Optimizing breeding schemes //

Manual / Field trials

Design and analysis of field trials - 1. Practical guidelines

- This manual provides guidelines on the design and statistical analysis of plant breeding field trails with a focus on breeders who do not have access to permanent biometric support.

- Published on 08/02/2022

Excellenceinbreeding.org/toolbox/tools/eib-breeding-scheme-optimization-manuals

CGIAR | Excellence in Breeding Platform

ExcellenceinBreeding.org

Design and analysis of field trials - 1. Practical guidelines

Author.

Christian R. Werner, Dr / c.werner@cgiar.org

Quantitative Genetics Specialist, Excellence in Breeding Platform

Editors.

Giovanny E. Covarrubias-Pazaran, Dr / EiB

Marlee Rose Labroo, Dr / EiB

Dorcus Gemenet, Dr / EiB

CGIAR | Excellence in Breeding Platform

Contents

Introduction

Field trials are used to evaluate genotypes for multiple target traits and to estimate or predict their genetic values. Measures of different types of genetic value are essential to select crossing parents for population improvement and candidate varieties for product development. Therefore, well-designed field trials and robust statistical analyses lay the foundation for high rates of genetic gain.

Considering the large variety of experimental designs and the complexity of the various analysis models, staying on top of good practice approaches can be challenging. Optimally, design and analysis is planned with a biometrician. Many programs, however, do not have access to full-time biometric support. Breeders then have to carry out the trial design and data analysis themselves - usually besides numerous other tasks - which often requires a pragmatic and time-efficient approach to conducting field trials.

This manual provides guidelines for the design and analysis of plant breeding trials for breeders who find themselves in this situation. While not necessarily aiming for the optimal or "best" approach, these guidelines ensure generally robust and accurate field trials while helping to avoid major mistakes. In particular, we focus on early-stage and late-stage testing and differentiate between across-location and within-location designs. We also give recommendations on check / replication strategies, on the analysis of (multi-environment) field trial data, and on modeling genotype-by-environment (GxE) interaction.

Additionally, a brief introduction to the theoretical background of field experimentation, multi-environmental trials and GxE interaction is given in Design and analysis of field trials – 2. Theoretical background. This introduction provides a basic understanding of the theory and concepts necessary to follow the guidelines presented below.

Recommendations for the design of plant breeding trials

Different breeding objectives require different trial designs

Based on the objective of plant breeding trials, we can distinguish between early-stage trials and late-stage trials.

Early-stage plant breeding trials usually test hundreds to thousands of genotypes with limited seed or planting material available in one or a few locations. Replication of some or all genotypes is often not possible. The main objective of early-stage testing is to select those genotypes which improve the population for multiple key traits. In a breeding program that is optimized for short generation intervals, early-stage trials serve two purposes:

- Selection of crossing parents which ensure high and sustainable genetic gain.
- Advancement of superior genotypes to late-stage trials.

Late-stage plant breeding trials usually test a relatively low number of advanced genotypes at multiple locations, referred to as *multi-environment trials (METs)*. The main purpose of late-stage testing is prediction of the true value of the tested genotypes within the *target population of environments (TPE)* for all traits included in the product profile. In a breeding program that is optimized for short generation intervals, late-stage trials focus on variety development[1]. Selection of candidate varieties is usually conducted relative to one or several benchmark varieties.

Early-stage and late-stage plant breeding trials usually differ with regard to their experimental designs. To manage the different numbers of genotypes, replicates and

[1] While selection of parents in late-stage trials (and sometimes even recycling of old varieties) is still common practice in many plant breeding programs, we advocate the adoption of strategies which strictly implement selection of new parents at early testing stages

locations in early-stage and late-stage trials, various **across-location (MET) designs and within-location designs** can be used (**Table 1**).

The across-location (MET) design of a trial is characterized by the number and type of test locations. The objective of testing in multiple locations is to obtain an accurate prediction of a genotype's true value within the entire TPE based on the sampled test environments.

The within-location design of a trial is characterized by the number of the tested genotypes, the number of (in)complete replicates, and the arrangement of genotypes within replicates. Within-location designs are used to obtain an accurate prediction of a genotype's true value within the tested location. In an MET, the predictions of a genotype's true value within tested locations are combined to obtain a prediction of the true value across the entire TPE.

Table 1. Classification of across-location and within-location designs.

Across-location design	Within-location design
• Single location • Multiple locations (multi-environment trial; MET)	• Fully replicated • Partially replicated (P-rep) • Unreplicated

In this manual, we will present guidelines for the design and analysis of early-stage and late-stage breeding trials. While there is no clear-cut differentiation between early-stage and late-stage trials, we consider this classification helpful to provide breeders with general recommendations on how to design robust and accurate field trials at different stages of a breeding program. We assume that most breeders are somewhat familiar with this classification and will be able to translate the concept of early-stage and late-stage testing into their breeding programs.

