

DOI: 10.1515/ffp-2016-0026

## Application of GGE biplot graphs in multi-environment trials on selection of forest trees

Krzysztof Ukalski<sup>1</sup> ✉, Marcin Klisz<sup>2</sup>

<sup>1</sup> Warsaw University of Life Sciences – SGGW, Faculty of Applied Informatics and Mathematics, Department of Econometrics and Statistics, Biometry Division, Nowoursynowska 159, 02-776 Warsaw, Poland, e-mail: krzysztof\_ukalski@sggw.pl

<sup>2</sup> Forest Research Institute, Department of Silviculture and Genetics, Braci Leśnej 3, Sękocin Stary, 05-090 Raszyn, Poland

### ABSTRACT

In the studies on selection and population genetics of forest trees that include the analysis of genotype  $\times$  environment interaction (GE), the use of biplot graphs is relatively rare. This article describes the models and analytic methods useful in the biplot graphs, which enable the analyses of mega-environments, selection of the testing environment, as well as the evaluation of genotype stability. The main method presented in the paper is the GGE biplot method (G – genotype effect, GE – genotype  $\times$  environment interaction effect). At the same time, other methods have also been referred to, such as, SVD (singular value decomposition), PCA (principal component analysis), linear-bilinear SREG model (sites regression), linear-bilinear GREG model (genotypes regression) and AMMI (additive main effects multiplicative interaction). The potential of biplot method is presented based on the data on growth height of 20 European beech genotypes (*Fagus sylvatica* L.), generated from real data concerning selection trials and carried out in 5 different environments. The combined ANOVA was performed using fixed-effects, as well as mixed-effects models, and significant interaction GE was shown. The GGE biplot graphs were constructed using PCA. The first principal component (GGE1) explained 54%, and the second (GGE2) explained more than 23% of the total variation. The similarity between environments was evaluated by means of the AEC method, which allowed us to determine one mega-environment that comprised of 4 environments. None of the tested environments represented the ideal one for trial on genotype selection. The GGE biplot graphs enabled: (a) the detection of a stable genotype in terms of tree height (high and low), (b) the genotype evaluation by ranking with respect to the height and genotype stability, (c) determination of an ideal genotype, (d) the comparison of genotypes in 2 chosen environments.

### KEY WORDS

AMMI additive main effects multiplicative interaction, GGE biplot analysis, genotype  $\times$  environment interaction, multi-environment trial, PCA principal component analysis, SVD singular value decomposition

## INTRODUCTION

In a multi-environment trial (MET) analysis, the effect of genotype can be tested in different combinations, depending on whether there is one-year data taken into account concerning several environments (genotype  $\times$  environment), or a few years data concerning one environment (genotype  $\times$  year), or a few years data concerning several environments (genotype  $\times$  environment  $\times$  year). The complete data set required for the construction of biplot graphs consists of genotype means (table rows) across environments or across years or across environments in the years (table columns). The biplot graph allows the analysis of both variables in the table (rows and columns), relationships between columns or rows, as well as those concerning their interactions. The biplot analysis includes no hypothesis testing. The use of biplot graphs in data analysis involves the application of statistical methods that enable testing significance of differences between the means in rows or columns, as well as those between row  $\times$  column interactions.

Data compiled in the two-dimensional table of means are factorized using singular value decomposition (SVD) (Pearson 1901; Golub and Reinsch 1971). The earliest works concerning the application of this method in various scientific fields emerged 40 years ago (Gabriel 1971, 1972, 1978; Bradu and Gabriel 1978). Nowadays, the majority of statistical programs include procedures to carry out SVD and construct the biplot graphs.

Three models are used often in the biplot graph construction in reference to SVD:

$$M_{ij} = \bar{y}_{ij} - \mu - c_j = r_i + rc_{ij} \quad (1)$$

$$M_{ij} = \bar{y}_{ij} - \mu - r_i = c_j + rc_{ij} \quad (2)$$

$$M_{ij} = \bar{y}_{ij} - \mu - r_i - c_j = rc_{ij} \quad (3)$$

where:

$M_{ij}$  – expresses the matrix  $\mathbf{M}$  with  $k$ -rows ( $k$ -number of tested objects,  $i = 1, \dots, k$ ) and  $l$ -columns ( $l$ -number of tested environments,  $j = 1, \dots, l$ );

$\bar{y}_{ij}$  – the mean of  $i$ -th object in  $j$ -th environment;

$\mu$  – the overall mean;

$r_i$  – the main effect of  $i$ -th object;

$c_j$  – the main effect of  $j$ -th environment;

$rc_{ij}$  – the interaction between  $i$ -th object and  $j$ -th environment.

In the model 1, the overall mean  $\mu$  and the main effects of columns (environment main effects) are subtracted from the mean values  $\bar{y}_{ij}$  in the table of means. Next,  $\mathbf{M}$  matrix is factorized with the use of SVD. Equivalently, calculations can be performed on the two-dimensional table of means  $\bar{y}_{ij}$  using PCA (Pearson 1901, Hotelling 1933, 1936). When computing the values of the principal components using statistical software, it should be considered if the values are calculated based on covariance matrix or correlation matrix. The covariance matrix should be used in order to obtain correct results. Subsequent to PCA calculations, the values of principal components for genotypes should be divided by the root of the product of multiplication: genotype number and respective principal component Eigen value. Once the calculations are performed, the PCA results should be equivalent to those obtained with SVD. Analogous results for model 1 would be obtained using SREG linear-bilinear model (sites regression) (Cornelius and Seyedsadr 1997).

After the groundwork publications by Gabriel (1971) that showed the meaning of selection of SVD model to construct the biplot graphs of different kinds, further works of Kempton (1984), Gauch (1992), Cooper and DeLacy (1994) were directed towards studying the stability of crop plants' yield under various environments, by means of the biplot graphs. The results obtained showed that biplot graph displays worked very well in the analysis of the effects of genotypes (G), genotype environments (E), and in the analyses of genotype  $\times$  environment interaction (GE). The terminology "GGE biplot" (G – genotype effect, GE- genotype  $\times$  environment interaction effect) was introduced by Yan et al. (2000). The GGE biplot method was further developed by Yan (2001), Yan and Kang (2003), and Yan and Tinker (2005, 2006).

