

## Principal component analysis for functional data on grain yield of winter wheat cultivars

Mirosław Krzyśko<sup>1</sup>, Adriana Derejko<sup>2</sup>, Tomasz Górecki<sup>1</sup>,  
Edward Gacek<sup>3</sup>

<sup>1</sup>Faculty of Mathematics and Computer Science, Adam Mickiewicz University,  
Umultowska 87, 61-614 Poznan, Poland, e-mail: mkrzysko@amu.edu.pl,  
tomasz.gorecki@amu.edu.pl

<sup>2</sup>Department of Experimental Design and Bioinformatics, Warsaw University of Life  
Sciences, Poland, email: adrainaderejko@gmail.com

<sup>3</sup>The Research Centre for Cultivar Testing, Slupia Wielka, Poland,  
email: e.gacek@coboru.pl

### SUMMARY

The aim of this paper is to present a statistical methodology to assess patterns of cultivars' adaptive response to agricultural environments (agro-ecosystems) on the basis of complete Genotype x Crop Management x Location x Year (GxMxLxY) data obtained from 3-year multi-location two-factor trials conducted within the framework of the Polish post-registration trials (PDOiR), with an illustration of the application and usefulness of this methodology in analyzing winter wheat grain yield. Producing specific varieties for each subregion of a target region, from widely adapted varieties, may exploit positive genotype x location (GL) interactions to increase crop yields. Experiments designed to examine combinations of environment (E), management practices (M) and cultivars (G) also provide evidence of the relative importance of each of these factors for yield improvement. The evidence shows that variation due to E far outweighs the variation of grain yield that can be attributed to M or G, or the interactions between these factors, and between these factors and E (Anderson, 2010). This statistical method involves the use of functional PCA and cluster analysis.

A total of 24 cultivars were evaluated over 3 years in 20 environments using randomized incomplete split-block designs with two replications per trial. The methodology proved an efficient tool for the reliable classification of 24 winter wheat cultivars, distinguishing cultivar groups that exhibited homogeneous adaptive response to environments. It enables the identification of cultivars displaying wide or specific adaptation. The remaining cultivars were locally adapted to some testing environments, or some of them were not relatively adapted to the environments because they always yielded substantially below the environmental means. Performing earlier specific selection, or adopting distinct genetic bases for each agro-ecosystem, may further increase the advantage of specific breeding.

**Key words:** cultivar adaptation, functional data, grain yield, post-registration trials, principal component analysis, winter wheat

## 1. Introduction

Wheat is a major crop contributing to the world's food supply. Of the total supply of wheat, an average of 53% is consumed as food in the developed countries, and close to 85% in the developing countries (Denčić et al., 2011). It has long been recognized that wheat yields and other agronomic and quality traits vary considerably as a result of genotype, environment and their interaction (Allard and Bradshaw, 1964; Basford and Cooper, 1998; Trethowan and Crossa, 2007; Denčić et al., 2011).

The main objective of plant breeding in major crop species, including winter wheat, is to develop new cultivars demonstrating one of two adaptation patterns, called wide or specific (local) adaptation, to environments within a target production (growing) area (Sivapalan et al., 2000; Annicchiarico, 2002a,b; Rane et al., 2007). The adaptation patterns of each tested cultivar can be described by their yield responses (also called cultivar adaptive responses – Annicchiarico et al., 2006b, 2011; Annicchiarico and Iannucci, 2008; Gauch et al., 2008) across a wide range of environments and also years in the production area. Predicting repeatable cultivar adaptive responses requires the conduct of 3-year multi-environment trials with a set of offered cultivars, continued across representative test environments (locations) of the production area and years (Annicchiarico, 2002a,b; Trethowan and Crossa, 2007; Annicchiarico et al., 2010). On the basis of yield data from these trials it is possible to estimate (predict) both the genotypic mean yield and repeatable genotype x location (GL) interaction effects (Yan and Hunt, 1998; Trethowan et al., 2002; Annicchiarico, 2002b; Annicchiarico et al., 2006a, 2010). GxL interaction effects for a given cultivar produce repeatable cultivar adaptive responses (Ghaderi et al., 1982; Yan et al., 2007; Rodriguez et al., 2008; Annicchiarico, 2002b).

