Optimizing breeding schemes //

## Manual / Field trials

# Design and analysis of field trials - 2. Theoretical background

- This manual provides a brief introduction to the theoretical background of field experimentation, multi-environmental trials and GxE interaction.
- 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 – 2. Theoretical background

**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.

The design and the statistical analysis of trial data require a basic understanding of the key concepts and models commonly used in field experimentation. In particular, this involves understanding genetic value, the "environment", genotype-by-environment (GxE) interaction, and their effects on the relationship of plant phenotype to plant genotype. To obtain accurate estimates of genetic and non-genetic effects, basic principles of field experimentation (replication, randomization, and blocking) must be efficiently applied when testing either in a single or multiple environments, depending on the testing stage.

This manual provides a brief introduction to the theoretical background of field experimentation and trial analysis. Good-practice recommendations are presented in Design and analysis of field trials – 1. Practical guidelines.

# Field trials: our key to obtaining accurate estimates of genetic value from a phenotype

## Why do we need field trials?

In plant breeding programs, field trials are used to test genotypes for multiple traits and estimate (or predict) their genetic values[1]. We use those measures of genetic value to:

- Select crossing parents for population improvement; and
- Identify candidate varieties for product development.

Therefore, accurate estimates of genetic values are critical to drive genetic gain and ensure a high turnover of improved varieties. Field trials, however, generate phenotypic values, which are a result of genetic and non-genetic effects. Efficient field trial designs and their analysis are key components of plant breeding programs to divide phenotypic values into components attributable to the genotype and to non-genetic effects. Hence, field trials lay the foundation for accurate and reliable estimates of genetic value.

## The phenotype as a function of genotype and environment

Genetic value is determined by the entire set of expressed genes in an individual and their interactions (additive and non-additive genetic effects). Environmental effects cause the phenotypic value to deviate from the genetic value.

[1] We will use the term "genetic value" rather than "genotypic value". Although both are common in the literature, "genotypic value" may also refer to the value of a single locus, and we intend to avoid this confusion. For the sake of simplicity, the term "genetic value" will be used synonymously for the total genetic value and any of its components used to select genotypes (e.g., the breeding value).

As stated by Falconer and Mackay (1995):

> We may think of the genotype conferring a certain value on the individual and the environment the genotype is grown in causing a deviation of the observed phenotypic value from the underlying genetic value in one direction or the other. Hence, generally speaking, the effect of the environment is a source of error which reduced the precision of the estimated genetic value when based on the phenotypic value.

This relationship can be expressed in a simple model, where the phenotypic value (P) is described as a linear function of the genetic value (G) and the environment (E).

$$P = G + E$$

**Equation 1**

On an individual genotype's basis, the genetic value can also be expressed as a deviation from the population mean for a trait:

$$y_{ij} = \mu + g_i + e_{ij}$$

**Equation 2**

Where:

$y_{ij}$ is the $j^{th}$ observation on genotype i.

$\mu$ is the population mean.

$g_i$ is the genetic value of genotype i expressed as a deviation from the population mean.

$e_{ij}$ is the effect of the environment for observation j on genotype i.

## What is "the environment"?

### Environmental factors and their effect on the phenotype

In **models (2) and (3)**, our interpretation of a single environmental effect as a source of error stems from the fact that we have no information on what exactly causes the deviation of the phenotype from the genetic value. The environment, however, is not a single factor but involves a plethora of biotic and abiotic factors that affect the phenotypic value:

- Physical and chemical soil attributes, including those induced by soil tillage and crop cultivation practice.
- Climatic factors, such as precipitation, temperature, and sunlight.
- Biological organisms, such as weeds, pests, and pathogens.
- Measurement error.

The total of these factors and their interactions result in an environment that is highly specific to the combination of location and season a genotype is grown in. We refer to the variation due to the effect of the location as spatial variation, and the variation due to the effect of the season as seasonal variation (**Figure 1**).

### Spatial variation

Spatial variation is caused by the variability of biotic and abiotic factors at different geographical locations. We usually differentiate between spatial variation within experimental fields and spatial variation across experimental fields, as differences between environmental conditions tend to increase with physical distance. Spatial variation within experimental fields is primarily caused by soil heterogeneity and can complicate the comparison of genotypes. Spatial variation across experimental fields can result in substantial genotype-by-environment (GxE) interaction and affects how we define our target population of environments (TPE).

### Seasonal variation

Seasonal (or temporal) variation refers to the variability of biotic and abiotic factors over time at the same geographical location. Key factors causing this type of variation are usually (but not limited to) climatic factors, such as temperature, rainfall, and hours of sunlight. Seasonal variation may cause a genotype to produce significantly different phenotypes when grown in different years at the same geographical location.