In the second (2) model, the overall mean  $\mu$  and genotype main effects  $r_i$  were subtracted from the mean values  $\bar{y}_{ij}$  in the two-dimensional table of means. Next, the matrix  $\mathbf{M}$  was factorized by SVD. Equivalent results can be obtained with the use of the linear-bilinear model GREG (genotypes regression) (Cornelius et al. 1996; Cornelius and Seyedsadr 1997). The latter repre-

sents the re-parameterized model for the analysis of stability of the models by Finlay and Wilkinson (1963) and Eberhart and Russell (1966) that use genotype linear regression on environmental means (Cornelius et al. 1996; Cornelius and Crossa 1999; Liu and Cornelius 2001; Crossa 2012). The model 2 serves better to describe the environments than the model 1, as long as the first two principal components in the model 2 explain more overall variation as compared to model 1.

In the third (3) model, the matrix  $\mathbf{M}$  was obtained that consisted of the effects of interaction of rows (genotypes) and columns (environments)  $rc_{ij}$ . Similar results are achieved in the analysis of additive main effects and multiplicative interaction (AMMI) (Zobel et al. 1988; Gauch 1988, 1992).

MET data analysis for one trait should comprise 3 main aspects: (i) similarity analysis of environments and determination of mega-environments, (ii) selection of environment tested, (iii) evaluation of genotype stability (Yan and Kang 2003).

Long-term development of woody plants, as well as the fact that they undergo a juvenile stage that is characteristic of a rapid growth rate with low stability, are the decisive factors for a relatively rare use of biplot graphs in tree breeding studies. Nevertheless, in the recent years, there has been observed an increasing interest in this method (Murillo 2001; Kim et al. 2008; Ding et al. 2008; Correia et al. 2010; Taibi 2014; Sixto et al. 2015; Zhao et al. 2016), which is, among others, due to a considerably wide range of options of interpretation of data obtained in multi-environment genetic trials. The assessment of variability based on genotype  $\times$  environment interaction allows us to determine a level of genetic gain that is possible to achieve through genotype selection (Murillo 2001). Besides this, the calculation of genotype  $\times$  environment effect allows us to conclude on a range of seed transfers, followed by designating the seeding regions for genotypes and populations tested (Kim et al. 2008; Taibi 2014). Identification of the best provenance, in terms of a given trait or a group of traits, makes it possible to limit the selection process to the genotype promoted in tree plantations cultivated in short-term production cycles (Ding et al. 2008; Sixto et al. 2015). The use of biplot graphs to interpret the results of multi-environment trials at the narrowed selection criterion also lets the breeder select superior genotypes (Zhao et al.

2016). The distinction between genotypes specifically adapted to the chosen environments tested, as well as those stable in terms of the high value of the trait analyzed, allows us to elaborate the recommendations and strategies for reforestation/afforestation, breeding programs and conservation of genetic resources (Correia et al. 2010).

The aim of the study was to present the possibilities of using the biplot graphs in: (i) multi-environment analyses, (ii) selection of test environment, and (iii) evaluation of stability of forest genotypes.

## MATERIAL AND METHODS

The GGE biplot graphs were constructed using the data generated from the actual results of a selection trial conducted on half-sib families of *Fagus sylvatica*. The height of trees after 5 years of growth is discussed in the paper.

The trials on 20 beech genotypes were established in 5 environments in a randomized block design with 4 replications. Table 1 presents the means (of replications per environment) of beech heights in 20 genotypes investigated (G1, G2, ..., G20) on 5 experimental sites (E1, E2, ..., E5). The biplot graph analysis can be performed in a balanced design or in an unbalanced design with no missing cells (genotype-environment combinations should consist of at least 1 value). In the case of an incomplete data set, there is a need to perform additional complex calculations (Patterson and Thompson 1971, 1975; Piepho 1998). In the recent years, the restricted maximum likelihood method (REML) has been recommended for the analysis of unbalanced designs (Piepho 1998; Balzarini 2002; Piepho and Möhring 2006).

The GEE biplot method is effective only if the GE interaction is significant, thus, ahead of the biplot graph analysis, there has to be performed the combined ANOVA for the environments tested. Prior to ANOVA, it has to be decided which effects in the linear model are fixed and which are random. The choice of effect category depends on the type of experiment. When the genotypes under study are selected decisively with reference to a given trait, then the effect of genotype will be fixed (Saxton 2004). If the genotypes constitute representation of all the genotypes of a given species, the effect

of the genotype will be random (Ukalska and Kociuba 2013). Both approaches are applied in breeding studies on cultivated crops. In the experiments concerning the selection and genetics of forest trees, the effect of the genotype is most often fixed. The environments in MET data analysis can also be treated dually. If the experimental sites are selected so as to represent the widest possible spectrum of environmental conditions for a given species, the effect of environment should be treated as random (Saxton 2004). If at least one of the two main factor effects included in an interaction is random, the interaction term will be random as well. The choice of fixed and random effects influences the way of testing significance of the model effects.

**Table 1.** Mean height of 20 beech genotypes (G1, G2, ..., G20) observed in 5 environments (E1, E2, ..., E5)

| Genotypes | Environments |        |        |        |        |
|-----------|--------------|--------|--------|--------|--------|
|           | E1           | E2     | E3     | E4     | E5     |
| G1        | 70.24        | 105.2  | 126    | 155.48 | 89.98  |
| G2        | 75.05        | 90.62  | 136.7  | 154.74 | 154.98 |
| G3        | 63.18        | 119.82 | 141.08 | 158.7  | 100.49 |
| G4        | 85.35        | 94.13  | 132.58 | 139.44 | 139.7  |
| G5        | 68.6         | 96.95  | 136.66 | 155.57 | 110.44 |
| G6        | 79.93        | 99.42  | 127.69 | 148.89 | 148.07 |
| G7        | 70.4         | 97.13  | 132.79 | 139.41 | 105.14 |
| G8        | 84.88        | 111.13 | 147.34 | 148.12 | 112.66 |
| G9        | 79.8         | 94.59  | 141.09 | 145.88 | 114.87 |
| G10       | 78.09        | 104.75 | 125.47 | 151.39 | 119.63 |
| G11       | 73.57        | 93.66  | 148.75 | 143.48 | 110.89 |
| G12       | 84.99        | 99.83  | 139.48 | 156.32 | 139.48 |
| G13       | 83.62        | 104.19 | 115.06 | 164.73 | 129.57 |
| G14       | 71.83        | 88.26  | 132.84 | 138.02 | 116.73 |
| G15       | 82.23        | 115.17 | 141.3  | 149.08 | 129.52 |
| G16       | 73.26        | 91.07  | 125.61 | 143.64 | 130.08 |
| G17       | 65.05        | 103.29 | 128.04 | 147.43 | 117.39 |
| G18       | 73.01        | 85.85  | 139.65 | 150.55 | 150.93 |
| G19       | 69.98        | 103.93 | 128.53 | 142.78 | 109.3  |
| G20       | 60.08        | 83.05  | 103.6  | 144.78 | 116.78 |

