The analysis of usefulness of locations on the basis of spring barley trials

Bogna Zawieja[1], Andrzej Bichoński[2], Jolanta Grala-Michalak[3]

[1]Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, 60-637 Poznań, Poland, e:mail: bogna.zawieja@up.poznan.pl
[2]Malopolska Breeding Company Polanowice, Zbożowa 4, 30-002 Kraków, Poland
[3]Faculty of Mathematics and Computer Science, Adam Mickiewicz University in Poznań, Umultowska 87, 61-614 Poznań, Poland

SUMMARY

The purpose of breeding experiments is to predict the best yielding lines to be registered. Unfortunately, the results obtained in different locations and years are often different. The main objective of this study was the evaluation and choice of experimental locations. The methods used included ANOVA, Andrews' curves, PCA, cluster analysis, coefficients of usefulness and heritability coefficients. The experimental data are derived from pre-preliminary and preliminary breeding experiments with spring barley (*Hordeum vulgare* L.), malting and fodder, conducted in the period from 2008 to 2013 at six experimental stations in Poland. The results showed that some of the locations were similar in respect of the analyzed coefficients, while some locations were unique. The most valuable locations were indicated as those which have the greatest contribution to the interaction and the greatest usefulness (the lowest joint usefulness coefficient). This is because, at the last stage of new variety cultivation, when new varieties are to be registered, they are evaluated in more variable experimental environments.

Key words: Andrews' curves, breeding trials, new genotypes, multivariate analysis, PCA.

1. Introduction

The purpose of breeding experiments is to predict the best yielding lines to be registered. The results obtained in different locations and years are often different, and so breeders need to pay special attention to environmental testing. The breeding trials, called pre-preliminary and preliminary, are usually conducted in several locations with contrasting environmental conditions (diverse climate, soil and others). In the pre-preliminary trials a selection of breeding lines is made.

Next, the selected lines are sown in the preliminary trials together with lines from different breeding stations. In a given year of research, in each of the locations the same lines are sown. However, many of these lines are sown only in one year, which is a disadvantage of these experiments. In subsequent years only control varieties are repeated (although it may happen that one or two of these varieties are changed, via the guidelines for conducting experiments at all locations). In recent years only preliminary trials have been required.

This paper presents a method of choosing the best location for breeding experiments. Considerations are based on measurements of yield (dt/ha) in experiments with spring barley breeding lines, conducted in the years 2008 to 2013.

Barley (*Hordeum vulgare L.*) is one of the oldest domesticated cereal crops. It is one of seven internationally grown cereal grains, currently ranking fourth in world production, behind maize (*Zea Mays L.*), wheat (*Triticum aestivum L.*) and rice (*Oryza sativa*) and ahead of sorghum (*Sorgum Moench*) and rye (*Secale cereale*) (according to the FAO Statistical Yearbook 2016). It is more productive and its yield is less variable than wheat. Therefore, it is widely used amongst farmers with limited and poor resources in less favorable climatic and soil conditions.

2.   Material and Methods

The experimental data are derived from pre-preliminary and preliminary breeding experiments with spring barley (malting and fodder) conducted in the period from 2008 to 2013 at six experimental stations in Poland (Figure 1). In Table 1 the numbers of studied breeding lines are listed. Table 2 shows the average yields over the years in question.

The average yields shown in Table 2 are illustrated in Figure 2. The columns (Fig. 2) show differences in yields as an effect of the year for the breeding lines (left panel) and the differences in crops as an effect of the breeding lines for years (right panel).

Figure 1. The placement of experimental stations (in Poland, the Middle-East Europe)

Table 1. Characteristics of data (numbers of replications in brackets)

Year	Location					
	M1	M2	M3	M4	M5	M6
2008	B (66)	B (66)	B (66)	B (66)	B (66)	B (66)
	P (103)	P (103)	P (103)	P (103)	P (103)	P (103)
2009	B (79)	B (79)	B (79)	B (79)	B (79)	B (79)
	P (114)	P (114)	P (114)	P (114)	P (114)	P (114)
2010	B (21)	B (21)	B (21)	B (21)	B (21)	B (21)
	P (38)	P (38)	P (38)	P (38)	P (38)	P (38)
2011	B (63)	-	B (63)	B (63)	B (63)	B (63)
	P (114)		P (114)	P (114)	P (114)	P (114)
2012	B (67)	B (28)	B (67)	B (67)	B (67)	B (67)
	P (133)	P (45)	P (133)	P (133)	P (133)	P (133)
2013	B (30)	B (30)	B (30)	-	B (30)	B (30)
	P (45)	P (45)	P (45)		P (45)	P (45)

P – the fodder barley breeding lines;   B – the malting barley lines;

Table 2. Mean yields in years

	2008			2009			2010		
Location	B	P	All	B	P	All	B	P	All
M1	56.4	58.0	57.4	59.1	58.3	57.4	64.0	65.5	65.0
M2	84.3	83.5	83.8	63.7	63.1	83.8	38.0	42.3	40.8
M3	57.1	44.2	49.3	62.2	65.7	49.3	72.4	68.5	69.9
M4	83.7	85.8	85.0	69.2	64.0	85.0	65.4	65.4	65.4
M5	80.3	74.1	76.5	76.3	73.5	76.5	66.6	68.4	67.7
M6	65.6	61.5	63.1	66.0	65.3	63.1	84.3	75.9	78.9
	2011			2012			2013		
Location	B	P	All	B	P	All	B	P	All
M1	38.7	36.8	37.5	59.4	59.2	59.2	61.6	61.5	61.5
M2	-	-	-	83.4	86.0	85.0	96.4	88.4	91.6
M3	61.1	48.0	52.7	56.3	74.8	68.6	62.6	58.0	59.9
M4	73.5	75.5	74.8	66.6	66.4	66.4	-	-	-
M5	76.7	83.3	80.9	84.3	83.7	83.9	72.6	75.9	74.6
M6	80.3	85.5	83.6	85.5	89.4	88.0	67.7	70.2	69.2

P – the fodder barley breeding lines;   B – the malting barley lines

Successive rows contain yields of malting barley (first row), yields of fodder barley (middle row) and mean yield for fodder and malting combined (last row).

Figure 3 shows histograms for all locations. At locations M1, M2 and M6 great variability of means is observed. At M3 there is small dispersion and small variability, while M5 exhibits high variation of data and small variability of means. Finally, high variation of data and high variability of means occur at M4.