Across-location design of plant breeding trials

Multi-environment testing to increase the repeatability of a genotype

The objective of testing in multiple locations is to obtain an accurate prediction of a genotype's true value within the entire TPE. To achieve this, **genotypes should be tested in as many locations as possible**.

From the equation of the repeatability (or broad-sense heritability; **Equation 1**), we learn that increasing the number of test locations is always advantageous over increasing the number of replicates within locations (assuming a well-defined TPE). In particular, we see that:

- Increasing the number of locations (*nEnv*) increases the repeatability (H^2) through reduction of the GxE interaction variance (σ^2_{GxE}) and the residual variance (σ^2_e).
- Increasing the number of replicates within locations (*nRep*) increases the repeatability (H^2) only through reduction of the residual variance (σ^2_e).

$$H^2 = \frac{\sigma^2_g}{\sigma^2_p} = \frac{\sigma^2_g}{\sigma^2_g + \frac{\sigma^2_{gxe}}{nEnv} + \frac{\sigma^2_e}{nEnv * nRep}}$$

Equation 1

Where:

H^2 is the repeatability (broad-sense heritability).

σ^2_g is the genetic variance.

σ^2_p is the phenotypic variance.

σ^2_{gxe} is the genotype-by-environment (GxE) interaction variance.

σ^2_e is the residual variance.

nEnv is the number of environments (locations or year x location combinations).

nRep is the number of replicates within environments.

An increase in repeatability is more or less equivalent to an increase in selection accuracy. The increase in repeatability resulting from testing in more locations, however, is not linear and will quickly approach a plateau in a well-defined TPE, as shown in **Figure 1**.