In the analysis of variance concerning MET carried out in randomized block design, the linear model is most often used:

$$y_{ijk} = \mu + r_k(e_j) + g_i + e_j + ge_{ij} + \varepsilon_{ijk} \quad (4)$$

where:

- $y_{ijk}$  – observed value of the trait for  $i$ -th genotype ( $i = 1, \dots, 20$ ) in  $j$ -th environment ( $j = 1, \dots, 5$ );
- $\mu$  – the overall mean;
- $g_i$  – fixed effect of  $i$ -th genotype;
- $e_j$  – random or fixed\* effect of  $j$ -th environment;
- $ge_{ij}$  – random or fixed\* effect of genotype  $\times$  environment interaction GE;
- $r_k(e_j)$  – random or fixed\* effect of  $k$ -th block in  $j$ -th environment;
- $\varepsilon_{ijk}$  – experimental error;

\* if the effect of environment is fixed, the effect of genotype  $\times$  environment interaction GE is fixed as well.

The construction and ways of using the biplot graphs are presented below, in the section Results and Discussion.

Statistical analyses and the graphs were performed with SAS 9.4 software (SAS Institute 2013), by means of the following procedures: GLM, PRINCOMP, GPLOT (Littell et al. 1996; Khattree and Naik 2000).

## RESULTS AND DISCUSSION

### Analysis of variance

In order to demonstrate differences between the fixed and random approaches, ANOVA was performed for the model 4, taking into account random (Tab. 2A) and fixed (Tab. 2B) effect of environment (Tab. 2). F statistics values in the mixed model were determined according to Hocking's approach (Hocking and Speed 1975).

Significant differences were found in relation to environments and genotype  $\times$  environment interaction. However, no significant differences were observed between the genotype means for the model with random environment effect (Tab. 2A). On the other hand, for the model with fixed environmental effect (Tab. 2B), significant differences were found between genotypes and environments and significant genotype  $\times$  environment interaction was also observed. As shown by the above example, the choice of type of effects in the assumed ANOVA model has an influence on the evaluation of significance of effects and the interpretation of results obtained.

**Table 2.** Results of analysis of variance for tree heights in 20 beech genotypes at 5 environments along with the percentage of variation explained

| Source of variation         | df  | EMS      | F       | P      | % (G + E + GE) |
|-----------------------------|-----|----------|---------|--------|----------------|
| Part A                      |     |          |         |        |                |
| Genotype (G)                | 19  | 670.73   | 1.49*   | 0.11   | 4.00           |
| Environment (E)             | 4   | 67804.00 | 129.81* | <0.001 | 85.22          |
| Block (Environment)         | 15  | 145.86   | 1.95    | 0.02   | -              |
| Genotype × Environment (GE) | 76  | 451.29   | 6.03    | <0.001 | 10.78          |
| Experimental error          | 285 | 74.82    | -       | -      | -              |
| Part B                      |     |          |         |        |                |
| Genotype (G)                | 19  | 670.73   | 8.96    | <0.001 | 4.00           |
| Environment (E)             | 4   | 67804.00 | 464.86* | <0.001 | 85.22          |
| Block (Environment)         | 15  | 145.86   | 1.95    | 0.02   | -              |
| Genotype × Environment (GE) | 76  | 451.29   | 6.03    | <0.001 | 10.78          |
| Experimental error          | 285 | 74.82    | -       | -      | -              |

A – random effect of environment (E); B – fixed effect of environment (E); df – degrees of freedom; EMS – expected mean squares; F – statistics are calculated according to Hocking’s approach.

\* Hocking’s approach

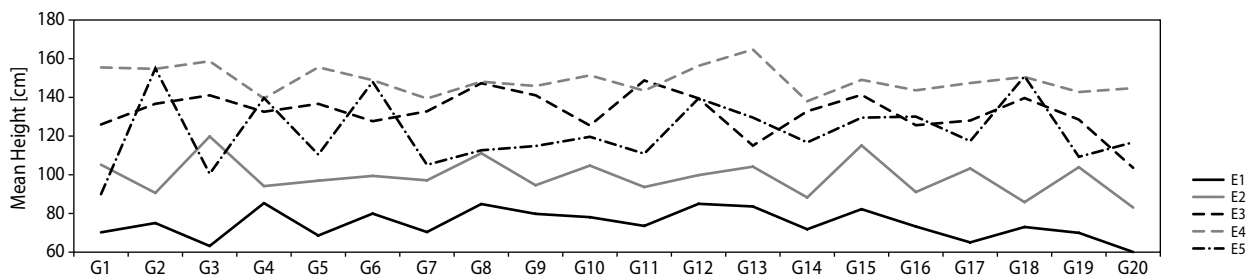
85.22% of the overall variation in height (G+E+GE) was explained by the effect of environment E, and the remaining 14.78% were explained by the effects of genotype G (4%) and interaction GE (10.78%) (Tab. 2). In majority of MET experiments, the effects E, G and GE explain 80%, 10% and 10%, respectively (Gauch and Zobel 1997), and as a rule, the environment variation

explains most of the overall phenotypic variance (Yan and Rajcan 2002).

