ANALYSIS OF OAT YIELD ENVIRONMENTS USING GRAPHICAL GGE METHOD

Krzysztof Ukalski[1], Tadeusz Śmiałowski [2], Joanna Ukalska[1]

[1] Biometry Division, Department of Econometrics and Statistics,
Warsaw University of Life Sciences-SGGW
Nowoursynowska 159, 02-776 Warsaw
[2] Department of Cereal Breeding and Quality Evaluation
Zawiła 4, 30-423 Cracow

Summary

Yield data of 29 covered grain oat cultivars tested across 6 environments and 14 naked grain oat cultivars tested across 5 environments during the 2008 growing season were analyzed using the GGE (i.e., G, genotype + GEI, genotype-by-environment interaction) biplot method. GGE biplot analysis is an efficient method which is based on principal component analysis (PCA) to fully explore multi-environment trials. It allows a visual examination of the relationships among the test environments. GGE-biplots for the environments were compared with the correlation coefficient respectively within 6 covered grain oat cultivar environments as well as within 5 naked grain oat cultivar environments. So-called ideal environment has been determined. The ideal environment among 6 environments of covered grain oat cultivars was Polanowice (POB). Polanowice (POB) turned out to be also the ideal environment among 5 environments of naked grain oat cultivars.

Key words and phrases: biplot, GGE effects, covered grain oats, naked grain oats, principal component analysis

Classification AMS 2010: 62H25

1. Introduction

Multi-environment yield trials (MEYTs) are used to identify superior genotypes in plant breeding programs. This is not an easy task due to the frequent presence of GEI (Hill et al. 1998, Annicchiarico 2002, Baker 2002; Kang 2002; Yan and Kang 2003). The measured yield of each genotype in each tested environment is a measure of an environment main effect (E), a genotype main effect (G), and the genotype×environment (GE) interaction (Yan and Kang 2003, Dehghani et al. 2006). The presence of GEI determines an immediate use of the means across environments. That is the reason why the differences between genotypes are ignored in the pattern of relative performance over the environments (Voltas et al. 2002, Slafer et al. 2002). Yan et al. (2000) presented standard biplots to select the best performing cultivars in subsets of sites. GGE-biplot, developed by Gabriel (1971), is constructed by the first two principal components (PC1 and PC2) derived from analysis of environment-centered yield data. GGE-biplot can be useful to identify discrimination and representation of test environments (Yan et al. 2001).

Research on genotypes across environments aim not only at identification of superior genotypes, but also at determining if the aimed region can be subdivided into different mega environments (Yan et al. 2000), in other words which environments are similar in a mean of yield of studied genotypes and which environment is ideal (i.e. has the biggest discriminating power in terms of the genotypic main effects and is the most representative among all environments). The GGE biplot methodology has been used to evaluate the test environments in soybean (Yan and Rajcan 2002), winter wheat (Yan and Hunt 2001), cotton (Blanche and Myers 2006), common bean (Kang et al. 2006), wheat (Morris et al. 2004) and in target cultivars within specific rice environments (Samonte et al. 2005).

The aim of this research was to use GGE-biplot in order to find the ideal environment and identify the most discriminating and representative environments for covered grain oat and naked grain oat separately.

2. Materials and methods

29 covered grain oat cultivars were grown in 6 environments (CHD – Kopaszewo, MAH – Małaszyn, POB – Polanowice, SOA – Sobiejuchy, STH – Strzelce, SKR – Skrzeszowice) and 14 naked grain oat cultivars were grown in 5 environments (the same environments except for SKR – Skrzeszowice) during the 2008 growing season. Details of general characterization of 6 environments are presented in the Table 1.

Table 1. Characterization of environments

Place (environment)	Region	Soil		Preceding Crop	Fertilizer	Date		Precipitation in growing season (mm)
		Type			NPK (kg/ha)	Sowing	Crop	
Kopaszewo CHD	Wielkopolskie	Wheat soil IV		Sugar beet	325.7	1.04	30.07	235.6
Małaszyn MAH	Lubuskie	Redzina and sand		Winter triticale	247.0	7.04	4.08	224.8
Polanowice POB	Małopolskie	Less		Sugar beet	181.0	28.03	6.08	315.5
Sobiejuchy SOA	Kujawsko-Pomorskie	Loamy sand III b		Sugar beet	210.0	31.03	30.07	315.5
Strzelce STH	Łódzkie	Brunisolic		Sugar beet	66.0	31.03	25.07	166.1
Skrzeszowice SKR	Małopolskie	Degraded chernozem		Winter wheat	160.0	1.04	12.08	193.4