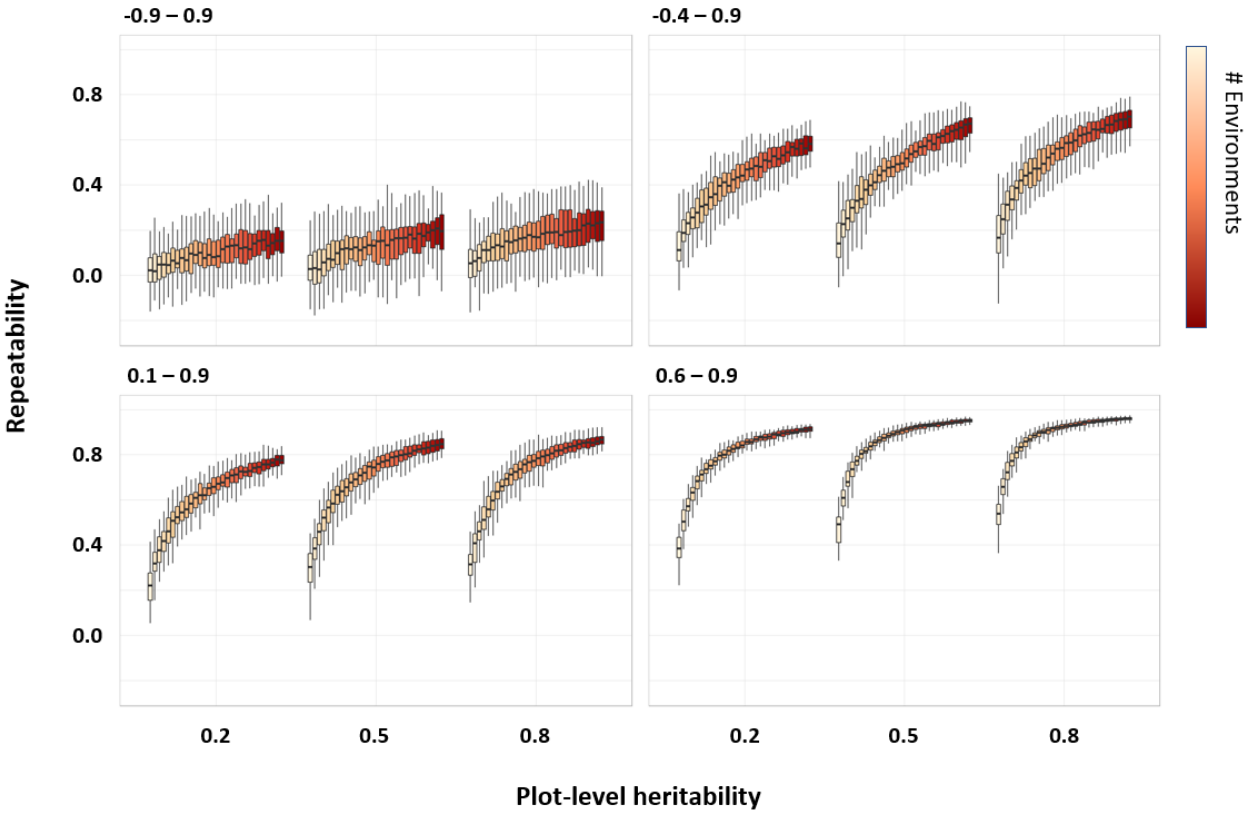

Figure 1. Effect of the number of test locations on repeatability (i.e., selection accuracy). A single quantitative trait was simulated for 300 genotypes. The genotypes were tested at 1 to 30 locations in two replications. Three different plot-level heritabilities and four different ranges of genetic correlations (shown top left of each figure) between the test locations were compared. Each scenario (number of environments x genetic correlation range x plot-level heritability) was simulated 50 times and observed repeatabilities are represented as boxplots for each of the 30 scenarios.

For a TPE with moderate to high genetic correlations between locations, we conclude:

- If the number of test locations is low (e.g., 5 or lower), testing in an additional location can substantially increase repeatability.
- If the number of locations is already high (e.g., 15 or higher), the increase in repeatability will be marginal, and this might not justify the costs associated with testing in an additional location.

For a TPE that contains locations with moderate to high negative correlations, we conclude:

- Increasing the number of test locations results only in a slow increase in repeatability. High repeatabilities are hard or impossible to achieve.
- This might indicate that the TPE is not well-defined (too broad), and selection based on a genotype's mean across the entire TPE is not an optimal strategy to identify candidate varieties.

General guidelines for efficient across-location trial designs

In general, we can say that increasing the number of locations is advantageous over increasing the number of replicates per location. This applies to early-stage as well as late-stage testing. However, depending on your current testing strategy, different approaches may be necessary to optimize your across-location designs:

- If replicates can be freely reallocated, increase the number of locations and reduce the number of replicates within locations. Especially in early-stage trials, testing in multiple locations is a key prerequisite for an accurate selection of parents.
- However, only reduce the number of replicates within locations if these replicates can be reallocated to additional test locations. If the number of locations cannot be increased, testing multiple replications within locations still increases repeatability[2].

[2] Although testing multiple replicates per location still increases repeatability, testing more than 3-4 replicates per location rarely pays off

- If the number of locations is low (e.g., up to 5), try to add more test locations. For small breeding programs with very few locations in particular, adding just one additional location can substantially increase the repeatability.
- If the number of locations is already high (e.g., 10 or more), the increase in repeatability and genetic gain will be limited and resources might be better invested elsewhere.

"Traditional" across-location designs assume that each genotype is tested at least once at each location. In that case, the number of test locations is restricted by the total number of replicates available per genotype. Sparse testing designs, however, allow us to test our genotypes at more locations than replicates per genotype are available.

Sparse testing: a special case of multi-environment testing

Sparse testing designs are multi-location (MET) designs in which not all varieties are tested at all locations (**Figure 2**). Sparse testing allows for the number of test locations to be higher than the number of replicates available per genotype. Therefore, it can be used to obtain a broader TPE sample and increase repeatability. We can make use of sparse testing designs in early-stage as well as late-stage testing. However, while late-stage trials usually test already at many locations, sparse testing is particularly useful in early-stage trials, where seed or planting material is limited. Likewise, it is ideal for participatory breeding, where each farmer can only test a small subset of the candidate varieties.

An efficient sparse testing design, however, can be complex. It requires that genotypes are somewhat evenly spread across locations and that all genotypes are connected using a genomic relationship matrix to predict their value in the locations they were not tested in[3].

[3] We strongly advise against using a pedigree matrix instead of a genomic relationship matrix for sparse testing, since pedigree relationships contain no information on the Mendelian sampling term.

This has the following reason: if the genotype is the unit of evaluation, to obtain information on its performance across multiple locations, the genotype itself has to be tested at all locations. However, if the marker allele becomes the unit of evaluation, most or all marker alleles (captured in haplotypes) will be tested at all locations, even if the individual genotypes are not. The mean performance of a genotype across all locations can then be predicted based on its marker genotype (its marker-based relationship with all other genotypes).