Cultivars having wide adaptation are defined as those that produce yields substantially above the environmental means and are among the few top-ranking cultivars at a majority of locations across a production area characterized by substantial variation in environmental conditions (Braun et al., 1996; Annicchiarico, 2002b; Rodriguez et al., 2008). Such cultivars produce relatively high and stable yields within the area (Annicchiarico, 2002b; Singh et al., 2007; Yan et al., 2007; Yang et al., 2009). Cultivars having specific adaptation are defined as those that produce yields substantially above the environmental means and are among the few top-ranking cultivars in the range of a sub-region (macroenvironment) within the target region, usually of limited environmental variation (Gauch and Zobel, 1997;

Annicchiarico, 2002b; Lillemo et al., 2005; de la Vega and Chapman, 2006) or in at least one environment within the target area (Annicchiarico and Iannucci, 2008; Annicchiarico et al., 2010). Usually, cultivars with wide adaptation have fairly high yield potential and stress tolerance, whereas specifically-adapted ones have maximum levels of either yield potential or stress tolerance (Annicchiarico, 2002b; Singh et al., 2007; Trethowan and Crossa, 2007; Ulukan, 2008). Although widely adapted cultivars are usually preferred, the merits of those with local adaptation are also recognized (Annicchiarico, 2002a,b; Zhang et al., 2006; Singh et al., 2007).

In Poland, predictions of repeatable cultivar adaptive responses of important crops are assessed in post-registration variety testing trials (called PDOiR trials), which are repeated across environments and years. These deliver essential information for effective cultivar recommendations for megaenvironments (locally adapted cultivars) or for large sub-regions including also the whole country (widely adapted cultivars). Many different statistical methods have been used to estimate or predict cultivar adaptive responses using PDOiR data, including largely graphical methods (Yan et al., 2007; Yang et al., 2009; Kozak, 2010a,b) and more advanced methods (Gauch 1992, 2006; Gauch and Zobel, 1997; Annicchiarico, 2002b; Gauch et al., 2008). Among these are advanced methods based on functional principal components.

However, in a case with a large number of assessed cultivars, the tool of nominal yield used in its classic form can be less effective, due to the many lines on the nominal yield graph and difficulties in clearly distinguishing them (Hausmann et al., 2000; Kozak, 2010b). A solution to this problem could be to group cultivar adaptive responses into homogeneous groups using cluster analysis.

The objective of this study is to illustrate the application and usefulness of functional principal components and cluster analysis to assess the genotype grain yield adaptive responses of recent Polish and foreign winter wheat cultivars across a major wheat-growing area in Poland under integrated crop management, using data obtained in the PDOiR trials.

## 2. Materials and methods

### 2.1. Experimental material

The data used in the study relate to the grain yield of 24 recent Polish and foreign winter wheat cultivars tested across 20 locations (called Exper-

imental Stations for Cultivar Testing) and repeated over three growing years (2007–2009). The cultivars were assessed under the post-registration variety testing trials (PDOiR trials) conducted within the nationwide PDOiR trials system developed by the Research Centre for Cultivar Testing in Słupia Wielka, near Poznań, Poland ([http://www.coboru.pl/English/index\\_eng.aspx](http://www.coboru.pl/English/index_eng.aspx)). The test locations were selected in such a way as to cover (represent) the main Polish wheat-growing area. In each trial, integrated crop management was used with N-rates of 40 kg ha<sup>-1</sup> less as compared with yield expectations and standard PK fertilizations for a given location and with pesticide use limited to seed treatment, without the use of growth regulators to prevent lodging. The seeding rate ranged from 450 to 550 grains/m<sup>2</sup> at the locations, depending on the cultivar. All experiments in macroenvironments were designed as a randomized incomplete split-block design with two replicates, with plot size 15 m<sup>2</sup> (10 × 1.5m). Mean values for varieties are calculated from single experiments. These experiments are incomplete, hence the mean values are correlated. This fact is ignored.

The cultivars, the breeding companies which produced them, and the years of their release are given in Table 1. Among the 24 tested cultivars, 15 come from Polish breeding companies, and the remaining 9 from German, French and UK companies. The locations of the Experimental Stations for Cultivar Testing (SDOOs), whose names and geographical position are indicated in Figure 1, were well-scattered across the main Polish common wheat-growing area, and are thus representative of that area.