The highest variation in yield over the years of the study is recorded at M2 (Fig. 3), and the lowest variation at M1 (for both malting and fodder lines). Other locations have similar results. The smallest differences between crops at different locations are found in 2009, and the highest in 2011 (and for malting lines in 2010 also). The highest yields were obtained in the year 2008 at the location M4, in the year 2009 at M5, and in the next three years (2010–2012) at M6. The average for all years varies quite markedly (23.03 dt/ha) at M2 (right panel), while the average for all locations (left panel) varies by only 9.44 dt/ha. This means that fluctuations of yields are greater among locations than among years.

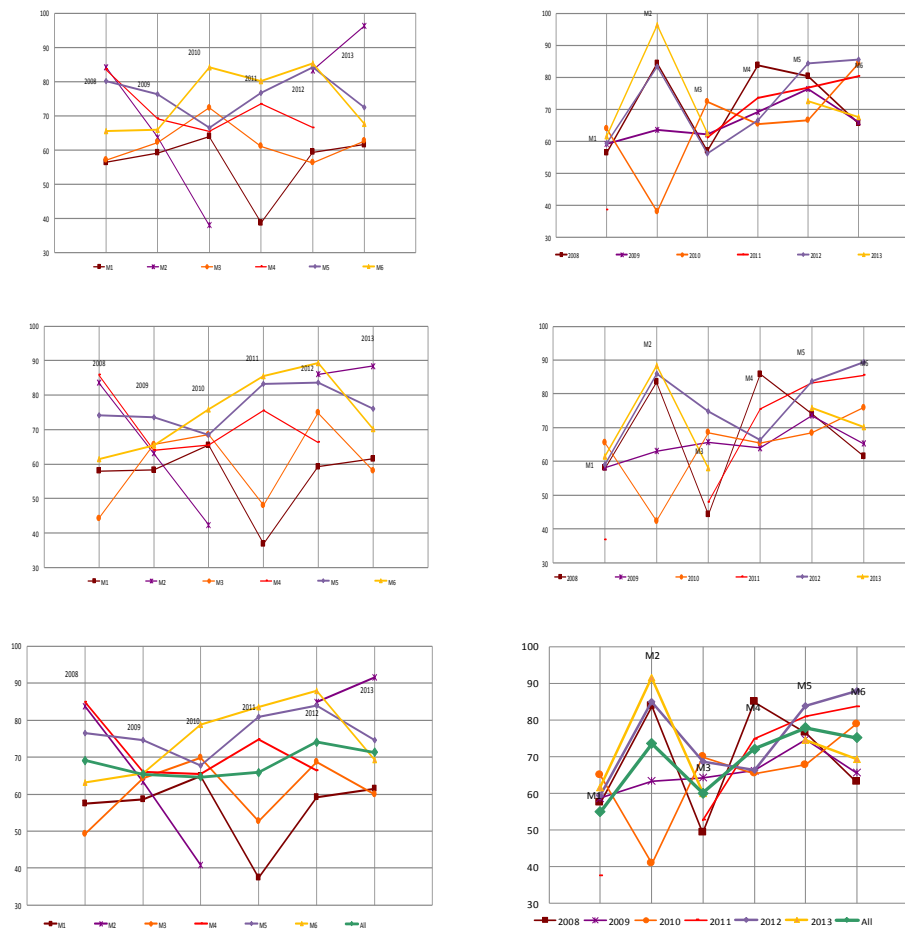Observations in the locations are presented on the box plot figure and surface plot (Fig. 4).

Figure 2. The mean yields in locations (left panel) and the mean yields in years (right panel); Form the top: Malting barley, Fodder barley and all types together
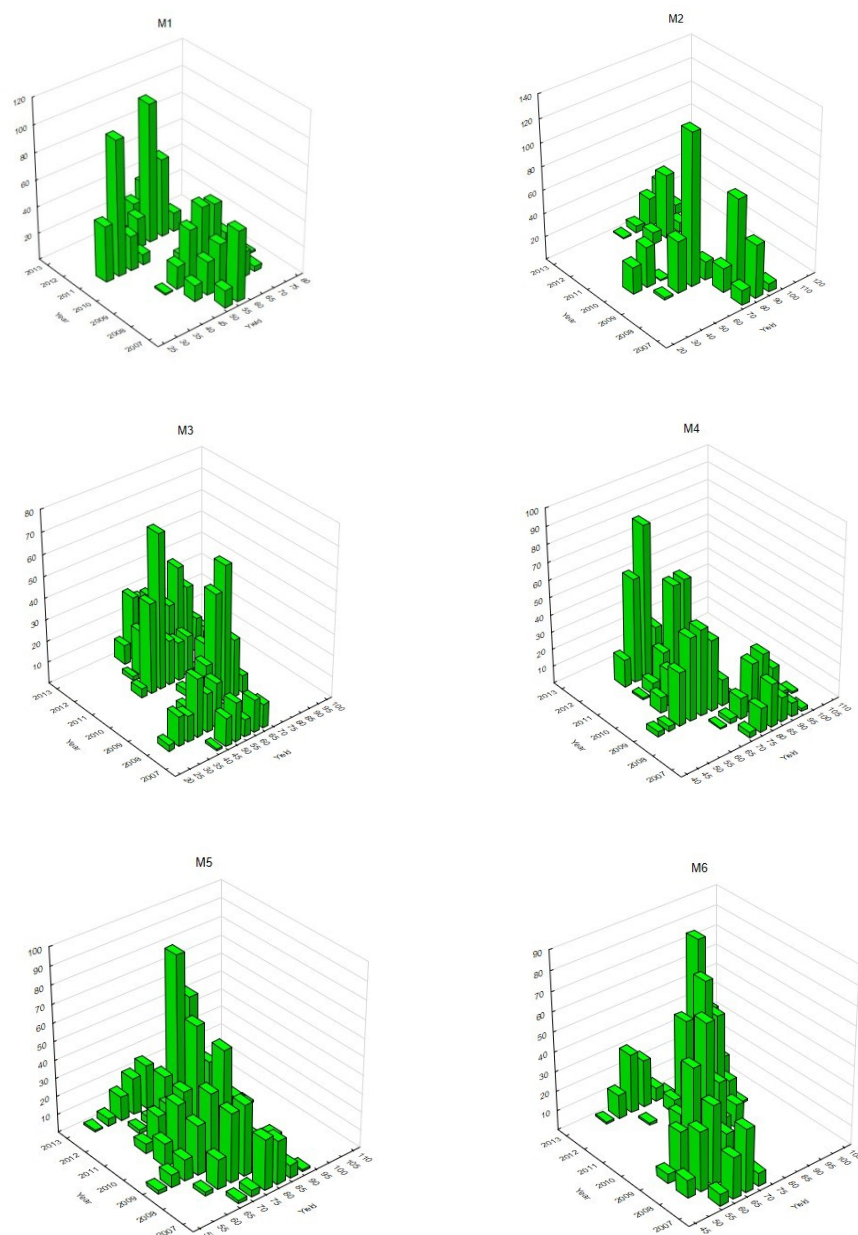
Figure 3. Histograms for locations. From the top: malting barley, inside: fodder barley, at the bottom: all types together
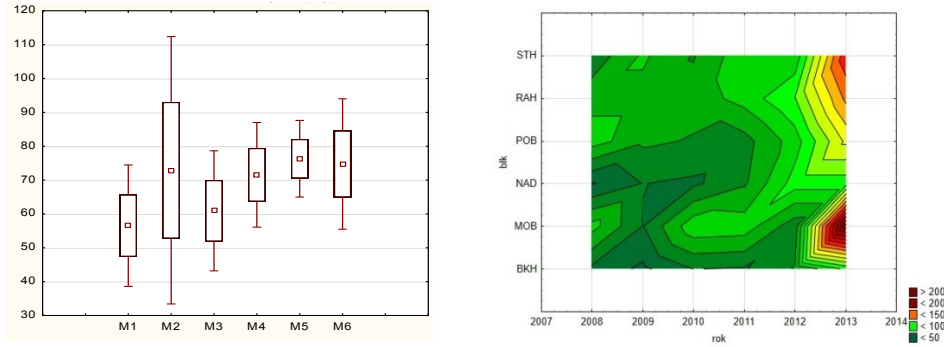
Figure 4. a) Box plots for the mean yields ($\pm$ standard deviations) and whisker plots for confidence intervals; b) The surface plot of mean yields in locations