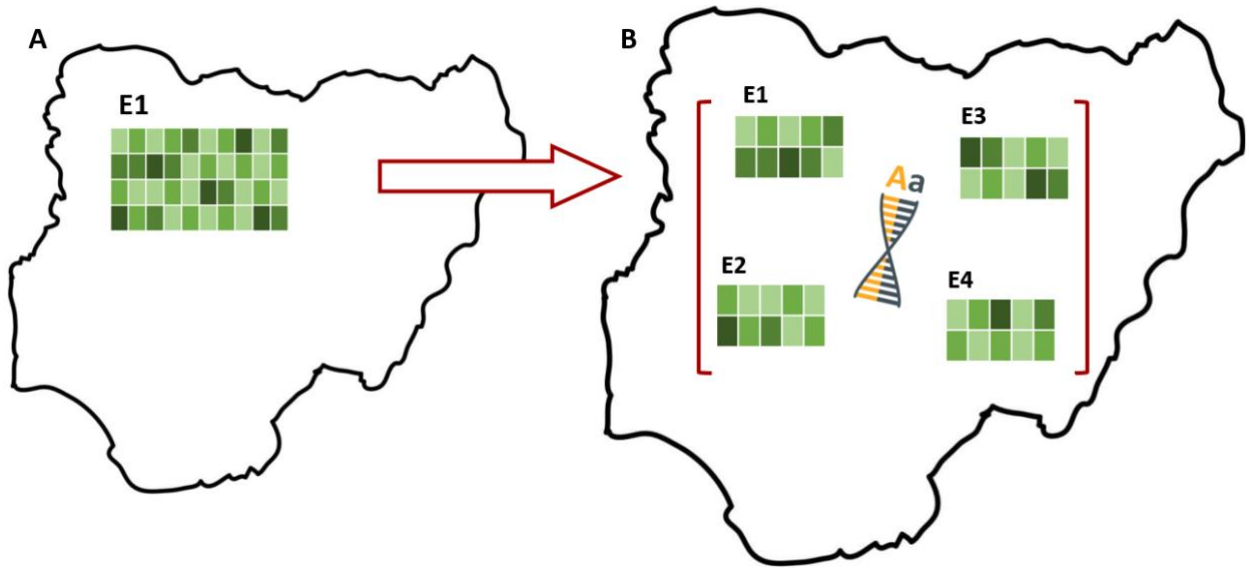

Figure 2. Transition from an early-stage single-location augmented design (A) to a multi-location sparse testing design (B). Although genotypes are unreplicated, sparse testing allows testing the breeding germplasm in multiple locations. Connectivity between unreplicated genotypes in different locations is achieved using a genomic relationship matrix. Therefore, the marker allele (haplotype) becomes the unit of evaluation rather than the individual plant genotype.

Sparse testing designs applied in early-stage testing provide a strong basis for accurate selection and reduction of the generation interval in breeding programs with genomic prediction. If you consider implementing sparse testing, we highly recommend that appropriate designs are planned in consultation with a biometrician.

Within-location design of plant breeding trials

Replication within locations in early-stage and late-stage trials

An overwhelming number of within-location (single-site) designs exist to optimize field trials for different numbers of genotypes and replicates, assumptions about field heterogeneity, different (incomplete) block sizes and arrangements, as well as different research questions. This makes it difficult to choose a suitable design at each stage of a breeding program. While a field trial is optimally designed in consultation with a biometrician, a rather pragmatic approach is required in programs where permanent biometric support is not available.

Based on the replication level, within-location designs can be classified into three types:

- Fully replicated
- Partially replicated (p-rep)
- Unreplicated

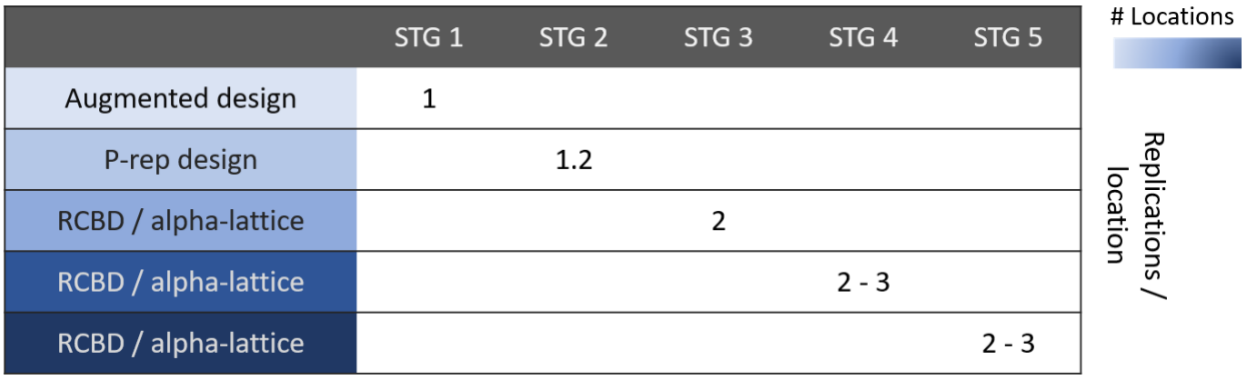

Figure 3. A simplified example of a conventional testing strategy for a plant breeding program. At early stages (STG1 & STG2), a high number of genotypes is tested in unreplicated or partially replicated designs at a single or a few locations. At later stages (STG3 – STG5), a reduced number of genotypes is tested in 2-3 full replications at multiple locations using a randomized complete block design (RCBD) or an alpha-lattice.