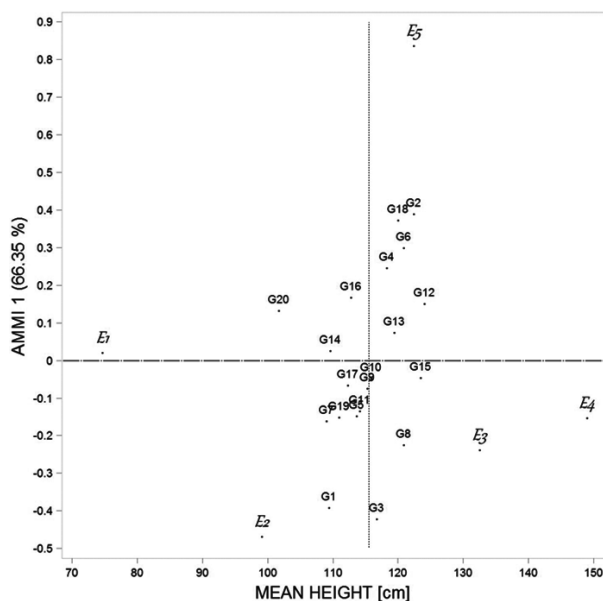
The chart (Fig. 1), which shows the mean tree height of 20 beech genotypes examined in 5 environments, does not assist either the analysis of environments or the evaluation of genotypes. In order to preliminary characterize the environments and genotypes, a biplot graph can be constructed using the AMMI procedure, or by SVD calculations in accordance with the model 3 (Fig. 2). The first component AMMI1 explained 66.35% of variation. The horizontal axis describes the tree height, and the vertical axis shows the values of AMMI1. The higher the absolute value of AMMI1, the higher is the value of genotype × environment interaction. The genotypes were least stable in E5 environment (the highest AMMI1 absolute value). The highest mean heights were observed in the genotypes in E4 environment, and the lowest – in E1 environment. The unstable genotypes (with the highest absolute AMMI1 values) were as follows (in descending order): G3, G1, G2, G18, G6, G4, G8 and G16. The stable genotypes (AMMI1 absolute values close to 0), were as follows (in descending order): G14, G15, G10, G17, G9, G13, G20 and G11.

Kim et al. (2006) used the AMMI graph for the preliminary assessment of the mean height of 36 Japanese pine (*Pinus densiflora*) provenances in 11 localities in South Korea. The first AMMI component (IPCA1) based on the average height increment let the authors determine a level of provenance adaptation to the growth environments, as well as to distinguish specifically adapted populations. The AMMI procedure was also used by Taibi (2014), in the study on the black pine (*Pinus nigra ssp. Salzmannii*) and the Aleppo pine (*Pinus halepensis* Mill.).

The AMMI model is inadequate for the construction of subsequent biplot graphs. A key difficulty ari-



**Figure 1.** Mean tree height in 20 beech genotypes in 5 environments E1, ..., E5



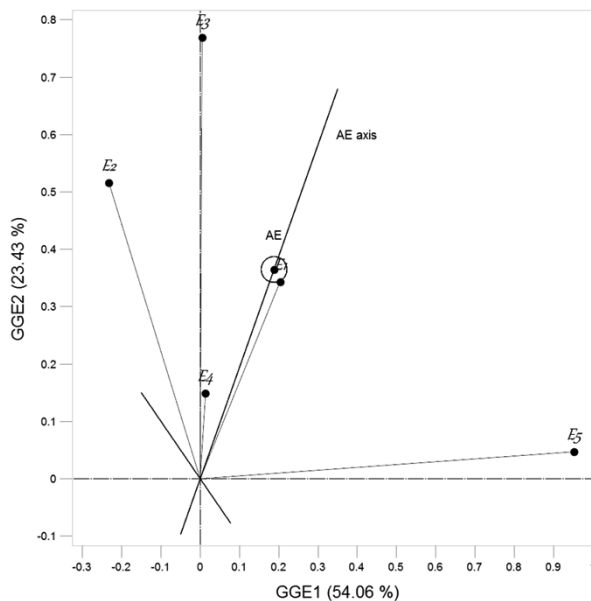
**Figure 2.** Biplot of the first component axis AMMI1 versus mean height

ses from different unit scales on the graph axes. The GEE biplot graphs constitute much better solutions for MET data analysis as regards one trait. The GGE biplot graph values can be obtained by SVD in accordance with model 1 or by using PCA, or else – SREG linear – bilinear model. In the present study, further calculations were performed using PCA. The results of the latter showed that the first principal component GGE1 explained 54.06% of the total variation, and the second (GGE2) explained 23.43%. In total, both principal components explained more than 77.49% of the overall variation. According to Jolliffe (1972, 1973) and McCabe (1984), such value of variation as explained by two first principal components is sufficient enough to effectively explain differences between the studied objects.

**Environmental similarity analysis**

The AEC method (Yan and Hunt 2001; Yan 2002) was used to examine similarities between environments on the GGE biplot graph (Fig. 3). The average environment AE was determined based on the principal components GGE1 and GGE2 for 5 environments. Next, the AE axis was combined between AE point and the coordinate system centre.

The position of environments with respect to the average environment (AE) axis shows that E1 envi-



**Figure 3.** Average environment coordination (AEC) in view of GGE biplot graph based on principal components for environments. AE is “the average environment”

ronment was the closest to AE as regards tree height stability (Fig. 2 points toward the same conclusion). Regarding the beech trait studied, E5 environment was the most unstable (the most distant from the AE axis; Fig. 2 points toward the same conclusion).

Correlations between environments in the GGE biplot graph were determined based on the angles between the vectors of environments (Tab. 3). Also, the correlation coefficients between the expressions of GE matrix product were determined, where G was the matrix of the principal components for genotypes, and E was the transposed matrix of the principal components for environments (Yan and Kang 2003).

**Table 3.** Correlation coefficients (above diagonal) and  $a_{ij}$  angles (below diagonal) between pairs of environments

| Environments | E1    | E2     | E3     | E4     | E5      |
|--------------|-------|--------|--------|--------|---------|
| E1           | –     | 0.58** | 0.86** | 0.90** | 0.55*   |
| E2           | 54.90 | –      | 0.91** | 0.87** | -0.37ns |
| E3           | 30.99 | 24.61  | –      | 0.99** | 0.06ns  |
| E4           | 25.44 | 29.46  | 4.84   | –      | 0.14ns  |
| E5           | 56.53 | 111.42 | 86.81  | 81.96  | –       |

Significant at 0.01(\*\*), 0.05(\*) probability levels. ns – not significant.