## 2.2. Principal component analysis for functional data

For a given variety of winter wheat and a given location (experimental station), the grain yield data are recorded in the form of the following sequence:

1. first intensity of crop management, in 2007
2. second intensity of crop management, in 2007
3. first intensity of crop management, in 2008
4. second intensity of crop management, in 2008
5. first intensity of crop management, in 2009
6. second intensity of crop management, in 2009.

**Table 1.** Winter wheat cultivars tested in post-registration trials (PDOiR trials) carried out at locations in 2007–2009

Cultivar	Year of release	Breeding company
Alcazar	2006	Secobra Recherches
Anthus	2002	KWS Lochow GmbH
Batuta	2003	DANKO Hodowla Roślin sp. z o.o.
Bogatka	2001	DANKO Hodowla Roślin sp. z o.o.
Boomer	2003	RAGT Seeds Ltd.
Dorota	2001	RAGT Seeds Ltd.
Flair	1998	Saatzucht Hans Schweiger and Co. oHG
Kobiera	1999	Hodowla Roślin Rolniczych - Nasiona Kobierzyc
Kris	1997	RAGT Seeds Ltd.
Legenda	2002	Poznańska Hodowla Roślin sp. z o.o.
Ludwig	2001	Saatzucht Donau Ges.m.b.H. and CoKG
Mewa	1996	DANKO Hodowla Roślin sp. z o.o.
Nadobna	1999	Poznańska Hodowla Roślin sp. z o.o.
Nutka	1998	Hodowla Roślin Strzelce sp. z o.o. Grupa IHAR
Rapsodia	2000	RAGT Seeds Ltd.
Rywalka	1999	Hodowla Roślin Strzelce sp. z o.o. Grupa IHAR
Satyna	2000	Hodowla Roślin Rolniczych - Nasiona Kobierzyc
Smuga	2001	DANKO Hodowla Roślin sp. z o.o.
Sukces	1996	Hodowla Roślin Strzelce sp. z o.o. Grupa IHAR
Tonacja	1998	Hodowla Roślin Strzelce sp. z o.o. Grupa IHAR
Trend	2000	KWS Lochow GmbH
Turnia	1998	Małopolska Hodowla Roślin-HBP sp. z o.o.
Wydma	2002	Hodowla Roślin Smolice sp. z o.o. Grupa IHAR
Zyta	1996	Hodowla Roślin Strzelce sp. z o.o. Grupa IHAR

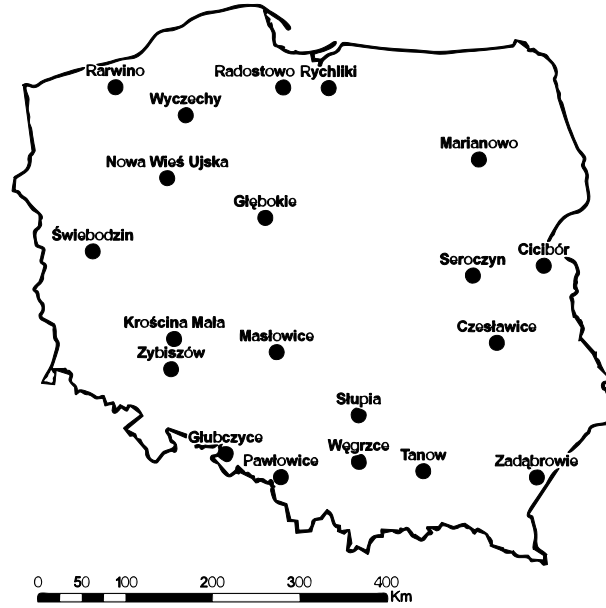
This sequence is then supplemented with the data relating to other locations. In this way the given winter wheat variety is characterized by a spatial and temporal sequence composed of  $J = 6 \times 20 = 120$  numbers.