The data came from the preliminary experiments, which were conducted in incomplete block design with three replications. The two-way ANOVA model was performed for covered and naked grain oat cultivars separately. The linear model for observations obtained in the experiment is as follows:

$$y_{ghk} = \mu + G_g + E_h + B_{k(h)} + GE_{gh} + \varepsilon_{ghk} \;, \tag{2.1}$$

where y_{ghk} is the observation of gth genotype (G_g) in the kth block ($B_{k(h)}$) in the hth environment (E_h), μ is the overall mean, GE_{gh} is interaction effects between main effects and ε_{ghk} represents a random error term.

GGE biplot analysis is an effective method, which is based on a principal component analysis (PCA), to fully explore multi-environment trials. The first two principal components (PC1 and PC2), which had been derived from PCA analysis of values of environment means, were conducted using SAS macro Biplot (Khattree and Naik 2000, SAS/STAT User's Guide 2004). Biplot macro utilizes the singular value decomposition (SVD) of the matrix **X**. SVD can be written using the rank factorization (Rao 1973) as

$$\mathbf{X} = \mathbf{GH'} \,, \tag{2.2}$$

where the *g* by *r* matrix **G** and the *h* by *r* matrix **H** are both of rank *r*. Each x_{ij} is thus expressed as $x_{ij} = \mathbf{g'}_i \mathbf{h}_j$, where $\mathbf{g'}_i$ is the ith row of G and $\mathbf{h}_j$ is the jth column of **H'**. Gabriel (1971) suggested that the *g*+*h* points can be plotted on

the plane giving a representation of the **X**'s *gh* elements when the rank of **X** is *r*=2. Mentioned representation is called a *biplot* of **X**. Hence, the singular value decomposition of **X** is

$$\mathbf{X} = \mathbf{U}\Delta\mathbf{V}' = \sum_{l=1}^{r} \delta_l \mathbf{u}_l \mathbf{v}'_l ,$$ (2.3)

where Δ is the *r* by *r* diagonal matrix with positive diagonal elements $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_r > 0$, **U** is the *g* by *r* matrix with columns $\mathbf{u}_1,\ldots,\mathbf{u}_r$ such that $\mathbf{U}\mathbf{U}' = \mathbf{I}$, and **V** is the *h* by *r* matrix with columns $\mathbf{v}_1,\ldots,\mathbf{v}_r$, such that $\mathbf{V}'\mathbf{V} = \mathbf{I}$. The approximation of dimension two for **X** using the method of least squares is

$$\mathbf{X} \approx \delta_1 \mathbf{u}_1 \mathbf{v}'_1 + \delta_2 \mathbf{u}_2 \mathbf{v}'_2 .$$ (2.4)

From equation 2.3 we get an approximation biplot for **X**, and the corresponding goodness of fit is measured by

$$w_2 = \frac{\delta_1^2 + \delta_2^2}{\sum_{l=1}^{r} \delta_l^2} .$$ (2.5)

The vectorial view of the GGE-biplot shows the interrelationships among the environments. The lines that connect the biplot origin and the markers of the environments are called environment vectors. The angle between the vectors of 2 environments is related to the correlation coefficient between them. Environments that are highly correlated make an angle close to zero (if the environments are positively correlated) or an angle close to 180 degrees (if the environments are negatively correlated). Environments that are weak correlated will make an angle close to 90 degrees. Thus, similarity or dissimilarity of environments could be graphically determined by the biplots by considering the angles.

Correlation coefficients among the environments were computed using SAS procedure CORR and GGE-biplot charts were done using SAS procedure GPLOT (SAS/STAT User's Guide 2004).

3. Results

The ANOVA showed that covered grain oat yield was significantly affected by environments (E), which explained 93.90% of the total (G + E + GEI) variation, whereas G and GEI, which were significant ($p < 0.0001$), accounted for 1.80% and 4.30%, respectively (Table 2). The results of analysis of variance naked grain oat yield are similar. Crucial E, G, GEI ($p < 0.0001$) explained 87.41%, 8.80%, 3.79% of the total variation, respectively (Table 3). Gauch and Zobel (1997) reported that in normal MEYTs E accounts are about 80% of the total variation while G and GE accounts are about 10% each.