3. Methods

3.1. ANOVA

To compare the environments in which the breeding experiments were performed, analysis of variance is carried out. In this analysis three factors are taken into consideration: year, location (environment) and type of experiment (with fodder or malting barley lines). The aim of this analysis is to determine whether the types of barley lines differ significantly. If this hypothesis is rejected, then both types of barley lines will be analyzed together. The model has the following form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{y}$ is the vector of N observations, $\boldsymbol{\beta}$ is an unknown vector of fixed-effects parameters with known design matrix $\mathbf{X}$, $\boldsymbol{\gamma}$ is an unknown vector of random-effects parameters with known design matrix $\mathbf{Z}$, and $\boldsymbol{\varepsilon}$ is an unknown random error vector whose elements are independent and homogeneous.

If the analysis of variance is conducted across years, then location, types (malting and fodder) and all interactions between these effects are considered as fixed effects, whereas year and its interactions with fixed effects are considered as random. If the effect of type proves to be irrelevant, then it can be omitted in further considerations.

When analysis of variance is conducted separately for each year, genotypes and types are assumed to be fixed, but location and genotype $\times$ environment (location) are considered as random.

3.2. Visual methods

- Andrews' curves.

The idea of coding and representing multivariate data by curves was suggested by Andrews (1972).

Each multivariate observation $X_i = (X_{i,1},...,,X_{i,p})$ is transformed into a curve as follows:

$$f_i(t) = \begin{cases} \dfrac{X_{i,1}}{\sqrt{2}} + X_{i,2}\sin t + X_{i,3}\cos t + ... + X_{i,p-1}\sin\left(\dfrac{p-1}{2}t\right) + X_{i,p}\cos\left(\dfrac{p-1}{2}t\right) & \text{for } p \text{ odd,} \\ \dfrac{X_{i,1}}{\sqrt{2}} + X_{i,2}\sin t + X_{i,3}\cos t + ... + X_{i,p}\sin\left(\dfrac{p}{2}t\right) & \text{for } p \text{ even,} \end{cases} \quad (2)$$

such that the observation represents the coefficients of Fourier's series for $t \in [-\pi, \pi]$. Outliers appear as single Andrews' curves that look different from the rest.

In our study, the role of $X_i = (X_{i,1},..., X_{i,6})$ is played by the mean yields in the years 2008–2013 years for the ith location, represented by the curves

$$f_i(t) = \frac{X_{i,1}}{\sqrt{2}} + X_{i,2}\sin t + X_{i,3}\cos t + X_{i,4}\sin 2t + X_{i,5}\cos 2t + X_{i,6}\sin 3t, \quad (3)$$

for $t \in [-\pi, \pi]$.

The order of the variables plays an important role. The last variables make only a small visible contribution to the curve, because they fall into the high-frequency part of the curve. To overcome this problem, Andrews suggested using the order obtained from Principal Component Analysis (PCA). Since means are considered as observations in time, it is proper to take only an increasing or decreasing order of years.

- Modified Andrews' curves (Khattree and Naik, 2002).
- Andrews' curves were modified to the following form

$$g_i(t) = \begin{cases} \dfrac{1}{\sqrt{2}}\{X_{i,1} + X_{i,2}(\sin t + \cos t) + X_{i,3}(\sin t - \cos t) + \ldots \\ \qquad + X_{i,p}\left[\sin\left(\dfrac{p-1}{2}t\right) - \cos\left(\dfrac{p-1}{2}t\right)\right]\} & \text{for } p \text{ odd,} \\[2mm] \dfrac{1}{\sqrt{2}}\{X_{i,1} + X_{i,2}(\sin t + \cos t) + X_{i,3}(\sin t - \cos t) + \ldots \\ \qquad + X_{i,p}\left[\sin\left(\dfrac{p}{2}t\right) + \cos\left(\dfrac{p}{2}t\right)\right]\} & \text{for } p \text{ even.} \end{cases}$$

Then the mean yields in 2008–2013 for the ith location are represented by the curves

$$g_i(t) = \frac{1}{\sqrt{2}}[X_{i,1} + X_{i,2}(\sin t + \cos t) + X_{i,3}(\sin t - \cos t) + X_{i,4}(\sin 2t + \cos 2t) + \\ + X_{i,5}(\sin 2t - \cos 2t) + X_{i,6}(\sin 3t + \cos 3t)]. \tag{4}$$

Andrews' method leads to a set of curves on a plane. Similarity between the shapes of the curves suggests similarity between environmental conditions.

