to keep up with the pace. I believe this is a valuable lesson that I learned from you for my future career. I respected your efforts to have me in your unit to continue and finish my journey. It was "a dream come true" for me to have wandered in the beautiful campus, Hohenheim. Thank you for giving me the invaluable opportunity to attend the Summer School from TUM 2019. Also, the chance to participate in Biometrical courses in SLU Alnarp, where I had the opportunity to meet Jose Crossa from CIMMYT. I learned many new things in biometry/biostatistics subjects and humanity from you. You were very patient and constructive in guiding me when I was stuck. You gave me a small candle with light and let me walk with my way in the dark to find other candles until I found all the hidden candles. You always bring ideas when there is an obstacle. Every time there was a new related-publication, you always sent it to me, which broadened my perspective. At the same time, I like the unforeseen jokes in the emails you sent me. Working with you is one of the most memorable parts of this journey. I hope we can have many collaborations in our beloved genotype×environment interaction and biostatistics topics.

I also want to mention two other important persons, **Fred** and **Marcos**. **Fred**, thank you for the invaluable opportunity to conduct my master thesis in Biometris, WUR. You allowed me to pursue my passion in your chair group. It was the first step that leads to this point where I am now. Also, I thank you that with your support, I am at this stage. I am very honoured that you participated in my Doctoral journey as an external reviewer and examiner. **Marcos**, you had trained me in many aspects of statistical genetics when you supervised my major and minor theses. You always support me since I was searching for a PhD position until I finished this journey. I hope we can collaborate in the future as you are now in the breeding industry.

This doctoral trip would not have been as colourful if I had not had the invaluable opportunity to meet all my fellow colleagues in SLU. **Cigdem**, my officemate for two years in SLU, thanks for all the great moments. You have never given up asking me to walk with you from SLU to Uppsala Central Station, especially in the winter time. You were always there when I needed a friend to talk to and complaining. I know I was always annoying for you when I was bored. That being said, you were a great officemate. When I moved to Hohenheim, I was happy and sad at the same time. But at least, I visited you a couple of times, and I was delighted to attend your Licentiate defence. I hope we can meet again soon and I hope you will also arrive at the finish line of your study. It is also my pleasure to acknowledge **Claudia** and **Razaw**. Thank you **Claudia** for allowing me to attend conferences and taking courses. You allowed me to attend a memorable Biometrics conference in Rothamsted research station, in the UK, which is the oldest agricultural research

station. **Razaw**, your role like my "mother" when I was in SLU, and you helped me a lot when I was struggling with daily life issues. I remember the trip to Barcelona to attend the XXIXth International Biometrics Conference in 2018. We had some shopping schedule besides the conference. Thanks for the books that you gave me when I left SLU. I really appreciated that. I also thank to my colleagues in Crop Production Ecology, **James**, **Xiangyu**, **Hui**, **Yayuan**, and **Elsa**; it has been a great time to have fika and discussion with all of you. You all are awesome.

I would like to thank *Biostatistik Leute* in Hohenheim. First, it goes to **Paul**. We met first time in Barcelona, and then we had a fruitful discussion on the statistical and plant breeding stuff. Actually, we already met in the XVIth Meeting of the EUCARPIA Section Biometrics in Plant Breeding, in Wageningen, September 2015. But, at that time, the moment has not come. I am glad to have you in my second paper. *Danke schön* Paul, for your contribution and allowed me to stay at your place when I visited Hohenheim for the first time. It was surprising that you offered me to stay at your place for two weeks. I am happy we have a chat group with **Muhammad** to share new statistical stuff, e.g., R packages and statistical memes.

The next one is **Nha**. I met you in a variety testing meeting in Poland in 2018. Then, we met again in Barcelona, and I enjoyed our trip to the Camp Nou Stadium. You picked me up in Stuttgart airport when I moved to Hohenheim and became my officemate. Thanks very much for your kindness. I enjoyed our discussions and debates in the office. It is broadening my perspective. I also appreciated your efforts to do jogging with you around our beautiful campus garden. Vielen Dank **Jens**. You helped me by visiting the WG and talked to the landlord, who was *sprechen Sie nur Deutsch*. So, I could get a place to live in Hohenheim. I learned from you regarding some aspects of statistical analyses and SAS Macro.

A special thanks to "my mentor," **Steffen**, the person with whom I mostly discussed my research topic. I highly appreciated our discussion regarding ASReml-R, SAS IML, and the random coefficient models. Thanks for helping me for checking the *Zusammenfassung* for this dissertation. I hope I can visit you in Denmark someday. It is also the occasion to thank **Waqas** and **Joseph**. **Waqas**, I really respected our discussion in the MET and machine learning topics for your habilitation. I wish you success in your scientific career. To **Joseph**, thanks for our chit-chat of general and statistic issues. We were the only the staffs that went to the office on the weekends. I like the stories you told me and admire you as a prolific scientist. Many thanks also to **Frau Akyildiz**. You helped me a lot with all the complicated administration stuff even before I officially studied here until I finished.