The successive values of this sequence will be  $y_j$ , where  $j = 1, 2, \dots, J$ . The sequence is smoothed using a continuous function  $x(t)$ , where  $t \in [0, T]$ . The function  $x(t)$  is taken to be of the form

$$x(t) = \sum_{k=0}^K c_k \varphi_k(t), \quad t \in [0, T],$$

where  $\{\varphi_k(t)\}$  is an orthonormal system of basis functions, that is, functions for which

$$\int_0^T \varphi_k(t) \varphi_l(t) dt = \begin{cases} 1, & \text{if } k = l, \\ 0, & \text{if } k \neq l. \end{cases}$$



**Figure 1.** Locations of the Experimental Stations for Cultivar Testing in Poland within the network of COBORU stations where post-registration trials (PDOiR trials) for winter wheat were carried out in 2007-2009

As an orthonormal basis of  $L_2(I)$  we took the orthonormal system of Legendre polynomials (Abramowitz and Stegun, 1965) in the space  $L_2([-1, 1])$ :

$$\tilde{p}_k(x) = \sqrt{\frac{2k+1}{2}} p_k(x),$$

where

$$p_{k+1}(x) = \frac{1}{k+1} [(2k+1)xp_k(x) - kp_{k-1}(x)], \quad k \geq 1, \\ p_0(x) = 1, \quad p_1(x) = x.$$

The coefficients  $c_0, c_1, \dots, c_K$  are selected by the least squares method based on the data  $y_1, y_2, \dots, y_J$ , i.e. so as to minimize the function:

$$S(c_0, c_1, \dots, c_K) = \sum_{j=1}^J (y_j - \sum_{k=0}^K c_k \varphi_k(t_j))^2.$$

In matrix notation, the function  $S$  has the form:

$$S(\mathbf{c}) = (\mathbf{y} - \mathbf{\Phi}\mathbf{c})'(\mathbf{y} - \mathbf{\Phi}\mathbf{c}),$$

where  $\mathbf{y} = (y_1, y_2, \dots, y_J)'$ ,  $\mathbf{c} = (c_0, c_1, \dots, c_K)'$  and  $\mathbf{\Phi}$  is a  $J \times (K + 1)$  matrix containing the values  $\varphi_k(t_j)$ , where  $k = 0, 1, \dots, K$ ,  $j = 1, 2, \dots, J$ . By differentiating  $S(\mathbf{c})$  with respect to the vector  $\mathbf{c}$  we obtain a system of normal equations in the form:

$$2\mathbf{\Phi}'\mathbf{\Phi}\mathbf{c} - 2\mathbf{\Phi}'\mathbf{y} = \mathbf{0}.$$

Hence the estimate of vector  $\mathbf{c}$  is equal to:

$$\hat{\mathbf{c}} = (\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{y}.$$

In this way, each of the 24 varieties of winter wheat was characterized by a continuous function

$$x(t) = \sum_{k=0}^K \hat{c}_k \varphi_k(t), \quad t \in [0, T].$$

The functions  $x_i(t)$  are called functional data (Ramsay and Silverman, 2005). The smoothness degree of the function  $x(t)$  depends on  $K$ , since small (large) values of  $K$  induce more (less) smoothed curves. The optimal number  $K$  of basis elements is determined using the Bayesian information criterion (BIC) for each function  $x_i(t)$  separately:

$$\text{BIC} = \ln\left(\frac{\mathbf{e}'\mathbf{e}}{2}\right) + (K + 1)\frac{\ln J}{J},$$

where  $\mathbf{e} = (e_1, e_2, \dots, e_J)'$ ,  $e_j = y_j - \sum_{k=0}^K \hat{c}_k \varphi_k(t_j)$ ,  $j = 1, 2, \dots, J$ . As the optimum value of  $K$  we take the value for which the BIC attains a minimum. Next, among the 24 values of  $K$  corresponding to all of the winter wheat cultivars, the modal value is selected, as a common value of  $K$  for all 24 functions  $x(t)$ .