- PCA (for means)

Principal component analysis (PCA) is a method in which a multidimensional space is projected onto a plane spanned by two variables being linear combinations of variables (Härdle and Simar, 2003). The multidimensional correlations among locations can be presented in the PCA plane.

- Cluster analysis (for means)

Similarity among the locations can be sought using the cluster method (Härdle and Simar, 2003). The single linkage clustering method with Euclidean distance was applied to the data. Data were represented by mean yields of barley for the years 2008–2013 years at the 6 locations (M1, …, M6).

3.3. Methods based on ANOVA according to Śmiałowski and Węgrzyn (2001)

- Coefficient of usefulness of the jth location in the lth year

On the basis of the ANOVA model (1), the following coefficients can be determined separately for all years:

$$w_{lj} = \sqrt[3]{\left| r_{lj} \right| \hat{h}_{lj} e^{-CV_{lj}/100}} \quad , \tag{5}$$

where r_{lj} is the coefficient of genetic correlations determined on the basis of model (1) separately in each year ($l = 1,...,r$),

$$r_{lj} = \frac{\left(\sum_{i=1}^{n_l} a_{li}^2 + \sum_{i=1}^{n_l} a_{li} g_{lij} \right)}{\sqrt{\left(\sum_{i=1}^{n_l} a_{li}^2 \right) \left[\sum_{i=1}^{n_l} \left(a_{li} + g_{lij} \right)^2 \right]}} \quad , \tag{6}$$

where a_{li} are the ith genotype effect estimators, and g_{lij} are the interaction effect estimators of the ith genotype in the jth location. Next, $\hat{h}_{lj}^2$ are heritability (genetic determination) coefficients in a broad sense (Mądry et al., 2010; Zhe et al., 2010), and CV_{lj} are the coefficients of variation (see Śmiałowski and Węgrzyn, 2001).

The usefulness coefficient of the jth location over all years is measured by

$$w_j = \sqrt[3]{\left| r_j \right| \hat{h}_j e^{-CV_j/100}} \quad , \tag{7}$$

where $\hat{h}_j$ and CV_j are as above but are determined on the basis of model (1) for all years. The genetic correlations r_j are determined separately for each year using the formula

$$r_j = \frac{\sum_{l=1}^{r} \left(\sum_{i=1}^{n_l} a_{li}^2 + \sum_{i=1}^{n_l} a_{li} g_{lij} \right)}{\sqrt{\left(\sum_{l=1}^{r} \sum_{i=1}^{n_l} a_{li}^2 \right) \left[\sum_{l=1}^{r} \sum_{i=1}^{n_l} \left(a_{li} + g_{lij} \right)^2 \right]}}$$

(see Śmiałowski and Węgrzyn, 2001). The smaller the coefficient of usefulness, the more useful is the location.

- Method based on the division of sum of squares for genotype×environment interaction

On the basis of ANOVA models designated both jointly and separately for years, the sum of squares for genotype×environment interaction can be calculated. In the following, the part for each location can be designated in the sum of squares.

This can be done separately in each year of research as well as jointly for all years. The location having the highest contribution to the interaction is considered to be the best.

4. Results

4.1. ANOVA

The hypotheses of the equality of mean yields of barley (malting and fodder) and the equality of mean yields among genotypes, years and locations were verified by ANOVA using model (1). The effects of years, year $\times$ genotype interaction, year $\times$ location interaction and year $\times$ type interaction are random. The type $\times$ location interaction is fixed (see Table 3). As is to be expected based on the results shown in Figure 2, the type effect is not significant (p=0.4425, in Table 3).

Table 3. ANOVA 1

Source	Effect (S/L)	df Effect	MS Effect	df Error	MS Error	F	P
Type	Fixed	1	46.74	6.83	70.34	0.66	0.4425
Location	Fixed	5	20610	22.91	4760.37	4.33	0.0064
Year	Random	5	1635.38	25.66	2000.84	0.82	0.5485
Genotype	Fixed	609	59.54	179.11	30.67	1.94	<.0001
Type $\times$ Location	Fixed	5	44.38	1178	29.73	1.49	0.1893
Type $\times$ Year	Random	5	91.43	1178	29.73	3.08	0.0092
Location $\times$ Year	Random	23	3596.49	1178	29.73	120.96	<.0001
Genotype $\times$ Year	Random	162	30.72	1178	29.73	1.03	0.3791
Genotype $\times$ Location	Fixed	2844	31.43	1178	29.73	1.06	0.1313
Genotype $\times$ Type	Fixed	6	8.998	1178	29.73	0.30	0.9357

Since the effect of type proved to be negligible, it is not necessary to distinguish between fodder and malting barley in the further analysis. Therefore the fixed effect of type is omitted from the model (1); see Table 4. The random effect of year cannot be omitted, because its interactions are significant (with the exception of the interaction with the genotype effect). The main item of interest in this study – the location effect – is significant, although its interactions are not significant.

In the joint analysis of malting and fodder barley (Table 4) the fixed effects of location and its interaction with years are significant.

Table 4. ANOVA 2

Source	Effect (S/L)	df Effect	MS Effect	df Error	MS Error	F	p
Location	Fixed	5	25374	22.85	5819.56	4.36	0.0062
Year	Random	5	2312.68	23.31	2601.59	0.89	0.5043
Genotype	Fixed	609	69.95	203.79	43.81	1.60	<.0001
Location × Year	Random	23	3625.85	1195	30.40	119.29	<.0001
Genotype × Year	Random	177	39.53	1195	30.40	1.30	0.0080
Genotype × Location	Fixed	2844	31.47	1195	30.40	1.04	0.2400

Because the mean yields differ significantly (see Table 4) across years, the analysis of variances is performed for each year separately. Significant differences are determined using Tukey's method. Two objects are in the same group if and only if their confidence intervals projected onto the real number axis have a nonempty common part.