Fully replicated (multi-location) designs, such as randomized complete block designs (RCBD) or incomplete block designs are common at late-stage trials. Partially replicated and unreplicated designs in one or a few locations are common at early-stage trials (**Figure 3**). It should be noted, however, that this categorization is not written in stone. Partially replicated or unreplicated designs, for example, can also be efficiently applied in late-stage testing.

General guidelines for efficient within-location trial designs

Often, there is a competition between the number of replicates tested per location (within-location design) and the number of test locations (across-location design). As shown above, testing in more locations within the TPE is in general advantageous over testing more replicates within fewer locations. Therefore, **a good within-location design is characterized by a replication level that allows maximization of the number of test locations**.

Guidelines for early-stage within-location designs

The choice of an efficient within-location design for early-stage breeding trials depends on the number of replicates available per genotype. You may use:

- **Augmented design** with 5-10% check plots if the genotypes cannot be replicated.
- **Partially replicated design (p-rep)** with a 5-10% replication level if some genotypes can be replicated.
- **Multi-location augmented or p-rep design** if all genotypes can be replicated. If a p-rep design is used, different genotypes may be replicated at all locations (if possible).
- **A sparse augmented design or a sparse p-rep design** to further increase the number of test locations. This also allows multi-location testing even if the individual genotypes cannot be replicated. Note that sparse testing requires the utilization of a genomic relationship matrix (see sparse testing section).

For augmented and p-rep designs, randomization of (partially replicated) genotypes and/or check genotypes is mostly sufficient to ensure a robust design. While more sophisticated designs can be more efficient, the differences are often small and of secondary importance in practice (Hoefler et al. 2020). Optimized designs, however, often come at no extra cost and therefore it is advisable to make use of them, if possible. Modern software solutions which generate optimal or near-optimal model-based designs include:

- **OD** (Butler 2019): R package (free).
- **DiGGer** (Coombes 2020): R package (free).
- **CycDesigN** (VSN): standalone; implemented in the Breeding Management System (BMS) software suite (licensed).

Examples on how to generate experimental designs using OD and DiGGer will be made available here.

Guidelines for late-stage within-location designs

In the late stages of a plant breeding program, resources often allow for fully replicated testing in many locations. If all genotypes can be tested in at least two replicates at each location, you may use a common, established design of choice:

- Depending on the number of genotypes and the expected field heterogeneity, a randomized complete block design (RCBD) or an incomplete block design such as an alpha-lattice are robust within-location designs.
- Modern software solutions which generate optimal or near-optimal model-based designs, such as OD, DiGGer, and CycDesigN, might further improve the efficiency of a fully replicated design.
- No additional check genotypes to control for spatial variation are necessary.
- Testing more than 3 or 4 replications per location hardly pays off and therefore should be avoided.

- If the number of test locations is below 8-10[4], the number of replicates per location may be reduced to enable testing in more locations. Sparse testing can be used if the total number of replicates per genotype doesn't allow testing all genotypes in all environments (or if testing resources are limited).