Table 2. Analysis of variance of yield data of 29 covered grain oats genotypes across 6 environments

Source of variation	df	SS	MS	p-value	%(G+E+GEI)
Environment (E)	5	163412.14	32682.43	<0.0001	93.90
Block(B)	102	11164.35	109.45		
Genotype (G)	28	3125.54	111.63	<0.0001	1.80
GEI	140	7490.31	53.50	<0.0001	4.30
Error	246	5872.42	23.87		

Table 3. Analysis of variance of yield data of 14 naked grain oats genotypes tested across 5 environments

| Source of variation	df	SS	MS	p-value	%(G+E+GEI)
Environment (E)	4	40513.81	10128.45	<0.0001	87.41
Block(B)	55	2723.44	49.52		
Genotype (G)	13	4078.43	313.73	<0.0001	8.80
GEI	52	1756.41	33.78	<0.0001	3.79
Error	85	710.48	8.36		

The covered grain oat yield analysis of the first two PC's showed that PC1 and PC2 were significant factors, explaining 46.42% and 31.35% of GGE variation, respectively. They, together, accounted for 77.77%. This score was counted according to equation 2.5 and hence

$$w_2 = \frac{40.569^2 + 33.342^2}{40.569^2 + 33.342^2 + 20.177^2 + 15.95^2 + 8.5039^2 + 7.3711^2} = 0.7777 \, .$$

For naked grain oats genotypes the first two PC's explained 81.16% and 8.95% of GGE variation, respectively. Together, they accounted for 90.11%, and were calculated as follows:

$$w_2 = \frac{39.74^2 + 13.2^2}{39.74^2 + 13.2^2 + 11.032^2 + 7.611^2 + 3.5775^2} = 0.9011.$$

Jolliffe (1972) used some simulation studies and suggested that cutoff point of the singular values δ_i for dropping undesirable principal components should be 0.7.
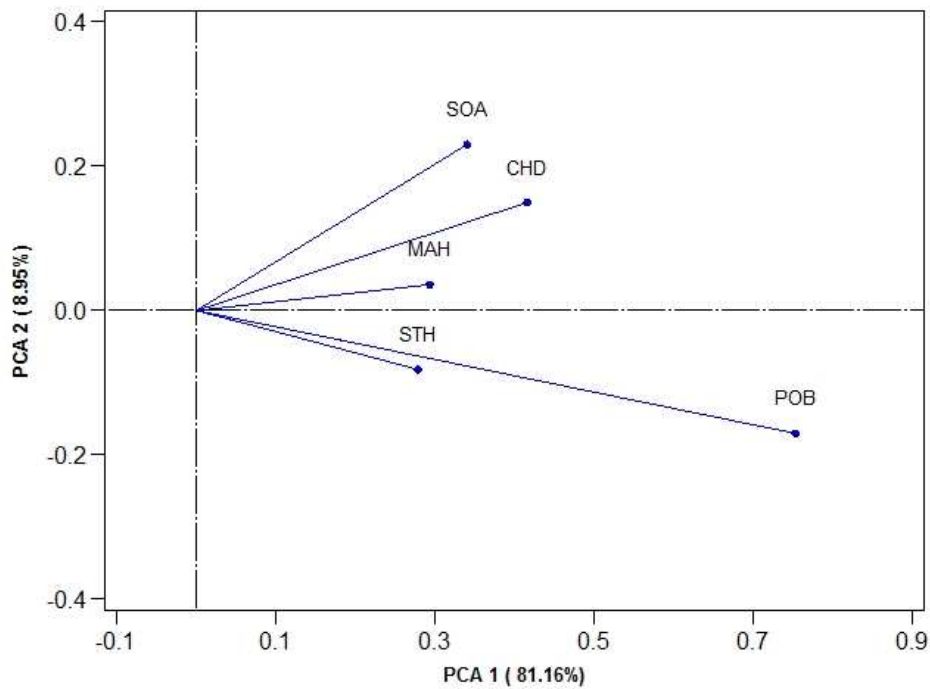
Fig. 1. GGE-biplot based on principal component for 5 environments of naked grain oats genotypes