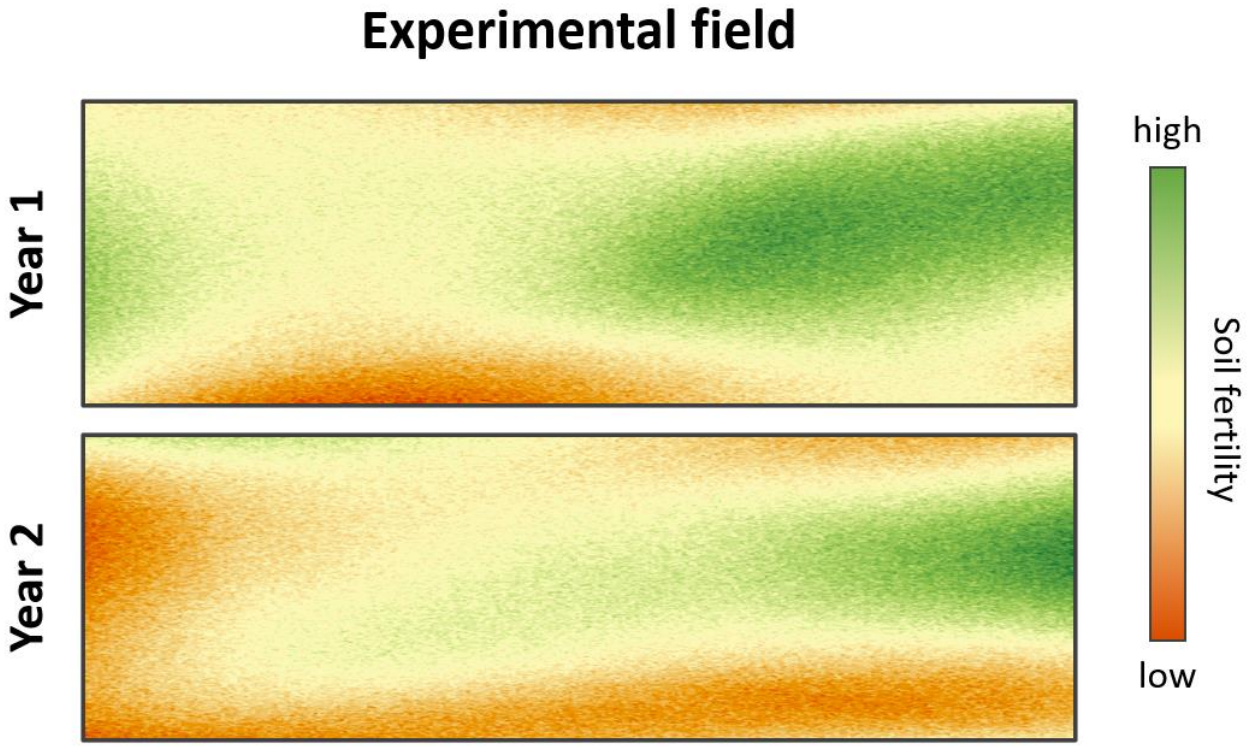



**Figure 1.** Spatial and seasonal variation of soil fertility in an experimental field at two different years. In the second year, a reduction of soil fertility can be observed at the left and bottom borders of the field.

Spatial and seasonal variation complicate the selection process in plant breeding programs for two reasons:

- The phenotypic performance of the same genotype can strongly vary when grown under different environmental conditions. This is an important phenomenon to consider when we derive genotypic values from phenotypic observations.
- A comparison of two or more genotypes grown in different environments may be unfair because phenotypic values might be significantly determined by differences in the environment rather than the underlying genetic values.

Phenotypes, however, are our primary source of information to derive genetic value. Therefore, to obtain an accurate predictor of the genetic value, we need to control and correct for spatial and seasonal variation as much as possible. We will now see how field trial designs and a proper statistical analysis enable us to achieve this.

# The principles of field experimentation

## Avoiding "post-mortem examinations"

In plant breeding trials, we routinely use a combination of experimental designs and statistical analysis to control and correct for environmental variation. In particular,

- Experimental designs help us to capture and control for environmental variation, and
- Statistical analysis helps us to correct for the factors of experimental variation captured by the experimental design.

While the design of a field trial always precedes its analysis, it is important that both measures are well coordinated. A good experimental design is ineffective without a suitable analysis. Likewise, a strong analysis will be of limited value if the basic principles of experimental design are ignored. As Fisher (1938) stated: "To consult the statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination. He can perhaps say what the experiment died of."