The two environments E3 and E4 were most similar to each other, whereas environments E1 and E2 were similar to them (Tab. 3, Fig. 3). The angles between E1, E2, E3 and E4 were less than 90 degrees, and the correlation coefficients between the environment pairs were significant (Tab. 3). Therefore, it was concluded that the environments E1, E2, E3 and E4 constituted one mega-environment. The vector of E5 environment formed almost right angles with E3 and E4 vectors (Tab. 3). E2 and E5 vectors formed an obtuse angle (Fig. 3), and the correlation in this case was not significant (Tab. 3). E5 was described as a separate environment, unlike mega-environment.

The SVD method was used in the study on 9 Mexican yellow pine *Pinus oocarpa* provenances, examined in 10 localities in South Africa (Murillo 2001). Then, the biplot graph was used to find similarities between the environments and 4 mega-environments were distinguished. This enabled the determination of achievable genetic gain with respect to the experimental system comprising all the environments jointly. Data analysis with the use of biplot graphs allowed for decreasing the effect of genotype  $\times$  environment interaction by grouping environments.

### Selection of test environment

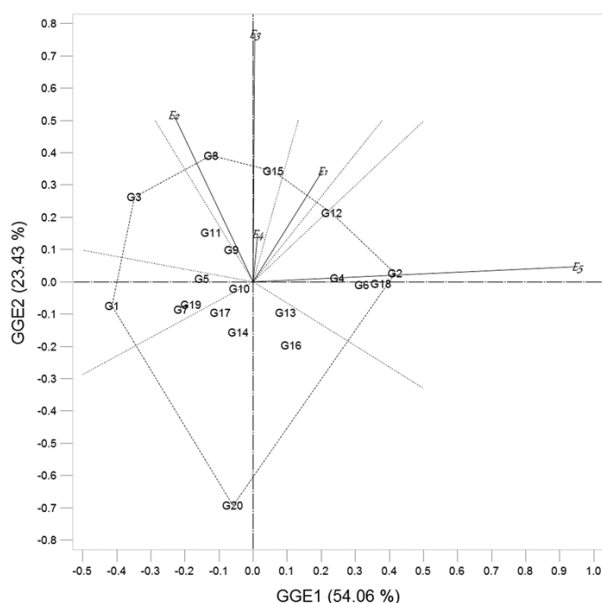
Application of the GEE biplot graph with the AE axis (Fig. 3) also allows to determine the environment where genotypes can be efficiently evaluated. The selected environment should characterize high ability to discriminate genotypes, as well as representativeness of the growing conditions (Yan 2002). The best in terms of genotype discrimination is the environment characteristic of the long vector, with its beginning at the centre of the coordinate system and its end at the environment point in the graph (in the present study environments E3 and E5 showed the longest vectors). In other words, such environment provides the best information on differences between genotypes. The smaller the angle between the environment vector and the AE line, the more representative is the environment (in the present study: E1) when compared to the environments with bigger angles (E5 and E2).

None of the environments tested in the present study was characterized at the same time by the longest vector and the smallest angle with the AE axis. This means that none of the five environments represented the ideal environment to perform the genotype selec-

tion, hence, the trials cannot be limited just to one test environment.

### Evaluation of genotype stability

The next GGE graph (Fig. 4) was used to find the genotypes with the highest GGE effect in each environment tested. The graph showed a convex polygon that was formed by linking the genotypes located outermost of the centre of the coordinate system. All the remaining genotypes were encompassed inside the polygon. The lines enclosing the polygon areas were perpendicular to the sides of the polygon or their extensions. A bunch of these half-lines (starting at the origin) split the plane into the areas. Within the areas, the positions of environments were observed, which enabled the depiction of the so called mega-environments within a given area (Gauch and Zobel 1997; Yan et al. 2000; Yan et al. 2001; Ukalski et al. 2010a, 2010b).



**Figure 4.** Polygon view of GGE biplot based on principal components for genotypes and environments

The genotypes and environments tested were located in 7 areas. Environments: E2, E3 and E4 were located in the sector formed by G15 genotype. Based on Fig. 3, showing that E1 belonged to the mega-environment together with E2, E3, E4 environments, and taking into account that the perpendicular lines enclosing the sectors showed that G15 was located in the sector

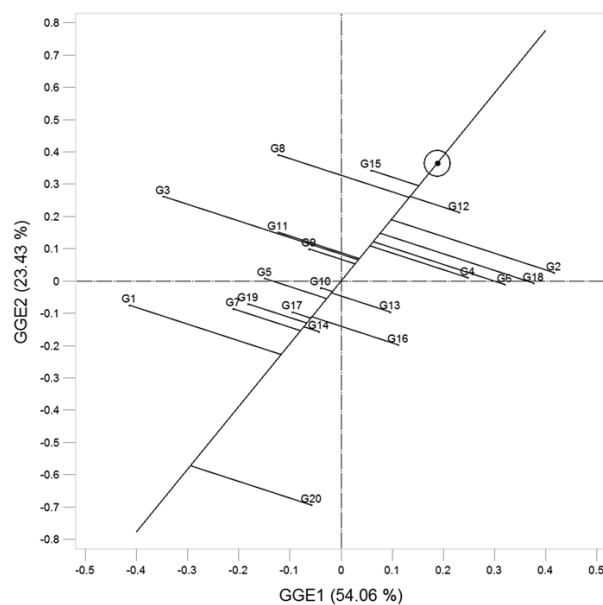
of E2, E3, E4, for simplification of conclusions, the two sectors were combined to form one sector. E5 environment was located in the sector formed by G2 genotype. In mega-environment E1, E2, E3, E4, the highest tree height means were obtained for the genotypes: G8, G15, G12, and the lowest for G20, G16, and G14. In E5 environment, the highest means were shown by G2, G18, G6, G4, and lowest by G1, G7, and G19.

The comparisons of 30 maritime pine *Pinus pinaster* genotypes derived from Portugal, France, Spain and Australia, were carried out with the use of a polygon on the GEE biplot graph (Correia et al. 2010). Adaptation capability of maritime pine genotype expressed by growth, quality and survival traits was tested in 5 environments in Portugal. The analysis of data obtained after 10 years of tree growth, carried out with the use of the biplot graph, allowed for defining the overall adaptation potential for the whole trial as well as specific adaptation patterns of the populations tested. In the study by Ding et al. (2008), from 165 to 216 Monterey pine *Pinus radiata* crosses were analyzed in 5 localities in Australia, with the use of polygons in the GGE biplot graphs. This let the authors define the crosses with the best increment within each of the tested environments. In both aforesaid studies, the AEC method was used in order to rank the genotypes, which is explicated below in Fig. 5. The biplot graphs constructed using the SREG method were presented by Sixto et al. (2015), in the study on poplar growth in short-term production cycles, conducted in 4 localities in Spain. The authors used this method in order to determine the best clones with respect to the produced biomass, building opinion on the results of multi-environment trial.