GGE-biplots, which were based on environment focused scaling, presented the pattern of environments of grain (Figure 1, Figure 2). For naked grain genotypes, environment PC1 resulted only in positive scores (Figure 1). This suggests that PC1 represents proportional genotype yield differences across 5 environments, which leads to a non-crossover GEI. POB is the most discriminating environment of the naked oat genotypes (the biggest score of PC1) but MAH is the most representative of the overall environments (score PC2 is close to zero). While for covered grain genotypes, environment PC1 had both positive and

negative scores (Figure 2) and it is cause of exist a crossover GEI where a significant change in genotype yield occurs from one environment to another are not the same. POB and SKR are the most discriminating environments of the covered oat genotypes but MAH is the most representative again. The first situation (PC1 scores are only positive) makes the identification of genotypes in environments with larger PC1 scores easier (Yan et al. 2000). In contrast to environmental PC1, PC2 had both positive and negative scores (Figure 1, Figure 2) so the crossover GEI was bigger, leading to disproportionate genotype yield differences across environments (Yan et al. 2000).
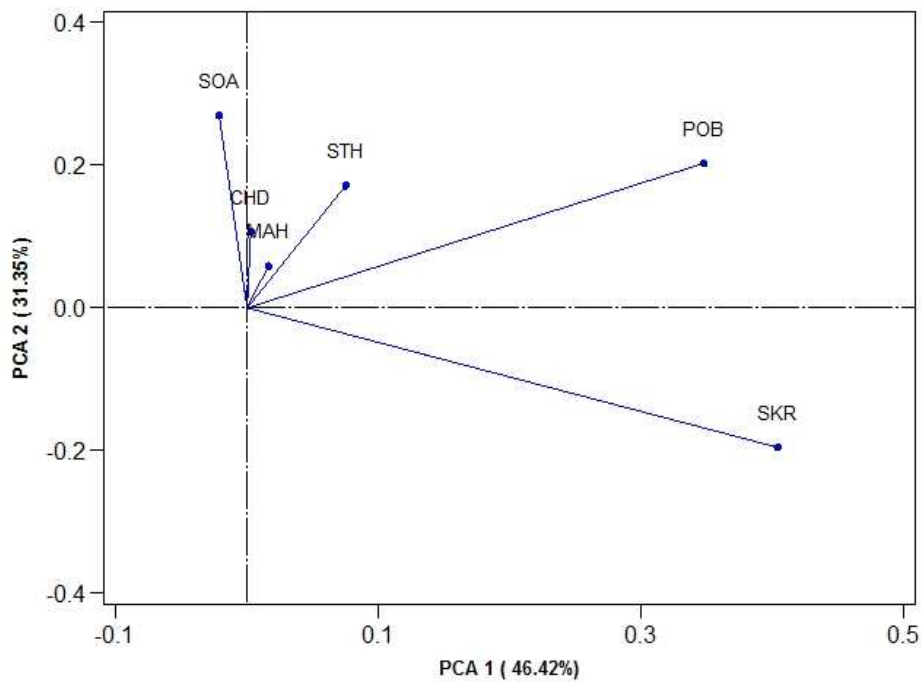
Fig. 2. GGE-biplot based on principal component for 6 environments of covered grain oats genotypes

The correlation coefficients among the 5 test environments of naked grain oats genotypes are presented in Table 4 and among the 6 test environments of covered grain oats genotypes in Table 5.

Table 4. Correlation coefficients (below diagonal) and their p-values (above diagonal) among the
5 test environments of naked grain oats genotypes

Correlation coefficients \ p-value	CHD	MAH	POB	SOA	STH
CHD	-	0.001	0.001	0.001	0.014
MAH	0.778	-	0.004	0.012	<0.0001
POB	0.776	0.715	-	0.008	0.002
SOA	0.779	0.646	0.669	-	0.092
STH	0.639	0.876	0.757	0.467	-

**Table 5.** Correlation coefficients (below diagonal) and their p-values (above diagonal among the
6 test environments of covered grain oats genotypes

Correlation coefficients \ p-value	CHD	MAH	POB	SKR	SOA	STH
CHD	-	0.118	0.050	0.110	0.038	0.011
MAH	0.296	-	0.125	0.856	0.005	0.555
POB	0.367	0.290	-	0.022	0.349	0.017
SKR	-0.302	-0.035	0.422	-	0.199	0.966
SOA	0.385	0.500	0.180	-0.245	-	0.056
STH	0.462	0.114	0.438	-0.008	0.357	-