Table 5. Homogenous groups for means in years

Year	Homogeneous groups	Increasing order of means
2008	(**M3**) (**M1**) (*M6*) (*M5*) (<u>M2</u>, *M4*)	**M3, M1,** *M6, M5,* <u>M2</u>, *M4*
2009	(**M1**) (<u>M2</u>, **M3**, *M4, M6*) (*M5*)	**M1,** <u>M2</u>, **M3,** *M6, M4, M5*
2010	(<u>M2</u>) (**M1**, *M4, M5*, **M3**) (*M6*)	<u>M2</u>, **M1,** *M4, M5,* **M3,** *M6*
2011	every object in different group	**M1, M3,** *M4, M5, M6*
2012	(**M1**) (*M4*) (**M3**) (<u>M2</u>, *M5, M6*)	**M1,** *M4,* **M3,** *M5,* <u>M2</u>, *M6*
2013	(**M3, M1**) (*M6*) (*M5*) (<u>M2</u>)	**M3, M1**, *M6, M5,* <u>M2</u>

The ordering of locations by increasing mean yield is the same in 2008 and 2013 (see Table 5). At location M1 the mean yield was the smallest or second smallest in all years. In the period 2010–2012 the highest mean yield was observed at M6, although in the remaining years this was a middle-ranked location. The position of M2 in the ranking changed across the years (underlined font in Table 5).

4.2. Principal Component Analysis of location

PCA was performed for the mean yield in all locations (variables). The measure of how well the first principal components explain variation is given by the cumulative relative proportion of variance. The first three main components explained 88.54% of the total variability (Figure 5); thus the 3-dimensional space is the most informative projection of the data.
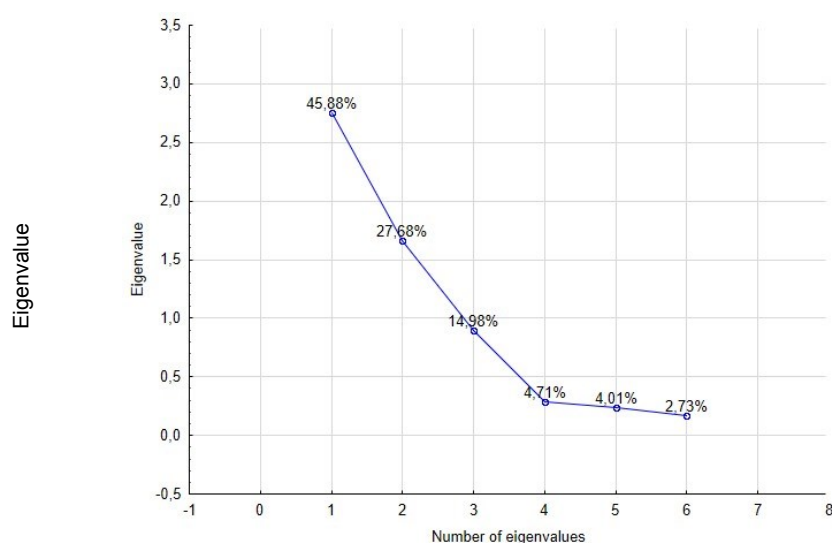

Figure 5. Scree plot of PCA

In the plane of the first two components (Fig. 6) the locations M1, M2, M3, M4 and M5 are placed at the vertices of a regular pentagon. This suggests that the yields in 2008–2013 at different locations are not correlated with each other. The only exception is M2, which is not correlated with the second component. Moreover, it is noticed that the yields at M1 and M5 or at M3 and M4 react in opposite ways to environmental conditions. The same result is obtained from the values of loadings for factor 1 (see Table 6), which have similar absolute value but different signs. The main role in factor 2 is played by the value of loading for M6 (–0.903663). The second (1x3) projection shows that M3 and M4 are

collinear, and M2 and M6 nearly collinear. A similar inference leads to the conclusion that M4 and M5 are collinear on a (2x3)-plane (Fig. 6).
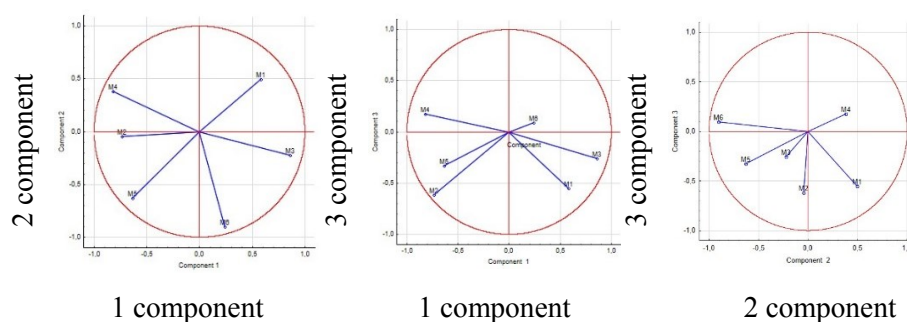

Figure 6. Three (1x2, 1x3 and 2x3) projections of 3-dimensional space of principal components onto the plane

Table 6. Factor loadings

Location	Factor 1	Factor 2	Factor 3
M1	0.581087	0.497349	-0.552877
M2	-0.736877	-0.044171	-0.616830
M3	0.857784	-0.221634	-0.256216
M4	-0.821130	0.383596	0.175970
M5	-0.635766	-0.631393	-0.328202
M6	0.241100	-0.903663	0.092718

The greatest variety of yields was observed at M2, even though in PCA missing values in 2011 were replaced with mean values, which would be expected to reduce the variability.

4.3. Cluster analysis for years and genotypes as replications

Cluster analysis was performed using the simple linkage method and the Euclidean metric. The result of the analysis (Fig. 7) divides locations into three groups: (M1, M3) (M2) (M4, M5, M6). Locations in the same group have similar average values of observations.
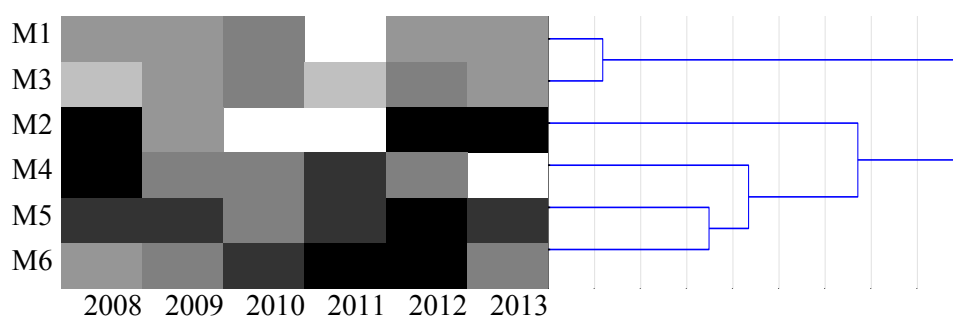
Figure 7. Hierarchical tree and graph of heat