In order to evaluate the genotypes by ranking with reference to the mean value of the tested trait and trait stability, the AEC method can be used in the GGE biplot graph in a different way than it is presented in Figure 3 (Fig. 5, Yan 2001). The length of the line segment between the origin and the orthogonal projection of the genotype location onto the AE axis, is proportional to the mean genotype value and represents the main effect of the genotype G. Genotype ranking in descending order is as follows: G15 > G12 > G8 > G2 > G18 > G6 > G4 > G11 > G3 > G9 > Mean > G10 > G13 > G5 > G16 > G17 > G19 > G14 > G7 > G1 > G20.

The length of the line segment between the genotype location and its orthogonal projection onto the AE

axis is a measure of genotype stability. In the present study, starting from the most stable, the genotypes G12, G9 and G15 showed the heights above the overall average and the highest stability. G17, G10, G14, G19, G7, G5, G13 genotypes were also stable, but their height was below the overall average. The unstable genotypes can be determined in an analogous way: G8, G2, G18, G6, G4 and G11 with the height above the overall average and G3, G20, G1, G16 – with the height below the overall average.

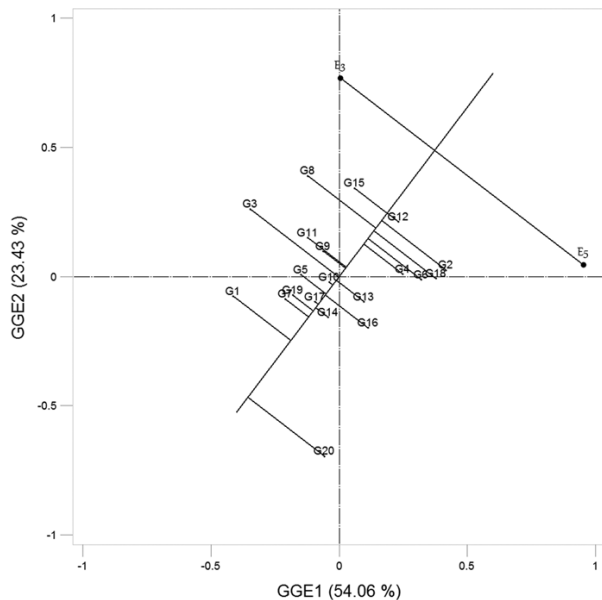


**Figure 5.** Means and stability of genotypes in view of GGE biplot based on principal components

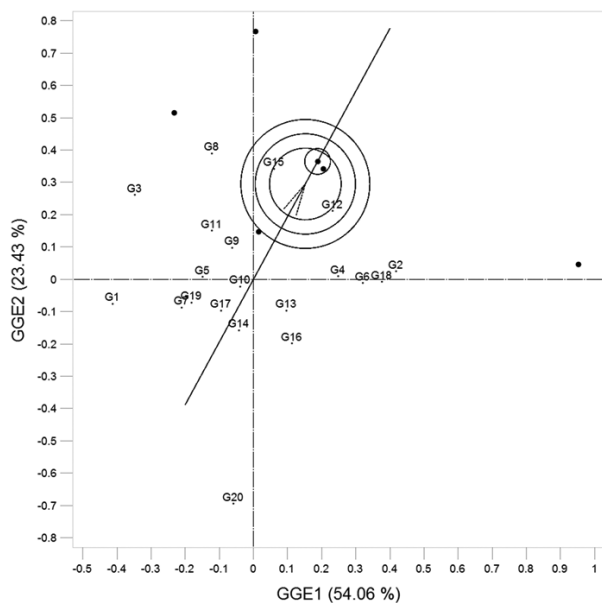
Using the GGE biplot graph, the mean genotype heights in two selected environments can also be compared (Yan 1999). For example, in Fig. 6, the two environments E3 and E5 were selected and linked with a straight line. Then, the origin was linked to the latter with a perpendicular line. The genotypes at the same side of the perpendicular line, where E3 environment was located, achieved greater heights in this environment than those on the other side. G15, G8, G11, G9, G3, G5, G19, G7, G1 genotypes belonged to the group with greater heights. In E5 environment, the greater heights were observed in the genotypes: G2, G18, G6, G4, G13, G16, and G20. On the other hand, in both environments, the genotypes: G10, G17, G12, G14, located close to the perpendicular line, showed no substantial height difference.



rences. The greatest differences were observed in the genotypes distant to the perpendicular line that is in G8, G3 and G20.



**Figure 6.** Comparison between environments E3 and E5, based on GGE biplot



**Figure 7.** Comparison of all genotypes with the ideal genotype

The biplot graphs allow for seeking the ideal genotype that should show the high mean value of the trait

and good stability in the environments tested, bearing in mind that both measures are equally important (Yan 2001). In order to find the ideal genotype, the end of the vector positioned on the AE axis should be determined, as the highest value among the perpendicular projections of genotype locations on the AE axis (Fig. 7). The AE axis determines the zero value of the effect of GE interaction in the average environment AE. To find the genotypes that are located closest to the end of the vector, circles are constructed with centres at points located at the end of the vector on the AE axis and with optional radius. Within the first circle, G15 genotype was located. Hence, this was the ideal genotype in terms of achieving the great mean height and good stability in the environments tested. In the second circle, there G12 genotype was located, with the great mean height, but less stable in comparison with G15.