## Replication, randomization and blocking: the pillars of robust experimental designs

The basic principles of a statistically valid experimental design comprise:

1. Replication
2. Randomization
3. Blocking (local control)

These three "principles of field experimentation" (**Figure 2**) were introduced by R.A. Fisher (1925; 1926) nearly 100 years ago and still remain valid today. Although there is a vast number of experimental designs used in plant breeding trials at different stages and for different purposes (e.g., Cochran and Cox 1992; Hinkelmann and Kempthorne 2005), all designs trace back to Fisher's early work.

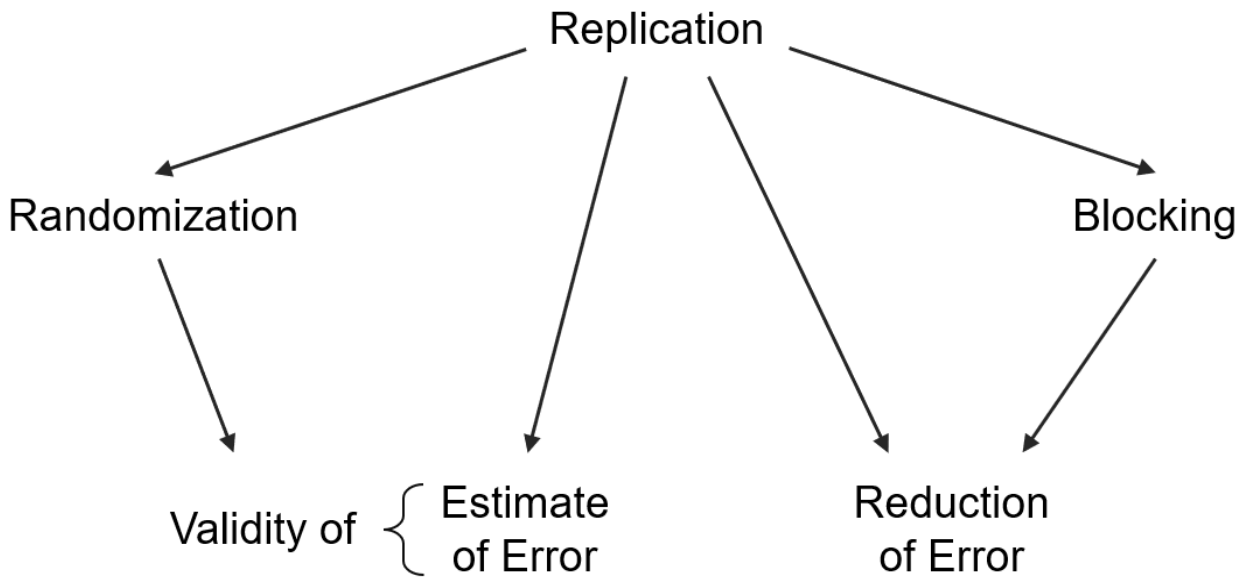



**Figure 2.** Fisher's diagram "The Principles of Field Experimentation" (adapted from Preece (1990)).

**Replication** is the repetition of an experimental treatment. In plant breeding, we mostly think of the repeated testing of selection candidates such as lines, hybrids, clones, or entire families when we refer to replication.

Replication serves two purposes, and both of them are important prerequisites to obtain accurate treatment effects:

- It allows us to obtain an estimate of the experimental error (standard error of the mean of the treatments), i.e., the variability of a treatment due to non-systematic environmental effects.
- It simultaneously reduces the experimental error (standard error of the mean of the treatments). While a single phenotypic measurement, for example, can be strongly affected by random environmental conditions, the mean of multiple replicates will approach the true genetic value and therefore increase repeatability.

Replication can be conducted within and across environments. Replication within an environment will increase the precision (repeatability) of the estimated genetic value only within that environment. Replication across environments will increase the precision (repeatability) of the estimated genetic value within the entire target population of environments.

**Randomization** is the random allocation of treatments to experimental units, such as the random allocation of genotypes to plots, rows, and columns. In a randomized design, each experimental unit has an equal probability of receiving a particular treatment. Thereby, randomization ensures that treatment effects and environmental effects are independent of each other. In a field trial, randomization provides protection against unknown effects of spatial trends on genotypic performance. Therefore, it helps to avoid confounding between treatment effects and environmental effects. Randomization is also a prerequisite to obtaining a statistically valid and unbiased estimate of the experimental error. Systematic treatment allocation can result in an inflated or deflated estimate of the error and, therefore, in a considerable loss of the precision of treatment estimates. Thus, in combination with replication, randomization helps to minimize selection bias and ensures a fair comparison of treatments (**Figure 2**).