4.4. Andrews' plots

Analysis was made of modified and unmodified Andrews' plots with two orderings of years: increasing, and by importance according to PCA. The two most interesting plots are presented in Figure 8. On the left plot, locations M2 and M4 differ from the others in frequency. This means that the most important years for showing the differences were 2008–2010. However, when years are
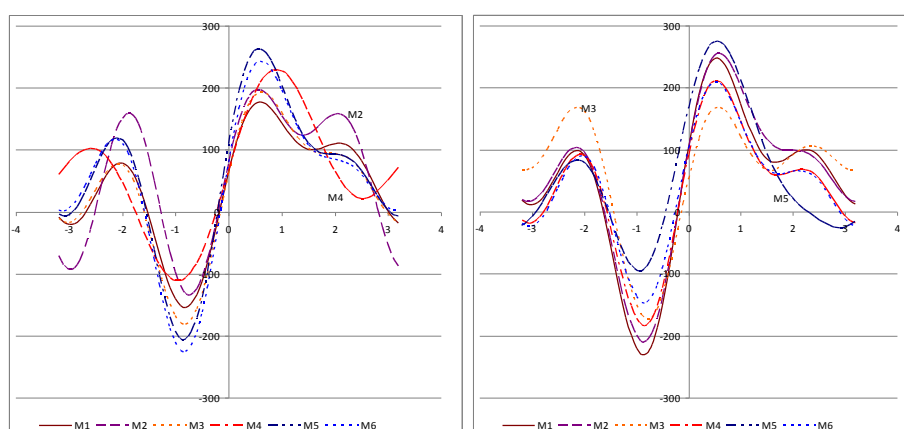

Figure 8. Modified Andrews' plot with increasing year data (left panel) and unmodified Andrews' plot with the years' order taken from PCA (right panel)

arranged by their importance according to PCA (2008, 2013, 2011, 2012, 2009, 2010) M3 is found to differ in amplitude (right panel). This is a consequence of the fact that the mean for M3 changes its position during the years (see Table 5). The analysis confirms the results of the analysis of homogeneous groups, where the locations are divided into three groups with M3, M2 and M4 as the respective representatives.

4.5. Coefficients of usefulness

Coefficients of usefulness were calculated using the formula (5). The least differentiated coefficients (Table 7) were observed in 2008 (maximum – minimum = 0.24) and at location M3 (0.44). The most differentiated coefficients were observed in 2010 (0.72) and at the locations M2 and M5 (0.60). The smallest average (over years) w_{lj} was found at location M2; thus the results obtained at that location may be very useful in the assessment of genotypes. In the years 2010 and 2012 the coefficients w_{lj} were the smallest at this location, and the w_{lj} were not the largest in any year. The location M1 may be considered the least useful for selecting genotypes (because it had the largest average w_{lj}, although in 2008 and 2009 the coefficients w_{lj} were the smallest and only in 2011 was the coefficient the largest). On the basis of the joint analysis for all years (formula 8) M3 was found to be the most useful location, and M2 and M5 were less useful. These results may be partially confirmed by the correlation coefficients calculated according to the formula (6). At location M5 (Table 7) the order of genotypes, in terms of yield, was the most similar to the order by average for all locations (for both the average over years and the joint analysis the correlation coefficients r_{lj} were greater than 75% - last two rows in Table 7). At locations M2 and M3 the order of genotypes in terms of yield was the most different from the order by average for all locations, but in 2013 for M2 and 2009 for M3 the coefficients r_{lj} were quite large (greater than 70%). At M4 the coefficients r_{lj} were the least differentiated (maximum – minimum = 0.16) and they were always smaller than 75%. At M3 the greatest differentiation was observed (0.66).

Table 7. Coefficients of usefulness of locations and correlation coefficients
between yields in the location and means yield

	Coefficients of usefulness						Correlation coefficients					
	M1	M2	M3	M4	M5	M6	M1	M2	M3	M4	M5	M6
2008	**0.35**	0.50	0.59	0.57	0.43	0.47	0.49	0.57	0.51	0.59	0.62	0.55
2009	**0.37**	0.42	0.61	0.47	0.76	0.56	0.68	0.43	0.80	0.72	0.84	0.73
2010	0.85	**0.13**	0.84	0.63	0.56	0.85	0.68	0.36	0.60	0.56	0.58	0.71
2011	0.78		0.43	**0.31**	0.35	0.53	0.52		0.22	0.58	0.73	0.67
2012	0.68	**0.00**	0.67	0.77	0.67	0.46	0.57	0.70	0.62	0.64	0.74	0.72
2013	0.91	0.73	**0.40**		0.95	0.92	0.78	0.80	0.14		0.86	0.83
Average	0.66	**0.36**	0.59	0.55	0.62	0.63	0.64	0.57	0.50	0.63	0.75	0.73
Jointly[1]	0.47	0.50	**0.20**	0.41	0.51	0.48	0.63	0.54	0.59	0.64	0.78	0.70

[1]Joint analysis of all years in the given location

Table 8. The percentage contribution of sum of squares to
genotype×environmental interaction

Year	Contribution to genotype×environmental interaction					
	M1	M2	M3	M4	M5	M6
2008	17.98	10.95	**34.95**	15.08	14.29	6.74
2009	**21.67**	17.95	15.81	16.97	16.89	10.71
2010	9.55	19.64	16.80	22.35	9.27	**22.39**
2011	6.74		**35.33**	13.39	32.42	12.12
2012	10.79	2.91	**53.10**	14.69	11.05	7.46
2013	8.08	**33.82**	26.47		19.77	11.87
Average	12.47	17.05	**30.41**	16.49	17.28	11.88
Jointly	17.09	**22.71**	18.81	15.85	6.74	18.79

The percentage contribution of locations, in the whole of the sum of squares, to genotype-by-environment interaction (Table 8) was dependent on year. The largest contribution came most frequently from location M3. At this location the joint coefficient w_j was the smallest. The joint percentage contribution was the largest for location M2 and the smallest for M5. For location M4 the percentage contribution was in one of the three highest positions for all years; moreover for this location the differences between maximum and minimum contributions was the smallest.