My Hohenheim colleagues, **Man**, **Thanh**, and **Yaqin**. Thanks very much for the great friendships. I really enjoyed our memorable dinner. Without you all, I might

not know which places were good for Vietnamese and Chinese food in Stuttgart. **Man**, thanks for regular Dönerteller and groceries. You are often surprised with a chat to have food outside. I also enjoyed our barbeque party at your place on my birthday. I will remember our complaints regarding the doctoral studies and publications. I hope I can visit Thailand and meet you. Success for your academic career in Thailand. **Thanh**, thanks for always joining food outside and allowing us to eat the take-away food at your place. I met you when we visited Cannstatter Volksfest, in Wasen, Bad Cannstatt. It was a nice occasion to have beers and food in Wassen. **Yaqin**, thanks for the Chinese dinner at your place, regular lunch in Mensa, and spent the time to jog around campus with **Nha**. I hope you will succeed in your doctoral study and your future career.

For my Indonesian buddies, **Theo**, **Ivo**, **Ilena**, and **Lydia**, thanks for this amazing friendship. **Theo**, I remember in the early times I was in Stuttgart, we discussed the statistical aspects of your thesis. I was honoured to attend your Bachelor thesis defence, where you had **Prof. Melchinger** as your examiner. Thanks for make me got acquainted with **Ivo**, **Ilena**, and **Lydia**. **Ivo**, thanks for your nasi Hainan and great dinners at your place with **Ilena** and **Lydia**. If there were no pandemic, I believe that this occasion will be regular with **Ilena** and **Lydia**. For **Ilena** and **Lydia**, without both of you, it would be "empty" in Ivo's kitchen. Both of you were always cheerful and brought a positive atmosphere.

Special thanks to **Kyrylo** and **Linh**, who are my "soulmate" since we studied in Wageningen. Thanks for this incredible friendship. **Kyrylo**, I remember when we were sitting down in Radix building, exhausted with our master thesis, talking about our next plan; to do or not to PhD, and in the end, destiny brought us to Germany. You started in Tübingen, and I began in Uppsala. An unexpected moment when I moved to Hohenheim. I appreciated our moments since our master's and doctoral studies. I wish you all the best for your postdoc in Düsseldorf, and keep in touch. **Linh**, I met you when **Kyrylo** invited me to Ede. Since that time, "two GMO guys and one sö Djörmon girl" were set. Thanks for your signature soup and hot chocolate. I enjoyed our incredible moments in Berlin and Tübingen with **Kyrylo**. I hope we all can reunite soon in Germany.

Thanks to my colleagues in Sweden. Kantorsgatan fellas; **Manu**, **Michael**, and **Sanka**, thanks for accepting me as a "refugee" before moving to Flogsta. I wish you all success in your career. I would like to thank **Henry**, who accommodated me and my stuff before moving to Germany. I wish you all the best in your study. For *lae* **Azhar**, *kokoh* **Stanley**, and **Olin**, thanks for bringing joy in our late-night "discussion" every time I visit Uppsala and in the group chat. I hope we keep in touch and meet again. I thank **Fahry** for our friendship and as a colleague from the

same building. I appreciated our discussion in the MVM building. I remember the funny moment we lifted your new sofa from your front gate to your house. I hope you and **Resti** will succeed in your study. I thank **Puspita**, who always spared her time to go shopping and had fancy fika with me when I went to Stockholm at the weekend.

My special thank for **Bu Juju**, who is the students' mother in Uppsala. Thanks very much for taking care of us. I enjoyed having Indonesian food, midsommar at your place, and picking mushrooms in the wood near your place. I appreciated your hospitality when students gathered at your place in the winter and summertime. Thanks for bringing us to Fjällnora to have ice-skating on the freezing lake. I wish you good health and hope we can meet again soon.

Many thanks to my colleagues in Indonesia, **Herr Nasrullah**, **Herr Supriyanta**, **Herr Taryono**, **Herr Basunanda**, **Frau Ambarwati**, **Taufan**, and **Habib**. Thanks for your support and patience to hear my critics regarding the misuse of statistical analyses in your institution. I thank **Herr Nasrullah**, who always supported me and started brainstorming the Biometrical topics. Thanks for introducing me to quantitative genetics and biometry fields. I thank **Herr Taryono** and **Herr Basunanda**, who motivated me to study in Germany. Now, there will be Humboldt, Gießen, and Hohenheim alumni in our circle. I remember your advice to be a person that is "*Hochbegabt*." I am very grateful for the life advice and support from **Herr Supriyanta** dan **Frau Ambarwati**. Thank you for being there. Thank you **Taufan** and **Habib** for many discussion regarding agronomy and statistics issue. You brought me the opportunity to play with your data.

A special thanks to **Reren**, who became a close friend after we met in Stockholm. Thanks for being a chat friend with broad topics, from science to politics. I wish you success in your doctoral studies and your scientific career. I thank **Wita** for a great friendship and for supporting each other. Success for your study in Dietitian. I believe you will be a great dietitian in the future. A big thanks to my dearest friend **Ria**. I am grateful to have you as a great friend since we were in Wageningen. You are like family to me, always supportive and chatty. I do miss our shopping behaviour. I hope to meet you, **Eddy**, and **little Bianca**, either in Europe or in Bali.