## SUMMARY AND CONCLUSIONS

The following points sum up the conclusions of the present study:

1. Studies on forests and agriculture, require the analyses of genotype  $\times$  environment interactions. An environment can refer to: research areas (i.e. localities) or observation years with changing weather conditions, and also, localities in the years.
2. The effects of genotype  $\times$  environment interaction can be analyzed based on a two-dimensional table, where the rows constitute mean values for the trait of the objects, and the columns constitute mean values for environments.
3. The biplot graph method enables a simple and clear analysis of genotype  $\times$  environment interaction.
4. The results necessary to construct the biplot graphs can be obtained by means of different calculation methods: SVD, PCA, AMMI, SREG or GREG. The choice of the method depends on the model (models 1–3), hence, the model and the method of the analysis rely on the aims of the biplot graph analysis:
  - a) evaluation of genotypes and their stability with respect to the mean value of the trait (AMMI),
  - b) evaluation of mega-environments (SVD, PCA, SREG, GREG),
  - c) characteristics of environments with high potential to discriminate the genotypes tested, as

well as those with the most representative conditions of growth in mega-environments (SVD, PCA, SREG, GREG),

- d) evaluation of the genotypes in terms of stability and mean value of the trait (SVD, PCA, SREG),
- e) evaluation of genotype  $\times$  environment interaction with the use of the polygon and its sectors, or the AEC method (SVD, PCA, SREG),
- f) comparison of genotype interaction in two selected environments (SVD, PCA, SREG),
- g) search for the ideal genotype which achieves high values of the trait, and is stable in all the environments (SVD, PCA, SREG).

The proposed choice of the models and methods used in the biplot graph analysis can be a helpful tool in studies concerning forest tree selection, as well as serve as inspiration to use this method in other fields of forest sciences. The biplot graphs allow clear demonstration of relationships within genotypes or environments, as well as those between genotypes and environments.

## ACKNOWLEDGEMENTS

The study was partly funded by the State Forests National Forest Holding as part of research project BLP-375.

## REFERENCES

- Balzarini M. 2002. Applications of mixed models in plant breeding. In: Quantitative genetics, genomics, and plant breeding (ed.: M.S. Kang). CABI Publishing, UK, 353–365.
- Bradu D., Gabriel K.R. 1978. The biplot as a diagnostic tool for models of two-way tables. *Technometrics*, 20, 47–68. DOI: 10.1080/00401706.1978.10489617.
- Cooper M., DeLacy I.H. 1994. Relationships among analytic methods used to study genotypic variation and genotype-by-environment interaction in plant breeding Multi-environment experiments. *Theoretical and Applied Genetics*, 88, 561–572. DOI: 10.1007/BF01240919.
- Cornelius P.L., Crossa J. 1999. Prediction assessment of shrinkage estimators of multiplicative models for multi-environment cultivar trials. *Crop Science*, 39, 998–1009. DOI: 10.2135/cropsci-1999.0011183X003900040007x.
- Cornelius P.L., Seyedsadr M. 1997. Estimation of general linear–bilinear models for two-way tables. *Journal of Statistical Computation and Simulation*, 58, 287–322. DOI: 10.1080/00949659708811837.
- Cornelius P.L., Crossa J., Seyedsadr M. 1996. Statistical tests and estimators of multiplicative models for cultivar trials. In: Genotype- by-Environment Interaction (eds.: M.S. Kang, H.G. Gauch, Jr). CRC Press, Boca Raton, Florida, 199–234.
- Correia I., Alia R., Yan W., David T., Aguiar A., Almeida M.H. 2010. Genotype  $\times$  environment interactions in *Pinus pinaster* at age 10 in a multi-environment trial in Portugal: a maximum likelihood approach. *Annals of Forest Science*, 67, 612p1–612p9. DOI: 10.1051/forest/2010025.
- Crossa J. 2012. From genotype  $\times$  environment interaction to gene  $\times$  environment interaction. *Current Genomics*, 13 (3), 225–244. Available at: <http://ejournals.ebsco.com.prxy4.ursus.maine.edu/direct.asp?ArticleID=473EADE10B8406E6D70F>.
- Ding M., Tier B., Yan W., Wu H.X., Powell M.B., McRae T.A. 2008. Application of GGE biplot analysis to evaluate Genotype (G), Environment (E), and G $\times$ E interaction on *Pinus radiata*: a case study. *New Zealand Journal of Forestry Science*, 38 (1), 132–142. Available at: [http://www.scionresearch.com/\\_data/assets/pdf\\_file/0007/5596/NZJFS\\_38\\_12008\\_Ding\\_et\\_al\\_132-142.pdf](http://www.scionresearch.com/_data/assets/pdf_file/0007/5596/NZJFS_38_12008_Ding_et_al_132-142.pdf).
- Eberhart S.A., Russell W.A. 1966. Stability parameter for comparing varieties. *Crop Science*, 6, 36–40. Available at: [http://www.sap.uchile.cl/descargas/fisiogenetica/Stability%20parameters%20for%20comparing%20varieties\\_Eberhart\\_Russell1966.pdf](http://www.sap.uchile.cl/descargas/fisiogenetica/Stability%20parameters%20for%20comparing%20varieties_Eberhart_Russell1966.pdf).
- Finlay K.W., Wilkinson G.N. 1963. The analysis of adaptation in a plant breeding programme. *Australian Journal of Agricultural Research*, 14, 742–754. DOI: 10.1071/AR9630742.
- Gabriel K.R. 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58, 453–467. DOI: 10.1093/biomet/58.3.453.
- Gabriel K.R. 1972. Analysis of meteorological data by means of canonical decompositions and biplots. *Journal of Applied Meteorology*, 11, 1071–1077.