A practical consequence is that close relatives should not be planted together in close proximity. While testing relatives together might have some logistical advantages, it is not usually an efficient strategy when selection is to be exercised across groups, as opposed to selection within groups (Piepho and Williams 2006). While there might be situations where complete randomization is difficult to realize, systematic testing of relatives closely together should be avoided when possible.

**Blocking**, or "local control", as originally referred to by R. A. Fisher (1926), is the process of reducing the experimental error by dividing a heterogeneous experimental area into more homogeneous subunits. A common example in plant breeding is the grouping of a field trial into blocks. Spatial trends in soil heterogeneity are often strong, which results in an increased experimental error. Differences in growing conditions, however, will generally be lower in

small subunits than across the entire experimental field. Blocking allows us to separate soil fertility effects from the experimental error since the difference between treatments can then to some extent be explained by the effect of a block. When genotypes are compared, phenotypic records can be corrected for these effects, resulting in increased accuracy of the estimated genetic values and reduced selection bias. Blocking factors are the basis of field designs and include, for example, replicates, (incomplete) blocks, rows, and columns.

## Going from one to multiple environments

A simple example of randomization, replication and blocking applied in practice is a randomized complete block design (RCBD). In an RCBD, each block gets all genotypes, and the entire test set of genotypes is replicated in multiple blocks. Within complete blocks, treatments are randomized to ensure independence of treatments and components of soil heterogeneity. In this way, differences in performance due to differences in fertility among blocks will be reduced, and the phenotypes become more reliable estimates of genetic values (Fisher 1926). An example of an RCBD is given in **Figure 3**.

Replication, randomization, and blocking are fundamental principles to control and correct for spatial variation within experimental fields. Replication, however, is not restricted to a single environment. To increase the accuracy of estimated genetic values, genotypes should be tested across multiple locations and years. However, testing resources in plant breeding programs are limited, and at each stage of testing, there will be a trade-off between the number of tested genotypes and the level of replication. While in early testing stages, unreplicated and partially replicated (p-rep) designs in one or a few locations are common to select promising candidates among a large number of genotypes, late-stage trials test a small set of potential candidate varieties across multiple locations and years. Conducting and analyzing such "multi-environment trials" (METs) requires introducing the concepts of ***Genotype-by-environment interaction (GxE)*** interaction and the ***Target population of environments (TPE)***.

We will define these two concepts in the next section and explain why their introduction requires an extension of the phenotype decomposition model given in **Equation 1** and **Equation 2**.

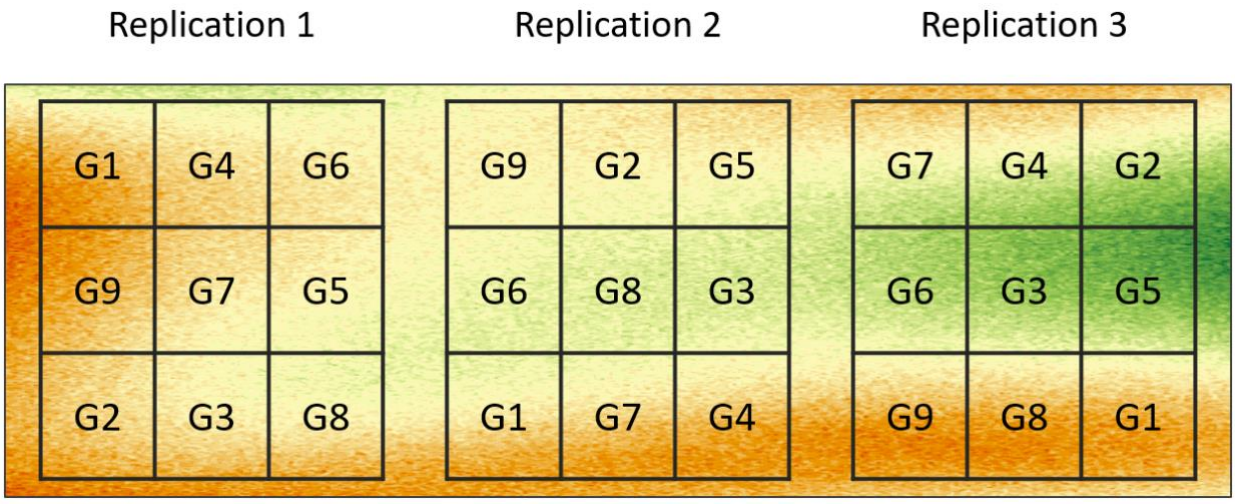



**Figure 3.** Illustration of Fisher's principles of field experimentation using a randomized complete block design (RCBD). Nine genotypes are tested in three replications across three blocks. Each block gets all genotypes, and genotypes within blocks are randomized.