Figure 9 shows coefficients of usefulness (left panel) and percentage contributions of sum of squares to genotype×environment interaction (right

panel). The smallest values of w_{lj} for M2, M3 and M4 and the highest percentage contribution of sum of squares for M3 are clearly visible.
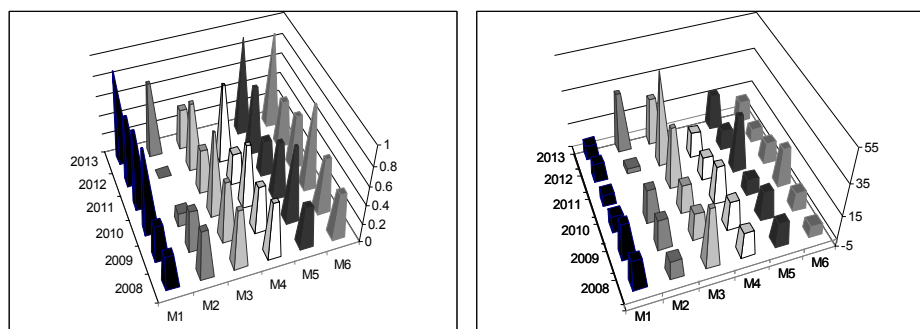

Figure 9. Coefficients of usefulness of locations (left panel); The percentage contribution of sum of squares to genotype × environmental interaction (right panel)

5. Discussion

In this study we have presented some methods for the assessment of locations where breeding trials with spring barley are conducted. The results from a six-year period (from 2008 to 2013) were considered. Because of the very high cost of these trials, there exists a tendency to reduce the number of locations. The presented methods made it possible to identify unique locations and locations for which the results were similar. Exclusion of locations with similar results should not have much impact on the choice of genotypes.

Understanding the causes of genotype × environment (GE) interactions is one of the most important tasks in crop breeding programs. Many authors have tried to eliminate the influence of environment on the crops. As a stability analysis, they have applied the AMMI model with additive main effects and multiplicative interaction, but the results have varied. For example, in (Tumuhimbise et al., 2014) the effect of GE interaction was found to be negligible for some genotypes of cassava (*Manihot esculenta*) and significant for others. Cassava is a storage root crop grown by most smallholder farmers partly because of its flexibility in

harvesting time and ability to perform well in drought-prone areas under poor management. Unfortunately cassava presents substantial differential genotypic responses under varying environmental conditions. Trials were conducted at three diverse locations in Uganda. The data were first analyzed independently for the locations, and then the error variances for the environments were tested for homogeneity (using Hartley's F test). The differences were not significant, and AMMI analysis of variance with two components in PCA was conducted across the locations.

Similar analyses have been made of malt barley (*Hordeum vulgare L.*) in six locations in Ethiopia (Mehari et al., 2014), sunflower (*Heliantus annuus*) in two locations in Pakistan (Khan et al., 2013) and Hungarian vetch (*Vicia pannonica CRANTZ.*) in five locations in Turkey (Sayar et al., 2013). The experiments were performed in randomized complete block designs. The testing locations were diversified for the purpose of discriminating the genotypes. Genotype-by-environment interaction was reported as the most important source of variation for the measured yield of the crops.

The results of AMMI analysis are often presented in a biplot, which displays both the genotype and environment values and their relationships using the singular vector technique. A biplot (Kandus et al., 2010) was used to discover which genotypes of maize (*Zea mays* L.) obtained the highest and the lowest yields in each environment and to distinguish mega-environments. To determine a mega-environment in a graphical form, the extreme genotypes of the biplot were joined to form a polygon, and then perpendicular lines were drawn on each side of the polygon through the origin. The mega-environments (Kandus et al., 2010) contained three locations (two in two years and one in one year) considered as five environments. The same analysis was conducted on sweet potato root (*Ipomoea batatas*) yield in Kenya (Kivuva et al., 2014), but only in two locations (although there were eight environments), and on wheat (*Triticum aestivum L.*) in Serbia in five locations (Zečević et al., 2009).

The AMMI and biplot analyses were sometimes complemented with so-called 'which-won-where' information, which provided an effective method to select

suitable winning genotypes according to stability in an analysis of soya bean (*Glycine max*) yield in four locations in Canada (Zhe et al., 2010).

Fan et al. (2001) reported, based on 15 locations for regional trials of rice (*Oryza sativa*), that the evaluation of locations based on GE interaction will depend upon a proper year-level in order to cope with unexpected GE interaction, thus arbitrary conclusions based on data from only one or two years would be avoided. They proposed a method of selecting test locations based on GE interaction. The data contained measures of winter wheat (*Triticum aestivum* L.) yield in 21 years in the United States, and allowed the authors to construct six regions for reducing GE interaction.

Another method of searching was proposed by Mohammadi et al. (2015). Cluster analysis is applied to separate the environments into four groups with similar discriminating ability among genotypes, and the genotypes into five groups with similar patterns in yield performance. This was done for seven locations in three years (21 different environments). Next, AMMI analysis was performed with a biplot based on the first two components of GE interaction from PCA. Finally, the clusters were marked on the biplot and inferences about durum wheat (*Triticum durum*) in Iran were obtained from both methods.

Many authors have estimated heritability coefficients and genetic coefficients of variability. Yildirim and Çaliskan (1985) calculated these in a study of potato (*Solanum Tuberosum L.*) in Turkey, Costa et al. (2000) in an analysis of rubber trees (*Hevea*) in Brazil, and Cooper et al. (1995) in a study of wheat (*Triticum aestivum L.*) yield in Australia.

Twelve locations used in the analysis of barley (*Hordeum vulgare L.*) yield were examined by Lin and Binns (1985). They used the standard method (AMMI) but operated on contrasts with checks, not on raw data.

6. Conclusions

In this paper the interaction is taken as a whole and not as a product of the effects of genotype×years and genotype×location in ANOVA. The reason is that the

point of interest here is the usefulness of locations, not the identification of reasons for GE interaction.

From the analysis of variance it can be concluded that: (1) The type of barley (fodder or malting), interaction between type and locations and genotypes have no significant influence on the yield (Anova 1). (2) The random effects of years were not significant (Anova 2) in comparison with the (stable) effect of location, which was significant. (3) At location M1 the yield was smallest in three years of the six and second smallest in the other three years. M2 was characterized by large differences in yield across years. (4) The homogeneous groups were different in each of year of the research.

From graphical methods it follows that: (1) From cluster analysis the yields at M1 and M3, and at M5 and M6, were similar. M2 formed a separate group in terms of yield. (2) PCA showed that 46% of the variation is covered by factor 1, built on M2, M3 and M4, while factor 2, constructed on M5 and M6, accounts for 28% of the variation. Factor 3, covering 15% of the variation, was based on M1 and M2. (3) Thus M2, M3 and M4 are the most differentiated locations. (4) The yields at M1, M2 and M4 may determine the remainder (they were different from the rest and characteristic in their groups). Generally M4 and M3, as well as M1 and M5 or M2 and M6, were collinear, and hence it is enough to analyze the results of only one location from each pair. M1, M2 and M4 are best chosen as representative locations. (5) According to the Andrews' plot, M2 and M4 differed from the other locations, but when the years were ordered according to the contribution of variables to the first main components, the curve for M3 had a different shape.