- DOI: [http://dx.doi.org/10.1175/1520-0450\(1972\)011<1071:AOMDBM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1972)011<1071:AOMDBM>2.0.CO;2).
- Gabriel K.R. 1978. Least squares approximation of matrices by additive and multiplicative models. *Journal of the Royal Statistical Society, Series B*, 40, 186–196. Available at: <http://www.jstor.org/prxy4.ursus.maine.edu/stable/2984752>.
- Gauch H.G. 1988. Model selection and validation for yield trials with interaction. *Biometrics*, 44, 705–715. DOI: 10.2307/2531585.
- Gauch H.G. 1992. Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs. Elsevier, Amsterdam, Netherlands.
- Gauch G.H., Zobel R.W. 1997. Interpreting mega-environments and targeting genotypes. *Crop Science*, 37, 311–326. DOI: 10.2135/cropsci1997.0011183X003700020002x.
- Golub G.H., Reinsch C. 1971. The singular value decomposition and least squares solutions. In: Handbook for Automatic Computation (eds.: J.H. Wilkinson, C. Renisch). Springer-Verlag, Berlin, 134–151.
- Hocking R.R., Speed F.M. 1975. A full-rank analysis of some linear model problems. *Journal of the American Statistical Association*, 70, 706–712. DOI: 10.2307/2285959.
- Hotelling H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417–441, 498–520. DOI: 10.1037/h0071325.
- Hotelling H. 1936. Relations between two sets of variates. *Biometrika*, 27, 321–377. DOI: 10.2307/2333955.
- Jolliffe I.T. 1972. Discarding Variables in a Principal Component Analysis. I: Artificial Data. *Applied Statistics*, 21, 160–173. DOI: 10.2307/2346488.
- Jolliffe I.T. 1973. Discarding Variables in a Principal Component Analysis. II: Real Data. *Applied Statistics*, 22, 21–31. DOI: 10.2307/2346300.
- Khattree R., Naik D.N. 2000. Multivariate data reduction and discrimination with SAS software. SAS Institute Inc., Cary, NC.
- Kempton R.A. 1984. The use of biplots in interpreting variety by environment interactions. *Journal of Agricultural Science*, 103, 123–135. DOI: 10.1017/S0021859600043392.
- Kim I., Kwon H., Ryu K., Choi W.Y. 2008. Provenance by Site Interaction of *Pinus densiflora* in Korea. *Silvae Genetica*, 57 (3), 131–139.
- Littell R.C., Milliken G.A., Stroup W.W., Wolfinger R.D. 1996. SAS system for mixed models. SAS Institute Inc., Cary, NC.
- Liu G., Cornelius P.L. 2001. Simulations and derived approximations for the means and standard deviations of the characteristic roots of a Wishart matrix. *Communications in Statistics – Simulation and Computation*, 30, 963–989. DOI: 10.1081/SAC-100107791.
- McCabe G.P. 1984. Principal Variables. *Technometrics*, 26 (2), 137–144. DOI: 10.2307/1268108.
- Murillo O. 2001. Genotype by environment interaction and genetic gain on unbalanced *Pinus oocarpa* provenances trials. *Agronomia Costarricense* 25 (1), 21–32.
- Patterson H.D., Thompson R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58, 545–554. DOI: 10.2307/2334389.
- Patterson H.D., Thompson R. 1975. Maximum likelihood estimation of components of variance. Proceedings of 8th International Biometric Conference, 197–207.
- Pearson K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2 (11), 559–572. DOI: 10.1080/14786440109462720.
- 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., Möhring J. 2006. Selection in cultivar trials – is it ignorable? *Crop Sciences*, 46, 192–201.
- SAS Institute Inc. 2013. SAS/STAT® 13.1 User's Guide. Cary, NC: SAS Institute Inc.
- Saxton A.M. 2004. Genetic analysis of complex traits using SAS. SAS Institute, Inc. Cary, N.C.
- Sixto H., Gil P.M., Ciria P., Camps F., Cañellas I., Voltas J. 2015. Interpreting genotype-by-environment interaction for biomass production in hybrid poplars under short-rotation coppice in Mediterranean environments. *GCB Bioenergy*. DOI: 10.1111/gcbb.12313.
- Taibi K. 2014. Title of dissertation “Integrated approach for addressing assisted population migration programs in forest management to climate change: Out-planting performance, genotype by environment interactions, physiological and molecular response”. The Polytechnic University of Valencia, Spain.

- Ukalska J., Kociuba W. 2013. Phenotypal diversity of winter triticale genotypes collected in the Polish Gene Bank between 1982 and 2008 with regard to major quantitative traits. *Field Crops Research*, 149, 203–212. DOI: 10.1016/j.fcr.2013.05.010.
- Ukalski K., Śmiałowski T., Ukalska J. 2010a. Analysis of oat yield environments using graphical GGE method. *Colloquium Biometricum*, 40, 81–93.
- Ukalski K., Śmiałowski T., Ukalska J. 2010b. Analiza plonowania i stabilności genotypów owsa za pomocą metody graficznej typu GGE. *Żywność. Nauka. Technologia. Jakość*, 3 (70), 127–140.
- Yan W. 1999. A study on the methodology of cultivar evaluation based on yield trial data with special reference to winter wheat in Ontario. Ph.D. thesis, University of Guelph, Guelph, Ontario, Canada.
- Yan W. 2001. GGE biplot: a Windows application for graphical analysis of multi-environment trial data and other types of two-way data. *Agronomy Journal*, 93, 1111–1118. DOI: 10.2134/agronj2001.9351111x.
- Yan W. 2002. Singular-value partitioning in biplot analysis of multi-environment trial data. *Agronomy Journal*, 94, 990–996. DOI: 10.2134/agronj2002.0990.
- Yan W., Cornelius P.L., Crossa J., Hunt L.A. 2001. Two types of GGE biplots for analyzing multi-environment trial data. *Crop Science*, 41, 656–663. DOI: 10.2135/cropsci2001.413656x.
- Yan W., Hunt L.A. 2001. Interpretation of genotype x environment interaction for winter wheat yield in Ontario. *Crop Science*, 41, 19–25. DOI: 10.2135/cropsci2001.41119x.
- Yan W., Hunt L.A., Sheng Q., Szlavnic Z. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Science*, 40, 597–605. DOI: 10.2135/cropsci2000.403597x.
- Yan W., Kang M.S. 2003. GGE biplot analysis: a graphical tool for breeders, genetics and agronomists. CRC Press, Boca Raton, FL.
- Yan W., Rajcan I.R. 2002. Biplot analysis of test sites and trait relations of soybean in Ontario. *Canadian Journal of Plant Science*, 42, 11–20. DOI: 10.2135/cropsci2002.0011.
- Yan W., Tinker N.A. 2005. An integrated biplot analysis system for displaying, interpreting, and exploring genotype-x-environment interactions. *Crop Science*, 45, 1004–1016. DOI: 10.2135/cropsci2004.0076.
- Yan W., Tinker N.A. 2006. Biplot analysis of multi-environment trial data: Principles and applications. *Canadian Journal of Plant Science*, 86, 623–645.
- Zhao X., Xia H., Wang X., Wang C., Liang D., Li K., et al. 2016. Variance and stability analyses of growth characters in half-sib *Betula platyphylla* families at three different sites in China. *Euphytica*, 208, 173–186. DOI: 10.1007/s10681-015-1617-7.
- Zobel R.W., Wright M.J., Gauch H.G. 1988. Statistical analysis of a yield trial. *Agronomy Journal*, 80, 388–393. DOI: 10.2134/agronj1988.00021962008000030002x.