# Multi-environment trials and genotype-by-environment (GxE) interaction

## Replication within environments versus across environments

Multi-environment trials (METs) are used to test genotypes in multiple year x location combinations. They are common in late testing stages when the number of genotypes has already been reduced to a small set of candidate varieties. Since the phenotypic expression of quantitative traits is more sensitive to changing environmental conditions than that of qualitative traits, testing in multiple environments is especially important to obtain accurate estimates for key traits such as yield and yield stability, or quantitative resistance to biotic and abiotic stress.

Multi-environment trials require that resources are efficiently used for replication of genotypes within and across environments. Replication within environments is needed to obtain an estimate of the experimental error and increases the accuracy of estimates of genetic values within this particular year x location combination. Replication across environments increases the accuracy of estimated genetic values across the whole target population of environments (TPE).

## Target population of environments (TPE)

The target population of environments can be defined as the entire set of farms and future seasons in which the varieties produced by a breeding program will be grown (Atlin et al. 2011). Optimally, a TPE combines geographical locations with similar growing conditions and farmers with similar variety preferences. Thus, factors that are commonly used to delineate a TPE include environmental factors as well as socio-economic factors (such as the production system, the variety type, available technologies, and a farmer's subjective preferences).

To develop varieties that match the profile of a well-defined TPE, breeders need to identify a set of test locations that is highly representative of the TPE. While TPE alignment is an extensive topic on its own that cannot be covered here, we will briefly review the probably most important factor to be considered when defining a TPE: genotype-by-environment (GxE) interaction.

## Genotype-by-environment (GxE) interaction

Genotype-by-environment interaction (**Figure 4**) occurs when two or more genotypes differ in their relative performance across environments (Bernardo 2010). While multi-environment trials provide us with more information to predict how a genotype will perform within the entire TPE, testing in multiple year x location combinations adds an extra layer of complexity to the statistical analysis. A genotype's performance is likely to be different across years and locations and might strongly fluctuate when tested in multiple environments. This change in performance is not only due to the effect of the genotype and the environment (i.e., the main factors) but also due to an interaction between the genotypes and the environment.

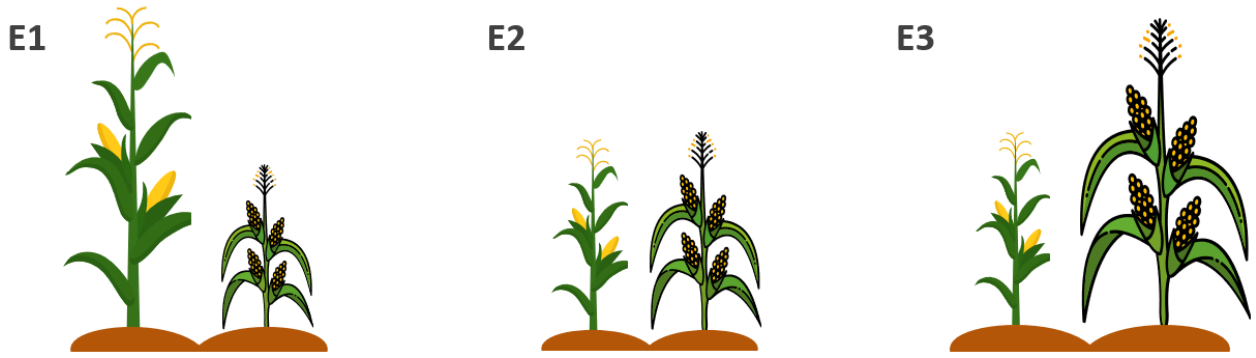



**Figure 4.** Simplified visualization of genotype-by-environment (GxE) interaction. Two maize genotypes are tested in three different environments (E1-E3). The difference in the relative performance of the genotypes across the three environments indicates GxE interaction.