The vectorial view of the GGE-biplot (Figure 1, Figure 2) shows the inter-relationships among the environments. Table 4 contains 10 correlation coefficients, 9 of them are significant at probability <0.05. All 5 test environments of naked grain oats genotypes were positively correlated, because all angles of among them were smaller than 90° (Figure 1). Table 5 contained 15 correlation coefficients among the 6 test environments of covered grain oats genotypes, only 5 of which were significant at probability < 0.05. The angles between the vectors of environments (for example POB and SKR, POB and STH, SOA and CHD, STH and CHD) which were correlated significantly are smaller then 90° (Figure 2). The rest of vectors of environments of covered grain oats genotypes have angels on the GGE-biplot (Figure 2) smaller (positively correlated, e.g. CHD and MAH), equal (not correlated, e.g. SKR and MAH) or greater (negatively correlated, e.g. CHD and SKR) than 90°. Some discrepancies between angles of mentioned vectors and correlation coefficients are expected because the biplot did not explain 100% of the GGE variation (Yan 2002).

The environments can be described as well discriminating genotypes in terms of the genotypic main effect (the test environment should have large PC1 scores) or representing all the environments (the test environment should have small absolute PC2 scores). Such an ideal environment is represented by an arrow pointing it (Figure 3, Figure 4). An average environment is defined by the average PC1 and PC2 scores of all environments, represented by a small circle and sign E (Figure 3, Figure 4). Thus, using the ideal environment as the center, concentric circles were drawn in order to help to visualize the distance between each environment and the ideal environment (Yan et al. 2000; Yan and Rajcan 2002).
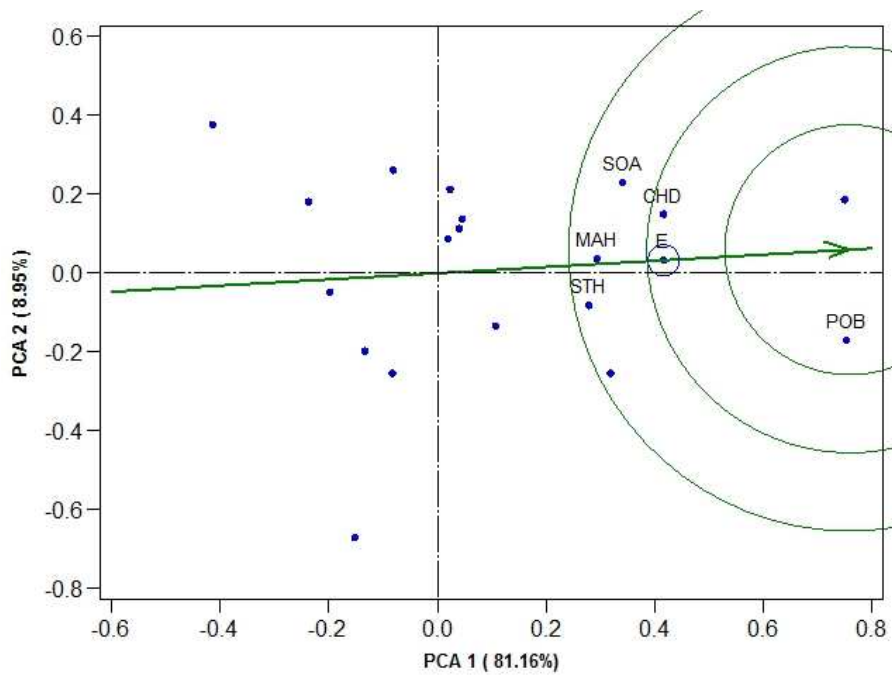
Fig. 3. The comparison of the environments to the ideal environment for naked grain oats genotypes

