Classification of Winter Rapeseed Cultivars and their Yield Characters with the Common Vector Approach

Nurdilek Gülmezoğlu[1], Zehra Aytaç[2] and M. Bilginer Gülmezoğlu[3]

[1]Department of Soil Science and Plant Nutrition, Faculty of Agriculture, Eskişehir Osmangazi University, 26480, Eskişehir, Turkey, e-mail: dgulmez@ogu.edu.tr
[2]Department of Field Crops, Faculty of Agriculture, Eskişehir Osmangazi University, 26480, Eskşehir, Turkey, e-mail: zehrak@ogu.edu.tr
[3]Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Eskişehir Osmangazi University, 26480, Eskişehir, Turkey, e-mail: bgulmez@ogu.edu.tr

Abstract In this study, five cultivars (Ceres, Zorro, Falcon, Express, and Samourai) of winter rapeseed were classified by using the common vector approach (CVA). For this purpose, seven yield characters (plant height, number of branches per plant, number of pods per plant, number of pods on main stem, number of seeds per pod, pod length and thousand seed weight) of each cultivar were used. The seven and six yield characters taken from five cultivars were classified by using CVA. 100% classification rate is guaranteed for the training set of both studies. For the test set, the classification of five cultivars has low performance, but the classification of seven and six yield characters gave satisfactory results. It is concluded that the CVA method was successful in the classification of different varieties belonging to any plant and/or of different characters belonging to any variety.

Keywords: Character classification, common vector approach, rapeseed classification.

1 Introduction

Rapeseed is an important oilseed crop in the agricultural systems of many arid and semiarid areas. Agronomic and quality advantages of new varieties have enlarged their production areas worldwide (Gül et al. 2007). Rapeseed in Turkey is mostly cultivated as a winter annual for oil production and rarely livestock feed. If planted in spring, they can be grown as summer crop but the seed yield would be decreased due to short growing season and lack of enough water at the end of growing season, thus, winter cropping is preferred. The canola cultivars are slow growing especially in winter and most of them will complete their life cycle in 210 to 270 days (Sharghi et al. 2011). There are wide variations among the cultivated canola cultivars with respect to seed and oil yields per unit area at different planting dates as well as irrigation regimes. The seed yield and maturity of canola is greatly influenced by fertility management, seeding rate and seeding date (Grant and Bailey, 1993).

Proceedings of the 7th International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2015), Kavala, Greece, 17-20 September, 2015.

Computer-based algorithms have been extensively used in agriculture in order to classify various plants and their characters or samples. Classification of plant varieties with computer algorithms has been become popular in recent years. The common vectors representing the invariant features of the plants can be extracted by eliminating the differences in each class of plants (Gülmezoglu 1999). Then these common vectors are used for the classification of varieties and characters of plants. Different methods were used in order to derive features or parameters from plant varieties (Wang et al., 1999; Zayas et al., 1996; Utku and Köksel, 1998; Delwiche & Massie, 1996; Neuman and Bushuk, 1987; Shuaib et al., 2010). Some classifications were analyzed for characters of rapeseed. Ali et al. (2012) analyzed for near infrared spectroscopy and principal component grain of rapeseed. Jankulovska et al. (2014) presented the use of different multivariate approaches to classify rapeseed genotypes based on quantitative traits. Some of these parameters were plant height, number of primary branches per plant, number of pods per plant, pod length, number of seeds per pod, seed weight per pod, 1000 seed weight, seed weight per plant and oil content. This model has been applied in agricultural sciences to identify the effect of yield character differences (Gülmezoğlu and Gülmezoğlu 2015). However, there is no information on the use in plant breeding programs.

In this study, we considered five cultivars (Ceres, Zorro, Falcon, Express, and Samourai) of winter rapeseed. Initially, these cultivars were classified by using the common vector approach (CVA). Secondly**,** seven yield characters (plant height, number of branches per plant, number of pods per plant, number of pods on main stem, number of seeds per pod, pod length and thousand seed weight) and six yield characters excepting number of branches per plant were classified by using CVA.

2 Material and Method

This research was carried out over three years during 2005 at the Faculty of Agriculture of Eskisehir Osmangazi University, Eskisehir (39^o 48′ N; 30^o 31′ E; 789 m in elevation). The field experiments included five winter rapeseed cultivars (Ceres, Zorro, Falcon, Express, Synergy and Samourai). The experiment was planned in a Randomized Complete Block Design with four replications. The individual plots were 3 m long and consisted of five rows. The cultivars were sown on the first week of September, using a seed rate of 10 kg ha $^{-1}$ in 40 cm spaced lines on a well prepared seed bed. The experiment was fertilized respectively with 150 kg N ha $^{-1}$ as ammonium nitrate: 33-0-0 and 50 kg P_2O_5 ha -1 as di-ammonium phosphate: 18-46-0. The plants were irrigated once during emergence and thinned at the rosette stage. The weeds were controlled by hand weeding.