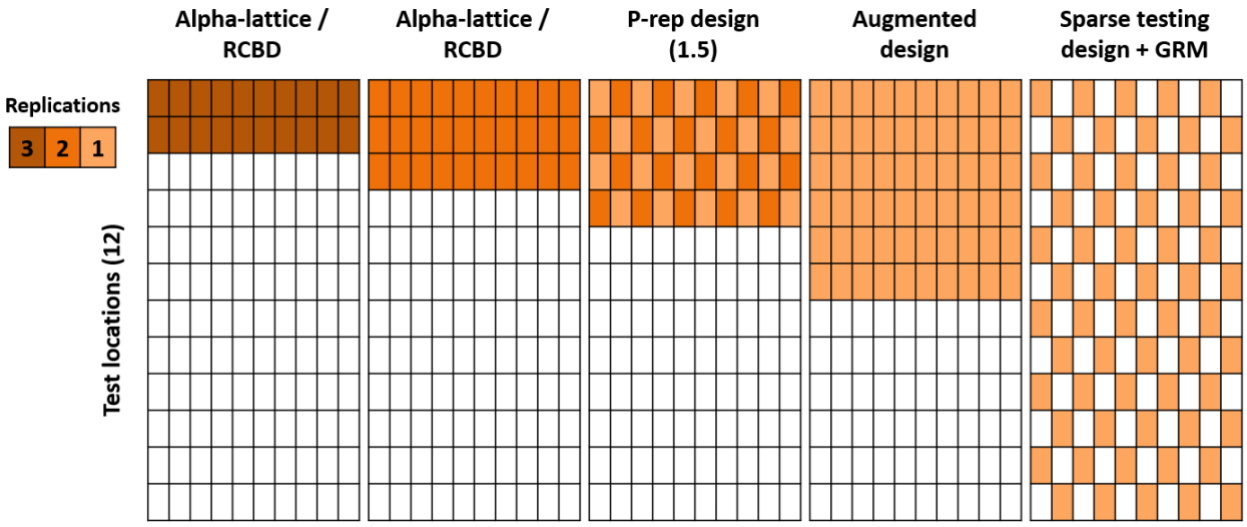

Figure 4. Comparison of different multi-environment testing strategies concerning the number of test locations and the number of replications per location. From left to right the number of replications per location (color of the squares) is gradually reduced while the number of test locations (number of rows) is increased. When a sparse testing strategy is used, a genomic relationship matrix is required to ensure connectivity between genotypes not tested in the same environment.

[4] Consider this a ballpark figure. As shown in **Figure 1**, the repeatability (and hence selection accuracy) approaches a plateau with an increasing number of environments. This trend depends on the genetic correlations between the tested environments (TPE). If the TPE is well-defined, less than 8-10 environments may be sufficient. While the optimal balance between repeatability and testing resources depends on the situation, 8-10 environments may be a reasonable reference number for many CGIAR breeding programs.

How many checks are needed?

Check genotypes fulfill several purposes in plant breeding trials. Breeders may use: 1) several benchmark varieties and local check genotypes to guide variety development; 2) dynamically stable check genotypes to connect trials across environments and measure genetic gain over time; and 3) replicated check genotypes to obtain an estimate of the residual variance.

Here, we will only provide guidelines on check strategies to obtain an estimate of the residual variance[5]. For a detailed description of how to optimize your check strategy, see the EiB manual on Estimating surrogates of genetic value.

- **Unreplicated designs** (e.g., augmented design): use 5-10% of the plots for replicated check genotypes. Checks can be randomly assigned to plots. A (model-based) design can be used to optimize check allocation. For example, an early-stage trial consisting of 1000 plots may test 950 unreplicated genotypes and allocate 50 plots to 3-4 check genotypes. An advanced trial consisting of 50 plots may test 40 unreplicated genotypes and allocate 10 plots to 4-5 check genotypes.
- **P-rep designs:** use 5-10% of the plots for partially replicated genotypes. These genotypes can be randomly assigned to plots. A (model-based) design can be used to further check allocation. Additional check genotypes are not necessary to obtain an estimate of the residual. In multi-location p-rep designs, replicate different genotypes at different locations (if possible). For example, a 1.1 p-rep early-stage trial consisting of 1000 plots may test 800 unreplicated genotypes (800 plots) and 100 replicated genotypes (200 plots). A 1.4 p-rep advanced consisting of 50 plots may test 30 unreplicated genotypes (30 plots) and 10 replicated genotypes (20 plots).
- **Fully replicated designs (e.g., RCBD, alpha-lattice):** the residual variance is estimated based on all replicated genotypes. Additional check genotypes are not necessary.

[5] These guidelines assume that a linear mixed model is used for trial data analysis (see below) and do not apply if a traditional ANOVA-based model is used.

Field trial analysis and evaluation of genotype-by-environment (GxE) interaction

Linear mixed models: the gold standard for the analysis of field trials

All guidelines presented in this manual assume that linear mixed models are used for field trial analysis. Linear mixed models enable flexible and accurate analysis of complex experimental data sets which makes them the gold standard of statistical analysis in plant breeding.

Compared to the traditional fixed linear (ANOVA) model, they offer multiple advantages, such as:

- Joint estimation of fixed effects and prediction of random effects.
- Accommodation of unbalanced (incomplete) data (i.e., not all genotypes are tested at all locations and/or years, or genotypes are tested at different replication levels)
- Flexible variance and covariance structures to model genetic effects, non-genetic effects, and genotype-by-environment interaction effects.
- Correlations (relationship) between genotypes can be exploited using a pedigree or genomic relationship matrix.