The estimated values  $\hat{c}_0, \hat{c}_1, \dots, \hat{c}_K$  are treated as the values of new statistical properties. For the properties so obtained, we construct principal components. Principal component analysis is based on evaluation of a covariance matrix  $\mathbf{\Sigma}$  built from the data for  $n = 24$  winter wheat cultivars. Let  $\hat{\mathbf{C}}$  be a matrix in the form

$$\hat{\mathbf{C}} = \begin{bmatrix} \hat{c}_{10} & \hat{c}_{11} & \dots & \hat{c}_{1K} \\ \hat{c}_{20} & \hat{c}_{21} & \dots & \hat{c}_{2K} \\ \dots & \dots & \dots & \dots \\ \hat{c}_{n0} & \hat{c}_{n1} & \dots & \hat{c}_{nK} \end{bmatrix},$$

where  $\hat{c}_{rs}$  is the  $s$ th coefficient for the  $r$ th cultivar,  $r = 1, 2, \dots, 24$ ,  $s = 1, 2, \dots$ . Then the estimator  $\hat{\Sigma}$  of the unknown matrix  $\Sigma$  has the following form:

$$\hat{\Sigma} = \frac{1}{n} \hat{\mathbf{C}}' \hat{\mathbf{C}}.$$

We next find the eigenvalues  $\lambda_k$  and corresponding eigenvectors  $\mathbf{u}_k$  of the matrix  $\hat{\Sigma}$ . The pair  $(\lambda_k, \mathbf{u}_k)$  will be called the  $k$ th principal configuration. Knowing the  $k$ th principal configuration for the random vector  $\mathbf{c} = (c_0, c_1, \dots, c_K)'$ , we are able to find the principal components for the stochastic process  $X(t)$  by determining the  $k$ th principal configuration  $(\sigma_k, u_k(t))$  for that process.

**Theorem 1.** (see Górecki and Krzyśko, 2012) *The  $k$ th principal configuration of a random vector  $\mathbf{c}$ , defined by  $(\lambda_k, \mathbf{u}_k)$ , is related to the  $k$ th principal configuration of the stochastic process  $X(t)$ ,  $(\sigma_k, u_k(t))$ , as follows:*

$$\sigma_k = \lambda_k, \quad u_k(t) = \mathbf{u}_k' \boldsymbol{\varphi}(t).$$

The functions  $u_k(t)$  are called eigenfunctions or weight functions. The  $k$ th functional principal component for the process  $X(t)$  is defined as the scalar product

$$\langle u_k(t), x(t) \rangle = \int_0^T u_k(t) x(t) dt.$$

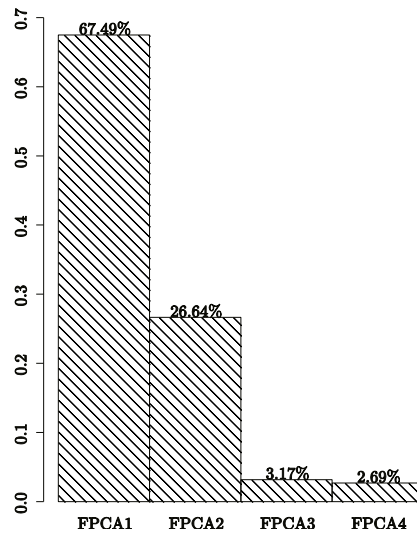
Hence the coordinate projection of the  $i$ th realization  $x_i(t)$  of the process on the  $k$ th functional principal component is equal to

$$U_{ik} = \hat{\mathbf{c}}_i' \mathbf{u}_k,$$

where

$$\hat{\mathbf{c}}_i = (\hat{c}_{i0}, \hat{c}_{i1}, \dots, \hat{c}_{iK})', \quad \hat{\mathbf{u}}_k = (\hat{u}_{k0}, \hat{u}_{k1}, \dots, \hat{u}_{kK})', \\ i = 1, 2, \dots, n, \quad k = 1, 2, \dots$$





**Figure 2.** Scree graph

The most frequently considered objects are presented on a plot of the first two functional principal components. In this case the criterion of goodness of the constructed functional principal components is the following expression (degree of explanation of the variation of winter wheat yields in the system of the first two functional principal components):