Among 5 environments of naked grain oat cultivars Polanowice (POB) (Figure 3) is the closest one to the arrow. This environment has a large PC1 score, while PC2 score is small. Among 6 environments of covered grain oats genotypes the closest one is also Polanowice (POB) (Figure 4).
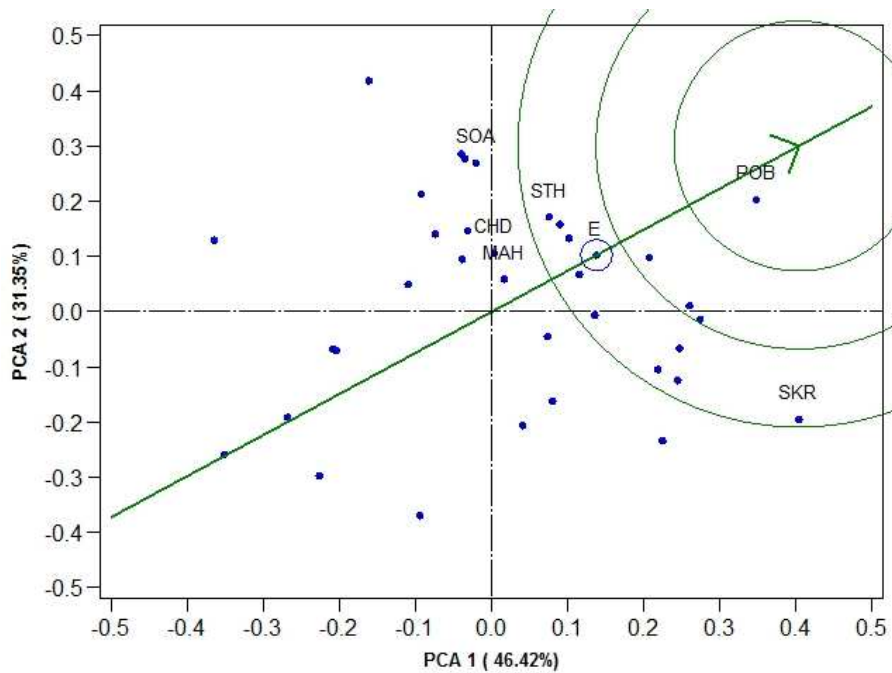
Fig. 4. The comparison of the environments to the ideal environment for covered grain oats genotypes

4. Conclusions

1. The mixed ANOVA showed that covered grain oat yield as well as naked grain oat yield was significantly affected by environments, which explained 93.90% and 87.41% respectively of the total (G + E + GEI) variation.

2. The covered grain oat yield analysis of the first two PC's showed that PC1 and PC2 were significant factors, explaining 46.42% and 31.35% of GGE variation, respectively. Together, they accounted for 77.77%. For 14 naked grain oats genotypes the first two PC's explained 81.16% and 8.95% of GGE variation, respectively. Together, they accounted for 90.11%.

3. All environments of naked grain oats genotypes were positively correlated in spite of they differed in crop conditions. Almost all the test environments of naked grain oats genotypes were significant at probability <0.05 except the correlation among SOA and STH environments. This pair of environments had the biggest angle on the GGE-biplot but it was smaller than 90°. The correlation among STH and MAH was the strongest and this pair of environ-

ments had small angle on the GGE-biplot and similarly PC1 scores (e.i. naked grain oats genotypes had similar yield in these two environments).

4. The vectors of test environments of covered grain oats genotypes had angles on the GGE-biplot smaller, equal or grater than 90°. Only 5 correlation coefficients among 6 environments of covered grain oats genotypes were significant at probability < 0.05. All environments of covered grain oats genotypes could be clustered into three groups according to the GGE-biplot and correlation coefficients. SOA, CHD and STH formed the first group, they had small angles on the GGE-biplot and similar PC2 scores (e.i. they had crossover interaction on the similar level). The second group was formed by POB and SKR. This pair of environments had small angle on the GGE-biplot and similar PC1 scores. Last group consisted of MAH environment in which covered grain oats genotypes achieved small yield but MAH was the most representative of the overall environments.

5. So-called ideal environment has been determined. The ideal environment among environments of covered grain oat cultivars and environments of naked grain oat cultivars was Polanowice (POB). Polanowice (POB) had a high PC1 score, hence a good discrimination of the genotypes. PC2 score was also high so the POB cannot be a representative environment.

References

Annicchiarico P. (2002). Genotype x Environment Interactions − Challenges and Opportunities for Plant Breeding and Cultivar Recommendations. FAO, Rome. *Plant Production and Protection Paper.* 174.

Baker R.J. (2002). Genotype-environment interactions.

    http:/homepage.usask.ca/~rjb609/gxe.html.

Blanche S.B., Myers G.O. (2006). Identifying discriminating locations for cultivar selection in Louisiana. *Crop Sci.* 46, 946–949.

Dehghani H., Ebadi A., Yousefi A. (2006). Biplot Analysis of Genotype by Environment Interaction for Barley Yield in Iran. *Agron J.* 98, 388-393.