Common free and licensed Linear Mixed Model software solutions include:

- **ASReml-SA** (Gilmour et al. 2021): standalone (licensed).
- **ASReml-R** (Butler 2020): R package (licensed).
- **sommer** (Covarrubias-Pazaran 2018): R package (free and open source).
- **lme4** (Bates et al. 2015): R package (free and open source).
- **SAS** (SAS Institute Inc.): standalone (licensed).

General guidelines for the analysis of field trial data

While finding the model with the best fit always depends on the dataset, a few general guidelines lay the foundation for a good and robust modeling approach. We suggest, however, that the design and the analysis of your field trials are planned and conducted with a biometrician, since a complete overview cannot be given here.

Guidelines for the analysis of single-location field trial data

- Variety candidates (unreplicated and replicated) are modeled as random effects. This is consistent with the aim of selection in early-stage trials (A. B. Smith, Cullis, and Thompson 2005) and allows for utilization of a pedigree or genomic relationship matrix.
- Check genotypes are modeled as fixed effects. They are not part of the breeding population and should be excluded when the genetic variance is estimated.
- Blocking factors (e.g., blocks, rows, and columns) are modeled as random effects (unless the number of factor levels is extremely low, in which case modeling as a fixed effect may be preferable). Make sure you keep track of row and range (column) information!
- A spatial term often enables improved modeling of residual effects by assuming a correlation structure of the residual across the experimental field. The two-dimensional autoregressive model (AR1:AR1; Gilmour et al. 1997) has proven to be an efficient and robust standard procedure.
- Including the "nugget" (random error term) additional to a correlated residual term (e.g., AR1:AR1) might also improve model fit.
- Strictly speaking, it is advisable to include blocking factors, correlation structures, and a nugget only if they improve model fit. In practice, however, they often improve the model and rarely result in a substantially reduced model fit. Nevertheless, we advocate a comparison of different models (if possible).

A general note on the analysis of multi-location (MET) field trial data and modelling genotype-by-environment interaction.

The analysis of multi-environment trials is more complex than the analysis of a single-location trial. While the same guidelines apply to the analysis of each environment individually (year x location combination), combining the data from multiple environments using a single-step or two-step approach requires a suitable GxE interaction model. This cannot be covered here in detail, and only a very brief overview of the most common GxE interaction models will be given. For a more detailed introduction on modeling GxE interaction effects see, for example, van Eeuwijk et al. (2001), Smith, Cullis, and Thompson (2005), and Smith and Cullis (2018).

Commonly used GxE interaction models include:

- Compound symmetry (Fig. 5A)
- Unstructured model (Fig. 5B)
- Factor analytic (FA) model

A

$$\begin{bmatrix} \sigma^2_{g+gxe} & \sigma^2_g & \sigma^2_g & \sigma^2_g \\ & \sigma^2_{g+gxe} & \sigma^2_g & \sigma^2_g \\ & & \sigma^2_{g+gxe} & \sigma^2_g \\ & & & \sigma^2_{g+gxe} \end{bmatrix}$$

B

$$\begin{bmatrix} \sigma^2_{g_{11}} & \sigma^2_{g_{12}} & \sigma^2_{g_{13}} & \sigma^2_{g_{14}} \\ & \sigma^2_{g_{22}} & \sigma^2_{g_{23}} & \sigma^2_{g_{24}} \\ & & \sigma^2_{g_{33}} & \sigma^2_{g_{34}} \\ & & & \sigma^2_{g_{44}} \end{bmatrix}$$

Figure 5. Variance-covariance structure of the compound symmetry model (A) and the unstructured model (B), two commonly used structures to model genotype-by-environment (GxE) interaction in multi-environment trials. GxE interaction is conceptualized to be the result of an imperfect genetic correlation between environments. Models differ in the number of variances and covariances to be estimated. While models with more variance components allow more realistic modeling, they are also computationally more demanding, which can result in convergence issues.

The **compound symmetry model** allows for fast and efficient analysis of MET data if we can assume a uniform genetic variance and GxE interaction variance within all environments, as well as a uniform genetic covariance between all environments, respectively. While the assumption of homogeneous variances is in most cases rather unrealistic, the compound symmetry model might be useful as a baseline model and to obtain a first, rough idea of how much GxE interaction variance there is. More complex models, however, are in general better suited for MET data.

The **unstructured model** fits a unique genetic variance for each environment and a unique genetic covariance between each combination of two environments. Therefore, it provides a realistic representation of trials conducted in multiple environments and also allows to predict environment-specific performance. However, when the number of environments is high, solving an unstructured model becomes computationally challenging and may produce many singularities due to the high number of variance components to be estimated (i.e., it may not converge).