Each cultivar was represented with seven yield characters which are plant height, number of branches per plant, number of pods per plant, number of pods on main stem, number of seeds per pod, pod length and thousand seed weight. Each character includes 20 plant samples which were taken from field study conducted during growing year.

As in all classification methods, CVA has training and testing stages. In the training stage, a common vector which represents common or invariant properties of each class is calculated and an in difference subspace for each class is constructed. Let the vectors $a_1^c, a_2^c, ..., a_m^c \in R^n$ be the feature vectors for a variety-class C in the training set where $m \leq n$. Then each of these feature vectors which are assumed to be linearly independent can be written as

$$\mathbf{a}_i^c = \mathbf{a}_{i,dif}^c + \mathbf{a}_{com}^c + \varepsilon_i^c \qquad \text{for } i=1,2,\ldots,m \qquad (1)$$

where the vector $\mathbf{a}_{i,dif}^c$ indicates the differences resulting from climatic effects and alien-pollination, and the vector $\mathbf{a}_{com}^c$ is the common vector of the variety or character class C, and ε_i^c represents the error vector (Gülmezoğlu et al. 2001). The common vector can be obtained from the subspace method. Let us define the covariance matrix of the feature vectors belonging to a variety or character class as

$$\mathbf{\Phi} = \sum_{i=1}^{m} (\mathbf{a}_i^c - \mathbf{a}_{ave}^c)(\mathbf{a}_i^c - \mathbf{a}_{ave}^c)^T \qquad (2)$$

where $\mathbf{a}_{ave}^c$ is the average feature vector of Cth class whose covariance matrix is to be calculated and T indicates the transpose of a matrix.

The eigenvalues of the covariance matrix $\mathbf{\Phi}$ are non-negative and they can be written in decreasing order: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Let $u_1^c, u_2^c, ..., u_n^c$ be the orthonormal eigenvectors corresponding to these eigenvalues. The first (m-1) eigenvectors of the covariance matrix corresponding to the nonzero eigenvalues form an orthonormal basis for the difference subspace B (Gülmezoğlu et al. 2001). The orthogonal complement, $B^{\perp}$, is spanned by all the eigenvectors corresponding to the zero eigenvalues. This subspace is called the indifference subspace and has a dimension of (n-m+1). The direct sum of two subspaces B and $B^{\perp}$ is the whole space, and the intersection of them is the null space. The common vector can be shown as the linear combination of the eigenvectors corresponding to the zero eigenvalues of $\mathbf{\Phi}$ (Gülmezoğlu et al. 2001), that is,

$$\mathbf{a}_{com}^c = \left\langle \mathbf{a}_i^c, \mathbf{u}_m^c \right\rangle \mathbf{u}_m^c + \ldots + \left\langle \mathbf{a}_i^c, \mathbf{u}_n^c \right\rangle \mathbf{u}_n^c \;\forall\; i=1,2,\ldots,m \qquad (3)$$

From here, the common vector acom is the projection of any feature vector onto the indifference subspace $B^{\perp}$. The common vector represents the common properties or invariant features of the variety or character class C. The common vector is independent from index i. Therefore, the common vector is unique for each class and all the error vectors ε_i^c would be zero.

During the classification stage, the following decision criterion is used:

$$distance = \underset{1 \leq C \leq S}{argmin} \left\| \sum_{j=m}^{n} \left\{ \left[\left(\mathbf{a}_x - \mathbf{a}_i^c \right)^T \mathbf{u}_j^c \right] \mathbf{u}_j^c \right\} \right\|^2 \qquad (4)$$

where $\mathbf{a_x}$ is an unknown or test vector and *S* indicates the total number of classes. If the distance is minimum for any class C, the feature vector $\mathbf{a_x}$ is assigned to class C.

Classification algorithm given above can be summarized as follows:

Step 1: Construct feature vectors by using samples taken for each character belonging to any cultivar. Be sure that number of samples in each feature vector (or dimension of each feature vector) is greater than number of feature vectors (or characters) for each cultivar.

Step 2: Find the covariance matrix (Eq. (2)) for each cultivar by using feature vectors belonging to that cultivar.

Step 3: Find the eigenvalues λ_i and corresponding eigenvectors u_i for each covariance matrix.

Step 4: Find the common vector (Eq. (3)) for each cultivar by using the *(*n-m+1) eigenvectors corresponding to zero eigenvalues.

Step 5: When an unknown feature vector $\mathbf{a_x}$ is given, classify this vector by using Eq. (4).