$$\frac{\lambda_1 + \lambda_2}{\sum \lambda_i} 100\%,$$

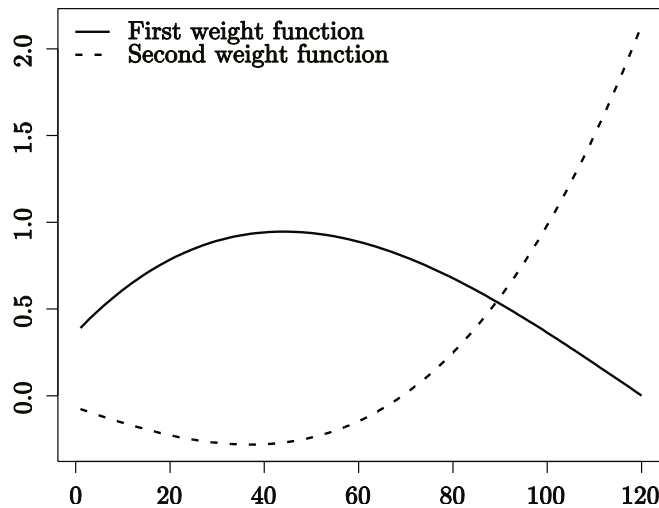
The greater the value of the expression, the greater is the variability shown by the first two functional principal components. In fact, the degree of explanation of the variation using the first two functional principal components relates to the functions  $x_i(t)$  and not the varieties that they represent. The degree of explanation of the variation for the varieties may be smaller, because it depends on the accuracy with which the real data for the varieties are expressed by these functions.

In the computational process the R program (R Core Team, 2013) was used. We also used the fda package (Ramsay et al., 2012).

### 3. Results

Figure 2 shows a scree graph for the functional principal component analysis. It can be seen that the first functional principal component explains 67.49% of the variation, and the second explains 26.64%. In total, the first two functional principal components explain as much as 94.13% of the variation. In this situation, a projection onto a plane constructed using the first two functional principal components would appear to be a very good solution. The algorithm selected 4 ( $K = 3$ ) elements of the orthonormal basis of Legendre polynomials.

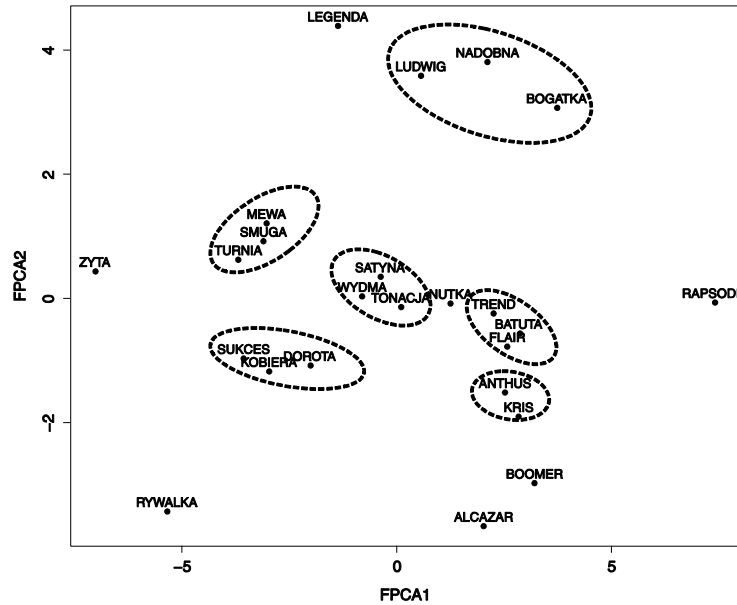
Figure 3 shows the first two weight functions  $u_k(t)$ ,  $k = 1, 2$  for the 4 elements of the orthonormal basis of Legendre polynomials.



**Figure 3.** The first two weight functions  $u_k(t)$ ,  $k = 1, 2$

Finally there remains the most important visualization, namely the projection of the data onto the plane of the first two functional principal components, which will show which cultivars of winter wheat behaved similarly during the period considered, in terms of the tested properties. If the points lie close to each other, this means that the cultivars behaved similarly in this regard. The graph is shown in Figure 4.