Factor analytic models provide a good balance between modeling complexity, efficiency, and model validity. Factor analytic models allow approximating the unstructured model in an astute and creative way which requires substantially fewer variance components to be estimated. As a result, factor analytic models have become very popular and are generally a good choice. Their interpretation, on the other hand, is a bit more complex and requires some experience.

Unfortunately, there is no “one size fits all” modelling approach to streamline the analysis of field trial data. Finding the “best” approach requires a comparison of different models to identify a model that combines a good model fit and computational efficiency. Furthermore, the objective of the field trial does also play a role. Predicting a genotype’s performance in various environments will require a different model than predicting a genotype’s mean performance across the entire TPE. Therefore, the analysis of multi-environment trials is best done by an experienced biometrician.

The genomic relationship matrix: a game-changer

Whole-genome genotyping and the potential to calculate marker-based relationships has fundamentally changed plant breeding. Although mainly seen as a prerequisite for genomic selection strategies, a genomic relationship matrix[6] can also be used to improve various components of a plant breeding program, field trials, and their analysis included.

When we connect our tested breeding material based on marker information, the marker allele becomes the unit of evaluation rather than the individual plant genotype. This has a important implication for the genetic evaluation: if two individuals share a large fraction of marker alleles, they are considered highly related. **Related individuals can be considered as partial replications**, and a high genomic relationship means that we can borrow a lot of extra information from a relative. This is a huge advantage, especially in early stages of a breeding program where individual plant genotypes are often unreplicated. When we introduce a genomic relationship matrix, information from many more or less closely related individuals across the entire field trial can be exploited.

In the case of a single population tested in a single environment, this can already result in a substantial increase in repeatability and selection accuracy. However, if we combine information from different crossing generations tested across multiple locations and years, we can build a training population that provides the basis for an efficient genomic selection strategy. Increased selection accuracies and reduced generation intervals are only two of the many advantages arising from this.

For most species, mid-density genotyping platforms with a few thousand markers are already available and become more and more affordable. Exploiting genomic information can be a game-changer and key driver of increased genetic gains in your breeding program.

[6] Marker information can be exploited either on the individual marker level (RR-BLUP; prediction of individual marker allele effects) or using a marker-based relationship (G-BLUP; all marker alleles are used to predict the relationship between individuals). Since both models are equivalent (under certain conditions), they are used somewhat interchangeably here.

References

Bates D, Mächler M, Bolker B, Walker S (2015) "Fitting Linear Mixed-Effects Models Using Lme4." *Journal of Statistical Software* 67 (1). https://doi.org/10.18637/jss.v067.i01.

Butler D (2019) *Od: Generate Optimal Experimental Designs. R Package Version 2.0.0.*

Butler D (2020) *Asreml: Fits the Linear Mixed Model. R Package Version 4.1.0.143.* www.vsni.co.uk.

Coombes N (2020) *DiGGer: DiGGer Design Generator under Correlation and Blocking. R Package Version 1.0.5.* http://nswdpibiom.org/austatgen/software.

Covarrubias-Pazaran GE (2018) "Software Update: Moving the R Package *Sommer* to Multivariate Mixed Models for Genome-Assisted Prediction." Preprint. Genetics. https://doi.org/10.1101/354639.

Eeuwijk FA van, Cooper M, DeLacy IH, Ceccarelli S, Grando S (2001) "Some Vocabulary and Grammar for the Analysis of Multi-Environment Trials, as Applied to the Analysis of FPB and PPB Trials." *Euphytica* 122 (3): 477–90. https://doi.org/10.1023/A:1017591407285.

Gilmour AR, Cullis BR, Verbyla AP (1997) "Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments." *Journal of Agricultural, Biological, and Environmental Statistics* 2 (3): 269. https://doi.org/10.2307/1400446.

Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R (2021) *ASReml User Guide Release 4.2 Functional Specification, VSN International Ltd, Hemel Hempstead, HP2 4TP, UK, www.vsni.co.uk*.

Hoefler R, González-Barrios P, Bhatta M, Nunes JAR, Berro I, Nalin RS, Borges A, et al (2020) "Do Spatial Designs Outperform Classic Experimental Designs?" *Journal of Agricultural, Biological and Environmental Statistics* 25 (4): 523–52. https://doi.org/10.1007/s13253-020-00406-2.

SAS Institute Inc., Cary, NC, USA. "*SAS*".

Smith AB, Cullis BR, 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 (6): 449–62. https://doi.org/10.1017/S0021859605005587.

Smith AB, Cullis BR (2018) “Plant Breeding Selection Tools Built on Factor Analytic Mixed Models for Multi-Environment Trial Data.” *Euphytica* 214 (8): 143. https://doi.org/10.1007/s10681-018-2220-5.

VSN International Ltd. "*CycDesigN*".