## The phenotype as a function of genotype, environment, and genotype-by-environment (GxE) interaction

Understanding GxE interactions is important for breeders to identify varieties that show high performance on average and also exhibit high performance stability across the entire TPE. To include GxE interactions in our field trial analyses, we need to extend the model presented in **Equation 2** to an appropriate multi-environment model:

**Population mean** **GxE interaction** of genotype *i* in environment *j*

$$y_{ijk} = \mu + g_i + t_j + gt_{jk} + e_{ijk}$$

**Phenotype** of **genotype *i*** in **environment *j***, **replication *k***

**Equation 3**

Where:

$y_{ijk}$ is the jth observation on genotype *i* in environment *k*.

$\mu$ is the population mean.

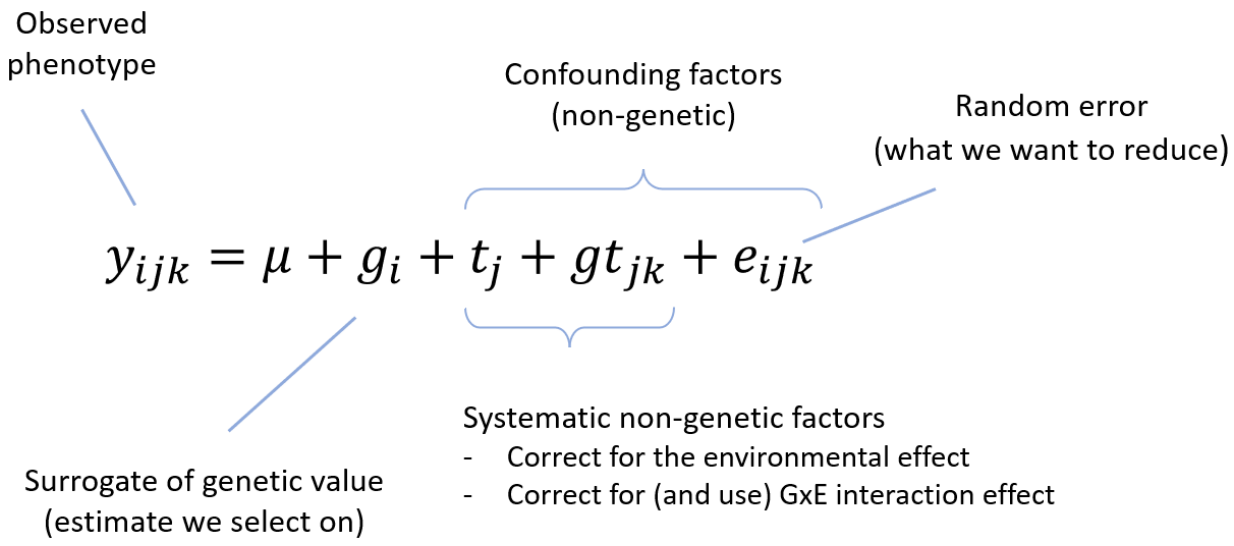
$g_i$ is the genetic value of genotype *i* expressed as a deviation from the population mean.

$t_j$ is the effect of environment j.

$g_{tij}$ is the genotype x environment interaction between genotype gi and environment tj.

$e_{ijk}$ is the residual for observation *k* on genotype *i* in environment *j*.

The genetic value of genotype *i* ($g_i$) in our TPE is what we want to estimate to select new parents and identify candidate varieties. While the observed phenotype of observation *k* on genotype *i* in environment *j* ($y_{ijk}$) is likely to deviate from the true genetic value of genotype *i* ($g_i$) due to several non-genetic confounding factors ($t_j$, $gt_{ij}$, $e_{ijk}$), the effect of the environment ($t_j$) and the GxE interaction ($g_{tij}$) on the phenotype can be estimated using an appropriate field design and a suitable statistical model:




**Equation 4**

The estimate of the GxE interaction term ($gt_{ij}$) tells us how stable a genotype performed across the different test environments. A relatively low GxE interaction indicates high stability of the genetic value ($g_i$) across the tested locations, and a high GxE interaction indicates that a genotype's performance strongly varied across environments. The effect of the environment ($t_j$) often refers to a specific year x location combination[2]. The residual of observation *k* on genotype *i* in environment *j* ($e_{ijk}$) contains all non-systematic non-genetic

[2] The model may be extended to include lower-level blocking factors such as year, location, block, row, and column effects. For illustration purposes we ignored these factors here, although they may help to further explain and reduce the residual variance.

factors which remain undefined and therefore cannot be estimated. This residual term (or random error) is what we aim to minimize through the modeling of systematic non-genetic factors.

## The genetic value, the environmental effect, and the genotype-by-environment (GxE) interaction effect

The genetic value, the effect of the environment, and the genotype-by-environment (GxE) interaction are not clear-cut values intrinsic to a genotype, a specific environment, and their interaction. They are properties of the TPE, they depend on the statistical model used to analyze MET data and they are interdependent.