Now, in order to form homogeneous groups of cultivars, we apply cluster analysis. In the first step, we construct the minimum spanning tree. Two



**Figure 4.** Projection on the first two functional principal components. Homogeneous groups contained in the ellipses

popular algorithms for constructing the minimum spanning tree are those of Kruskal (1956) and Prim (1957). Kruskal's algorithm is a greedy algorithm that adds edges in order of their length, so long as they do not result in a cycle. Prim's algorithm starts with a tree consisting of a single vertex, and keeps adding the smallest edge of the tree. The minimum spanning tree can be used for clustering. This idea is to break (remove edges from) the minimum spanning tree at edges that are greater than the common critical value  $\rho^*$ . We compute the mean  $\bar{x}$  and the standard deviation  $s$  of all edge lengths in the minimum spanning tree, and eliminate edges which are longer than

$$\rho^* = \bar{x} + s,$$

(when we used  $2s$  instead of  $s$  we obtained only one large cluster). The points which remain connected in the minimum spanning tree form pseudoclusters (Karoński, 1973). The distance between two cultivars  $i$  and  $j$  is

equal to the Euclidean distance of the form

$$d(\mathbf{S}_i, \mathbf{S}_j) = \sqrt{\sum_{k=i}^4 (S_i^{(k)} - S_j^{(k)})^2},$$

where the  $i$ th cultivar is represented by the vector  $\mathbf{S}_i = (S_i^{(1)}, \dots, S_i^{(4)})'$  of functional principal components. In Figure 4 are plotted the values of the first two functional principal components, with homogeneous groups outlined in the ellipses. The resulting functional principal components can also be called the interactive principal components. Included in them is in fact the adaptive response of the grain yield of 24 varieties to the environment, years and intensity of crop management. Cluster analysis allowed the separation of six groups of homogeneous varieties with respect to their adaptability (the groups of varieties within the ellipses). Other varieties form single isolated clusters, and these are located on the edge of the six groups.

#### REFERENCES

- Abramowitz M., Stegun I.A. (1965): Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications.
- Allard R.W., Bradshaw A.D. (1964): Implications of genotype-environmental interactions in applied plant breeding. *Crop Sci.* 4: 503–508.
- Anderson W.K. (2010): Closing the gap between actual and potential yield of rainfed wheat. The impacts of environment, management and cultivar. *Field Crops Research* 116: 14–22.
- Annicchiarico P. (2002a): Defining adaptation strategies and yield stability targets in breeding programmes. In: Kang M.S. (Ed.). *Quantitative genetics, genomics and plant breeding*. CABI, Wallingford, UK: 165–183.
- Annicchiarico P. (2002b): Genotype-environment interactions: challenges and opportunities for plant breeding and cultivar recommendations. *FAO Plant Production and Protection Paper No. 174*. Food and Agriculture Organization, Rome.
- Annicchiarico P., Bellah F., Chiari T. (2006a): Repeatable genotype-location interaction and its exploitation by conventional and GIS-based cultivar recommendation for durum wheat in Algeria. *Eur. J. Agron.* 24: 70–81.
- Annicchiarico P., Russi L., Piano E., Veronesi F. (2006b): Cultivar adaptation across Italian locations in four turfgrass species. *Crop Sci.* 46: 264–272.
- Annicchiarico P., Iannucci A. (2008): Adaptation strategy, germplasm type and adaptive traits for field pea improvement in Italy based on variety responses across climatically contrasting environments. *Field Crops Res.* 108: 133–142.