The following inferences may be made from the values of coefficients: (1) M5 has the maximal value of the genetic correlation coefficient, which denotes the lowest level of heterogeneity. M1, M2 and M3 contribute the greatest variety of results (their correlation with the mean yield is the weakest). (2) The values of heritability coefficients indicated the significant influence of genotype-by-environment interaction on the yield. (3) The usefulness coefficients are the smallest over the years at M3, and the largest at M2 and M5. (4) M3 made the

largest contribution to the interaction, whereas M5 had the smallest variety over the years. (5) The results obtained separately in each year w_{lj} were inconsistent with the results obtained from joint analysis w_j. (6) The locations M2 and M3 seemed to be the most useful to distinguish varieties, while M5 and M1 were the least useful. (7) The location M4 was the most stable as regards the contribution in the sum of squares for the interaction as well as correlation between yields at that location and average yield from all locations.

The following overall conclusions are drawn: (1) The majority of the applied methods indicate three groups of locations. The first group contains M1 and M3, which behave similarly, although M3 is slightly better than M1 (regarding usability, size of contribution to the interaction, and power of discrimination). This may be named the 'smallest yield group'. M2 forms the second group, which can be characterized by the term 'individual location', because the yield at M2 varies very much over the years and behaves differently compared with the other locations. Locations M4, M5 and M6 form a group that may be named the 'largest yield group'. M4 is the best representative of this group, because it is the most stable. (2) M3 had the largest contribution to the interaction and the lowest usefulness coefficient. Such locations are the most valuable, because at the last stage of new variety cultivation, when new varieties are to be registered with the Research Centre for Cultivar Testing (COBORU in Słupia Wielka, Poland) they are evaluated in more varied experimental environments. In that case a very important genotype×environment interaction may occur, and very good genotypes are still not predominant. It is advisable to examine genotypes in locations that cause the largest interaction (Pilarczyk et al., 2010).

REFERENCES

Andrews D. (1972): Plots of High-dimensional data. Biometrics. 28: 125-136.
Cooper M., Woodruff D.R., Eisemann R.L., Brennan P.S., DeLacy I.H. (1995): A selection strategy to accommodate genotype-by-environment interaction for grain yield of wheat: managed-environments for selection among genotype. Theor. Appl. Genet. 90: 492-502.

Costa R.B., de Resende M.D.V, de Araujo A.J., Gonçalves P.S., Martimns A.L.M. (2000): Genotype-environment interaction and the number of test sites for the genetic improvement of rubber trees (Hevea) in São Paulo State, Brazil. Gen. Mol. Biol. 23:179-187.

Fan L.J., Hu B.M., Shi C.H., Wu J.G. (2001): A method of choosing locations based on genotype $\times$ environment interaction for regional trials of rice. Plant Breeding. 120: 139-142.

Härdle W., Simar L. (2003): Applied multivariate statistical analysis, Springer, New York.

Kandus M., Almorza D., Boggio Ronceros R., Salerno J.C. (2010): Statistical models for evaluating the genotype-environment interaction in maize (Zea mays L.). ΦYTON. 79: 39-46.

Khan H., Rehman H.U., Bakht J., Khan S.A., Hussain I., Khan A., Ali S. (2013): Genotype $\times$ environment interaction and heritability estimates for some agronomic characters in sunflower. J. Anim. Plant. Sci. 23: 1177-1184.

Khattree R., Naik D.N. (2002): Association in contingency tables, correspondence analysis and (modified) Andrews plots. In: Huber-Carol, C., N. Balakrishnan, M.S. Nikulin, M. Mesbah (Eds.). Goodness-of-fit Tests and Model Validity. Birkhäuser, Boston: 311–326.

Kivuva B.M., Githiri S.M., Yencho G.C., Sibiya J. (2014): Genotype x environment interaction for storage root yield in sweet potato under managed drought stress conditions. J. Agr. Sci. 6: 41-56.

Lin C.S., Binns M.R. (1985): Procedural approach for assessing cultivar location data: pairwise genotype-environment interactions of test cultivars with checks. Can. J. Plant Sci. 65: 1065-1071.

Mądry W., Mańkowski D., Kaczmarek Z., Krajewski P., Studnicki M. (2010): Statistical methods based on linear models in applications for experimentation, genetics and plant breeding (in Polish). IHAR Państwowy Instytut Badawczy Radzików. Radzików, Poland.

Mehari M., Alamerew S., Lakew B. (2014): Genotype $\times$ environment interaction and yield stability of malt barley genotypes evaluated in Tigray, Ethiopia using the AMMI analysis. Asian. J. Plant. Sci. 13: 73-79.

Mohammadi R., Farshadfar E., Amri A. (2015): Interpreting genotype x environment interaction for grain yield of rainfed durum wheat in Iran. The Crop Journal. 3: 526-535.

Pilarczyk W., Bakinowska E., Bocianowski J., Zawieja B. (2010): Optimization of preliminary plant breeding trials with spring barley (in Polish). Biuletyn IHAR. 255: 13-26.

Sayar M.S., Anlarsal A.M., Basbag M. (2013): Genotype–environment interactions and stability analysis for dry-matter yield and seed yield in Hungarian vetch *(vicia pannonica crantz.)*. Turk. J. Field. Crops. 18: 238-246.

Śmiałowski T., Węgrzyn S.T. (2001): Usefulness of trial sites for testing winter rye breeding materials (in Polish). Biuletyn IHAR. 218/219: 409-417.

Tumuhimbise R., Melis R., Shanahan P., Kawuki R. (2014): Genotype x environment interaction effects on early fresh storage root yield and related traits in cassava. The Crop Journal. 2: 329-337.

Yildirim M., Çaliskan C.F. (1985): Genotype x environment interactions in potato (*Solanum tuberosum L.*) Amer. Potato. J. 62: 371-375.

Zečević V., Knežević D., Bošković J., Madić M. (2009): Effect of genotype and environment on wheat quality. Genetika. 41: 247–253.

Zhe Y., Lauer J.G., Borges R., de Leon N. (2010): Effects of genotype × environment interaction on agronomic traits in soybean. Crop Sci. 50: 696–702.