The genetic value ($g_i$) is defined as the mean of a genotype across all tested environments ($P_{i..}$) minus the overall mean of the population ($\mu$):

$$g_i = P_{i..} - \mu$$

**Equation 5**

Likewise, the effect of the environment ($t_j$) is defined as the mean of all genotypes tested within this environment ($P_{.j.}$) minus the overall mean of the population (μ):

$$t_j = P_{.j.} - \mu$$

**Equation 6**

Therefore, the genetic value of a genotype depends on the environments in which the genotype is measured, and the effect of the environment depends on the genotypes grown in it. These are very important properties of the two estimates, as they show that the genetic value and the effect of an environment are statistical properties rather than fixed values.

This realization also raises the question of whether a "true" genetic value or a "true" effect of the environment does exist. In theory, we obtain the true genetic value when we test a genotype in all possible environments. Likewise, we obtain the true effect of an environment when we test all possible genotypes in this environment. In practice this is not possible, and true values can only be approximated by testing as many genotypes in as many environments as possible. Another consequence of this definition of the genetic value is that it is dependent on the target population of environments. Therefore there is not a single objective genetic value, as the same genotype will have different genetic values in different TPEs.

The genotype-by-environment (GxE) interaction effect is defined as the difference between the mean phenotype of genotype *i* across all replications in environment *j* ($P_{ij.}$) and the sum of the population mean ($\mu$), the genetic value ($g_i$), and the effect of the environment ($t_j$).

$$(gt)_{ij} = P_{ij.} - (\mu + g_i + t_j)$$

**Equation 7**

The sum of the population mean, the genetic value, and the effect of the environment is the predicted mean phenotypic performance of genotype *i* in environment *j*. The GxE interaction effect then simply is calculated as the difference between the predicted and the observed phenotype ($P_{ij.}$). In comparison to the effect of the environment, which affects all genotypes in this environment equally, the GxE interaction effect is specific to the combination of a genotype and an environment.

## Patterns of genotype-by-environment (GxE) interaction

Genotype-by-environment interaction occurs in various patterns which can be classified into:

- Non-cross over GxE interaction (scaling).
- Cross over GxE interaction.
- A combination of both.

When testing two genotypes in two environments, six different reaction norm trends can be identified, which are presented in **Figure 5**.