I would like to express my gratitude to **my parents**. Thank you for letting me pursue my passion, whatever it takes. Without your support and patience to let me choose my path, I would not have achieved what I did. I appreciate how the way you teach me to be an independent and determined person. You always put the maximum effort into providing the best education. So, I am ready to face the hurdle and not give up when I struggle. I am grateful to have both of you open-minded. So I can be a person that open to other people with different cultures. I realised that not

many parents have it. Although I rarely visited you in Indonesia, you put your trust in me. It means a lot to finish this doctoral journey.

Many thanks to my Aunts, **Tante Nuning** and **Tante Triem** for their support and as the closest family member in Europe. **Tante Nuning**, you always gave advice and support me in my study and life matters. I really appreciate it. I hope I can visit you in the Netherlands soon. I thank **Tante Triem**, who always invited me to see her in Weinheim anytime. Thank you for spending Christmas at your place. I thank my cousin **Christian** as my big brother in Germany. I also thank my other cousin, **Claudia**. I am grateful to spend my time playing with your children, **Mira** and **Pascal**, when I and **Tante Triem** visited you and **York** in Karlsruhe.

To finalise these acknowledgements, the most extraordinary words go to **Clara**. It was unexpected to meet you when I was about to leave Sweden. You presented me with final beautiful memories in Uppsala, Stockholm, and Sigtuna. I find you are a very subtle and delicate person in my roller-coaster life. Moreover, you are very patient in dealing with my personality. With your attention, I can find new energy and confront the obstacles. You are always beside me in every step I take. I am deeply grateful to have you in my life journey. I am looking forward to the next page of this life story with you by my side.

"Und frische Nahrung, neues Blut
Saug ich aus freier Welt:
Wie ist Natur so hold und gut,
Die mich am Busen hält!

Die Welle wieget unsern Kahn
Im Rudertakt hinauf,
Und Berge, wolkig himmelan,
Begegnen unserm Lauf.

Aug, mein Aug, was sinkst du nieder?
Goldne Träume, kommt ihr wieder?
Weg, du Traum! so gold du bist:
Hier auch Lieb und Leben ist."

(Johann Wolfgang von Goethe, *Auf dem See*)

Affidavit

Annex 3

Declaration in lieu of an oath on independent work

according to Sec. 18(3) sentence 5 of the University of Hohenheim's Doctoral Regulations for the Faculties of Agricultural Sciences, Natural Sciences, and Business, Economics and Social Sciences

1. The dissertation submitted on the topic:

   Statistical methods for analysis of multienvironment trials in plant breeding: accuracy and precision

   is work done independently by me.

2. I only used the sources and aids listed and did not make use of any impermissible assistance from third parties. In particular, I marked all content taken word-for-word or paraphrased from other works.

3. I did not use the assistance of a commercial doctoral placement or advising agency.

4. I am aware of the importance of the declaration in lieu of oath and the criminal consequences of false or incomplete declarations in lieu of oath.

   I confirm that the declaration above is correct. I declare in lieu of oath that I have declared only the truth to the best of my knowledge and have not omitted anything.

Stuttgart, 17.02.2021
Place, Date

Signature

Curriculum vitae (Tabellarischer Lebenslauf)

Name:	Harimurti Buntaran
Date of birth:	July 2, 1990
Place of birth:	Pekalongan
Email:	harimurtibuntaran@outlook.com
Address:	Franziskaweg 20, 70599 Stuttgart

Professional Experience

02/2020 – 09/2021	Scientific Staff at the Biostatistics Unit, Institute of Crop Science University of Hohenheim, Stuttgart, Germany

Education

09/2017 – 09/2019	**Licentiate of Philosophy in Biometry** Swedish University of Agricultural Sciences Uppsala, Sweden
09/2014 – 08/2016	**Master of Science in Plant Sciences** Specialisation in Plant Breeding and Genetic Resources Wageningen University and Research Wageningen, The Netherlands
09/2008 – 04/2013	**Bachelor of Science in Agriculture** Major in Plant Breeding Universitas Gadjah Mada Yogyakarta, Indonesia
07/2005 – 06/2008	**High school**, St. Louis 1 Catholic Senior High School Surabaya, Indonesia

Stuttgart, 17.02.2021

List of publications

List of scientific publications included in this doctoral thesis:

1. Buntaran, H., Forkman, J., and Piepho, H.-P. 2021. Projecting results of zoned multi-environment trials to new locations using environmental covariates with random coefficient models: accuracy and precision. Theoretical and Applied Genetics 134:1513–1530. doi.org/10.1007/s00122-021-03786-2

2. Buntaran, H., Piepho, H.-P., Schmidt, P., Rydén, J., Halling, M., and Forkman, J. 2020. Cross-validation of stagewise mixed-model analysis of swedish variety trials with winter wheat and spring barley. Crop Science 60:2221–2240. doi: 10.2135/cropsci2018.10.0642

3. Buntaran, H., Piepho, H.-P., Hagman, J., and Forkman, J. 2019. A cross-validation of statistical models for zoned-based prediction in cultivar testing. Crop Science 59:1544–1553. doi: 10.1002/csc2.20177