- Annicchiarico P., Chiapparino E., Perenzin M. (2010): Response of common wheat varieties to organic and conventional production systems across Italian locations, and implications for selection. *Field Crops Res.* 116: 230–238.
- Basford K.E., Cooper M. (1998): Genotype x environment interactions and some considerations of their implications for wheat breeding in Australia. *Austr. J. Agric. Res.* 49: 153–174.
- Braun H.J., Rajaram S., van Ginkel M. (1996): CIMMYT's approach to breeding for wide adaptation. *Euphytica* 92: 175–183.
- de la Vega A.J., Chapman S.C. (2006): Defining sunflower selection strategies for a highly heterogeneous target population of environments. *Crop Sci.* 46: 136–144.
- Denčić S., Mladenov N., Kobiljski B. (2011): Effects of genotype and environment on breadmaking quality in wheat. *Int. J. Plant Prod.* 5: 71–82.
- Gauch H.G. (1992): Statistical analysis of regional yield trials. AMMI analysis of factorial designs. Elsevier Science, New York.
- Gauch H.G. (2006): Statistical analysis of yield trials by AMMI and GGE. *Crop Sci.* 46: 1488–1500.
- Gauch H.G., Piepho H.P., Annicchiarico P. (2008): Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop Sci.* 48: 866–889.
- Gauch H.G., Zobel R.W. (1997): Identifying mega-environments and targeting genotypes. *Crop Sci.* 37: 311–326.
- Ghaderi A., Adams M.W., Saettler A.W. (1982): Environmental response patterns in commercial classes of common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 63: 17–22.
- Górecki T., Krzyśko M. (2012): A kernel version of functional principal components analysis, *Statistics in Transition*, 13(3): 559–568.
- Karoński M. (1973): On a definition of cluster and pseudocluster for multivariate normal populations. In: Proceedings of the 39th Session of the International Statistical Institute. Vienna: 523–528.
- Kozak M. (2010a): Use of parallel coordinate plots in multi-response selection of interesting genotypes. *Commun. Biometry Crop Sci.* 5: 83–95.
- Kozak M. (2010b): Comparison of three types of  $G \times E$  performance plot for showing and interpreting genotypes' stability and adaptability. *Int. J. Plant Prod.* 5: 71–82.
- Kruskal J.B. (1956): On the shortest spanning subtree of a graph and the traveling salesman problem. In: Proceedings of American Mathematical Society 7: 48–50.
- Lillemo M., van Ginkel M., Trethowan R.M., Hernandez E., Crossa J. (2005): Differential adaptation of CIMMYT bread wheat to global high temperature environments. *Crop Sci.* 45: 2443–2453.
- Prim R.C. (1957) Shortest connection networks and some generalizations. *Bell System Technical Journal* 36: 1389–1401.

- R Core Team (2013): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Ramsay J.O., Wickham H., Graves S., Hooker G. (2012): fda: Functional Data Analysis. R package version 2.3.2. <http://CRAN.R-project.org/package=fda>.
- Ramsay J.O., Silverman B.W. (2005): Functional Data Analysis, Second Edition, Springer.
- Rane J., Pannu R.K., Sohu V.S., Saini R.S., Mishra B., Shoran J., Crossa J., Vargas M., Joshi A.K. (2007): Performance of yield and stability of advanced wheat genotypes under heat stress environments of the Indo-Gangetic Plains. *Crop Sci.* 47: 1561–1573.
- Rodriguez M., Rau D., Papa R., Attene G. (2008): Genotype by environment interactions in barley (*Hordeum vulgare* L.): different responses of landraces, recombinant inbred lines and varieties to Mediterranean environment. *Euphytica* 163: 231–247.
- Singh R.P., Huerta-Espino J., Sharma R., Joshi A.K., Trethowan R. (2007): High yielding spring bread wheat germplasm for global irrigated and rainfed production systems. *Euphytica* 157: 351–363.
- Sivapalan S., O'Brien L., Ortiz-Ferrera G., Hollamby G.J., Barclay I., Martin P.J. (2000): An adaptation analysis of Australian and CIMMYT/ICARDA wheat germplasm in Australian production environments. *Aust. J. Agric. Res.* 51: 903–915.
- Trethowan R., Crossa J. (2007): Lessons learnt from forty years of international spring bread wheat trials. *Euphytica* 157: 385–390.
- Trethowan R.M., van Ginkel M., Rajaram S. (2002): Progress in breeding wheat for yield and adaptation in global drought affected environments. *Crop Sci.* 42: 1441–1446.
- Ulukan H. 2008. Agronomic adaptation of some field crops: a general approach. *J. Agron. Crop Sci.* 194: 169–179.
- Yan W., Hunt L.A. (1998): Genotype by environment interaction and crop yield. *Plant Breed. Rev.* 16: 135–179.
- Yan W., Kang M.S., Ma B., Woods S., Cornelius P.L. (2007): GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Sci.* 47: 643–655.
- Yang R.C., Crossa J., Cornelius P.L., Burgueño J. (2009): Biplot analysis of genotype x environment interaction: Proceed with caution. *Crop Sci.* 49: 1564–1576.
- Zhang Y., He Z., Zhang A., van Ginkel M., Ye G. (2006): Pattern analysis on grain yield of Chinese and CIMMYT spring wheat cultivars grown in China and CIMMYT. *Euphytica* 147: 409–420.