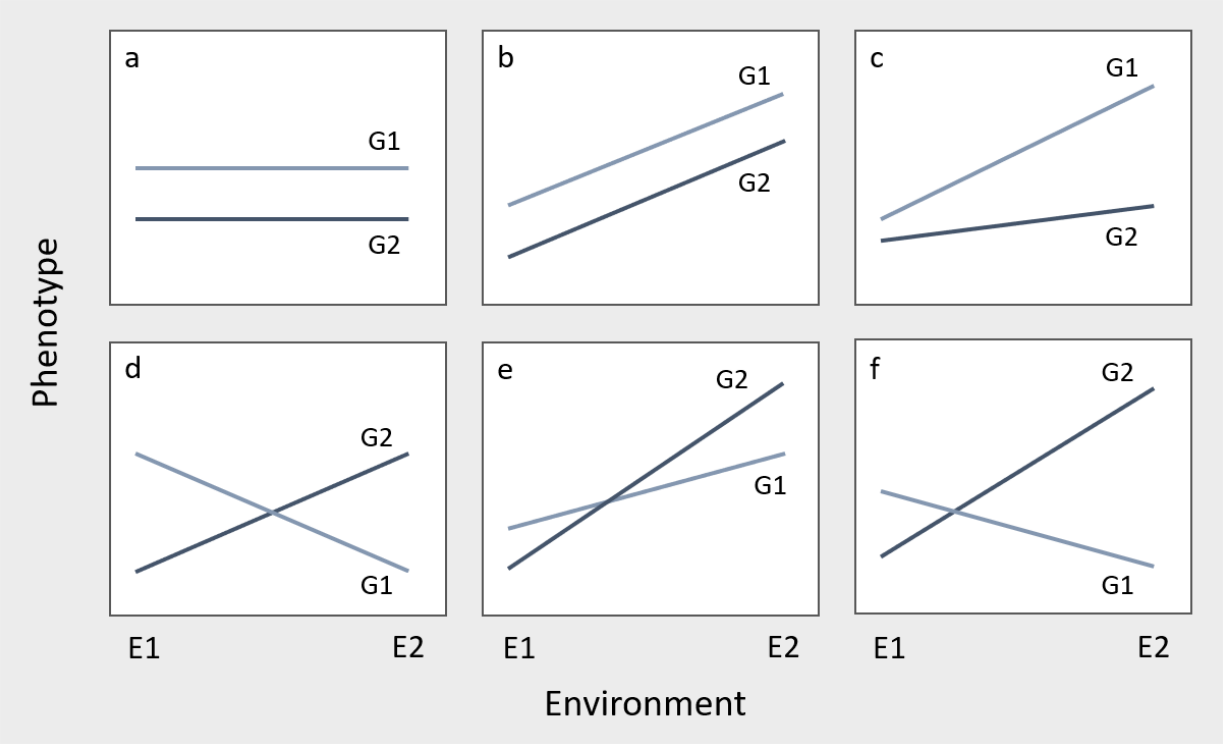



**Figure 5.** Patterns of genotype-by-environment (GxE) interaction.

a. No GxE interaction: both environments E1 and E2 have a similar effect. The phenotypic performance of genotypes G1 and G2 is not affected by the environment. This pattern might be observed in a highly controlled lab trial carried out in two different labs under similar conditions.

b. No GxE interaction: environment E2 has a higher effect than E1 and both genotypes G1 and G2 show higher phenotypic performance in E2 than in E1. The relative distance between the performance of the two genotypes is the same in both environments. This pattern might be observed when testing for a qualitative trait in two different environments, such as a high and a low disease pressure environment.

c. Non-cross over GxE interaction: environment E2 has a higher effect than E1, and both genotypes G1 and G2 show higher phenotypic performance in E2 than in E1. Additionally, the relative distance between the performance of the two genotypes is higher in E2 than in E1. This pattern might be observed when testing for a quantitative trait in two different environments, such as yield in a dry location and an irrigated location.

d. Cross over GxE interaction: there is a change in rank of G1 and G2, but the relative difference between the two genotypes is the same in both environments. In practice, this is an unlikely observation, as cross over GxE interaction usually occurs in combination with some degree of non-cross over GxE interaction (E. and F.)

e. Non-cross over + cross over GxE interaction: genotype G1 shows lower phenotypic performance than G2 in environment E1, but higher phenotypic performance in environment E2. Both genotypes show increased performance in environment E2, but the difference between their performance increases. This is a typical pattern observed for a quantitative trait in multi-environment trials.

f. Non-cross over + cross over GxE interaction: genotype G1 shows lower phenotypic performance than G2 in environment E1, but higher phenotypic performance in environment E2. In contrast to scenario (E.), G1 shows an increased performance in environment E2, and G2 shows a reduced performance in environment E2. This is also a typical pattern observed for a quantitative trait in multi-environment trials.

While non-cross over GxE interaction (C) is not a problem for the selection process, cross over GxE interaction is, because it changes the ranking of genotypes in different environments. Unfortunately, however, cross over GxE interaction is a common phenomenon. Besides breeding for performance stability, the most efficient strategy to reduce cross over GxE interaction is appropriate TPE alignment and strict TPE-oriented breeding. This allows breeders to exploit GxE interaction rather than just accepting it.

# References


Atlin GN, Kleinknecht K, Singh KP, Piepho HP (2011) "Managing Genotype x Environment Interaction in Plant Breeding Programs: A Selection Theory Approach." *J. Indian Soc. Agric. Stat.* 65 (2): 237–47.

Bernardo RN (2010) *Breeding for Quantitative Traits in Plants*. 2nd ed. Woodbury, Minn: Stemma Press.

Cochran WG, Cox GM (1992) *Experimental Designs*. 2nd ed. Wiley Classics Library. New York: Wiley.

Falconer DS, Mackay TFC (1995) *Introduction to Quantitative Genetics*. 4. ed., [16. print.]. Harlow: Pearson, Prentice Hall.

Fisher RA (1925) *Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.*

Fisher RA (1926) "The Arrangement of Field Experiments."

Fisher RA (1938) Presidential Address to the First Indian Statistical Congress.

Hinkelmann K, Kempthorne O (2005) *Design and Analysis of Experiments: Advanced Experimental Design*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471709948.

Piepho HP, Williams ER (2006) "A Comparison of Experimental Designs for Selection in Breeding Trials with Nested Treatment Structure." *Theoretical and Applied Genetics* 113 (8): 1505–13. https://doi.org/10.1007/s00122-006-0398-8.

Preece DA (1990) "R. A. Fisher and Experimental Design: A Review." *Biometrics* 46 (4): 925. https://doi.org/10.2307/2532438.