Gabriel, K.R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58, 453-467.

Gauch G.H., Zobel R.W. (1997). Interpreting mega-environments and targeting genotypes. *Crop Sci.* 37, 311–326.

Hill J., Becker H.C., Tigerstedt P.M.A. (1998). *Quantitative and ecological aspects of plant breeding.* Chapmann and Hall, London.

Jolliffe I.T. (1972). Discarding Variables in a Principal Component Analysis. I: Artifical Data. *Applied Statistics* 21, 160–173.

Kang M.S. (2002). Genotype-environment interactions: Progress and procpects. In: M.S. Kang (Ed), *Quantitative Genetics, Genomics and Plant Breeding*, CAB International Wallingford, UK 221–243.

Kang, M.S., Aggarwal V.D., Chirwa R.M. (2006). Adaptability and stability of bean cultivars as determined via yield-stability statistic and GGE biplot analysis. *J. Crop Improvement* 15, 97–120.

Khattree R., Naik D.N. (2000). *Multivariate data reduction and discrimination with SAS software*. SAS Institute Inc., Cary, NC.

Morris C.F., Campbell K.G., King G.E. (2004). Characterization of the end-use quality of soft wheat cultivars from the eastern and western US germplasm pools. *Plant Genet. Resources* 2, 59–69.

Rao C.R., (1973). *Linear Statistical Inference and Its Applications*. John Wiley&Sons, Inc., New York.

Samonte S.O.P.B., Wilson L.T., McClung A.M., Medley J.C. (2005). Targeting cultivars onto rice growing environments using AMMI and SREG GGE biplot analysis. *Crop Sci.* 45, 2414–2424.

SAS/STAT User's Guide, Version 9.1. (2004). SAS Institute, Cary NC.

Slafer G.A., Molina-Cano J.L., Savin R., Araus J.L., Romagosa I. (2002). *Recent Advances from Molecular Biology to Agronomy of Yield and Quality*. CRC Press.

Voltas J., Eeuwijk F., Igartua E., Garcia del Moral L.F., Molina-Cano J.L., Romagosa I. (2002). Genotype by environment interaction and adaptation in barley breeding: basic concepts and methods of analysis. In: *Barley science recent advances from molecular biology to agronomy of yield and quality* (Eds.: G.A. Slaver, J.L. Molina-Cano, R. Savin, J.L. Araus and I. Romagosa). The Haworth Press, NY, 205–241.

Yan W. (2002). Singular-value partitioning in biplot analysis of multienvironment trial data. *Agron. J.* 94, 990–996.

Yan W., Cornelius P.L., Crossa J., Hunt L.A. (2001). Two types of GGE biplots for analyzing multi-environment trial data. *Crop Sci.* 41, 656–663.

Yan W., Hunt L.A. (2001). Interpretation of genotype x environment interaction for winter wheat yield in Ontario. *Crop Sci.* 41, 19-25.

Yan W., Hunt L.A., Sheng Q., Szlavnics Z. (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci.* 40, 597–605.

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. (2002). Biplot analysis of test sites and trait relations of soybean in Ontario. *Crop Sci.* 42, 11–20.

ANALIZA ŚRODOWISK UPRAWY OWSA ZA POMOCĄ METODY GRAFICZNEJ TYPU GGE

Streszczenie

Do analizy plonu 29 rodów owsa oplewionego w 6 środowiskach (miejscowościach) i 14 rodów owsa nieoplewionego w 5 środowiskach wykorzystano metodę graficzną biplot typu GGE, (G - efekt główny genotypów + GEI efekt interakcji genotypowo-środowiskowej). Metoda graficzna biplot typu GGE, oparta na wynikach analizy składowych głównych PCA, pozwoliła na porównanie plonu rodów z wielu miejscowości (środowisk). Biploty typu GGE dla miejscowości porównano ze współczynnikami korelacji pomiędzy miejscowościami dla plonu rodów owsa oplewionego i nieoplewionego odrębnie. Wyznaczono tzw. „idealne środowisko". Idealnym środowiskiem, spośród 6 miejscowości, w których badano rody owsa oplewionego były Polanowice (POB), które okazały się również idealnym środowiskiem dla rodów owsa nieoplewionego.

Słowa kluczowe: analiza składowych głównych, biplot, efekty GGE, owies nieoplewiony, owies oplewiony

Klasyfikacja AMS 2010: 62H25