3 Results

In the first study, each of five cultivars forms one class in the CVA method. Seven characters each of which includes 20 samples for each cultivar or class form seven feature vectors of that class. Therefore, there are five classes and each class has seven feature vectors. When the feature vectors (characters) used in the training stage were tested, all classes (cultivars) were correctly classified, i.e., 100% correct recognition rate was obtained. When the "leave-one-out" strategy was used in the testing stage, that is, when six feature characters were used in the training stage and remaining one character was tested, 25.7% correct recognition rate was obtained as average of "leave-one-out" steps. The results obtained from this study are given in Table 1. The average score obtained in the test set is very low because samples included in the characters representing different cultivars are very close to each other.

In the second study, the characters were classified by using CVA. First of all, seven characters were considered and each of seven characters forms one class in the CVA method. Twenty samples taken from each cultivar for any character form one feature vector of that character. Therefore, there are seven classes and each class has five feature vectors. When the feature vectors, each of them includes 20 samples, used in the training stage were tested, all classes (characters) were correctly classified, i.e., 100% correct recognition rate was obtained. When the "leave-one-out" strategy was used in the testing stage, that is, when six feature characters were used in the training stage and remaining one character was tested, 77.1% correct recognition rate was obtained as average of "leave-one-out" steps. The results obtained for this study are given in Table 2.

Table 1. Correct recognition rates of five cultivars as percentage.

Varieties	Training set	Test set
Samourai	100	14.3
Zorro	100	42.9
Falcon	100	28.6
Ceres	100	0
Express	100	42.86
Average	100	25.7

Table 2. Correct recognition rates of seven yield characters as percentage.

Yield Characters	Training Set	Test Set
Plant Height	100	100
Number of Branches per Plant	100	0
Number of Pods per Plant	100	40
Number of Pods on Main Stem	100	100
Number of Seeds per Pod	100	100
Pod Length	100	100
Thousand Seed Weight	100	100
Average	100	77.1

Secondly, six characters excepting number of branches per plant were classified. All characters were correctly classified (100% correct recognition rate was obtained) in the training set and 90% recognition rate was obtained for the test set. These scores are remarkable because samples taken from cultivars for each character are close to each other and well represent that character. These results are given in Table 3.

Table 3. Correct recognition rates of six characters as percentage

Yield Characters	Training Set	Test Set
Plant Height	100	100
Number of Pods per Plant	100	40
Number of Pods on Main Stem	100	100
Number of Seeds per Pod	100	100
Pod Length	100	100
Thousand Seed Weight	100	100
Average	100	90

4 Discussion

It is known that varieties of different plants have been successfully classified by using various computer-based algorithms. Especially, classification of wheat varieties with computer algorithms has been become popular in recent years (Zayas et al. 1996, Utku and Köksel 1998, Neuman and Bushuk 1987, Gülmezoğlu and Gülmezoğlu 2015). Therefore, in this study, first of all, five rapeseed varieties were classified by using CVA method. In spite of 100% correct recognition rate in the training set, very low recognition rate (25.7%) was obtained in the test set. The reason is that samples included in the characters representing different varieties are very close to each other. Thus, common properties or invariant features of each variety cannot be correctly extracted and indifference subspace cannot be constructed efficiently.

Additionally, characters were classified by using CVA method. Initially, seven characters are applied to the classification process and 100% and 77.1% recognition rates were obtained for the training and test sets respectively. The reason of low score for the test set is that the samples belonging to number of branches per plant character are similar to samples belonging to other characters. When the number of branches per plant character is discarded, that is, when the remaining six characters are classified, 90% recognition rate is obtained for the test set.

5 Conclusion

It is concluded that the CVA method was very successful in the classification of different varieties belonging to any plant and/or of different characters belonging to any variety. Such classifications can be very helpful in assignment of unknown varieties or unknown characters to identify plant. When more specific characters are extracted for each variety of plants, good performance can be achieved from the classification process.

As a future work, number of varieties for any plant and the number of characters will be increased. Satisfactory results are also expected from this work.

References

1. Ali, I., Shah, S.A., Ahmed, H.M., Rehman K.U., Ahmad, M. (2012) Studies On Genetic Diversity For Seed Quality In Rapeseed (Brassica Napus L.) Germplasm of Pakistan Through Near Infrared Spectroscopy And Principal Component Analysis. Pak. J. Bot., 44: 219-222, Special Issue March 2012.
2. Delwiche, S.R., Massie, D.R. (1996) Classification of wheat by visible and near-infrared reflectance from single kernels. Analytical Techniques and Instrumentation 73(3) 399-405.

3. Grant, C.A., Bailey, L.D. (1993) Fertility management in canola production. Canadian Journal of Plant Science 73:651-670.
4. Gül, M.K. Egesel, C.Ö., Kahriman, F., Tayyar, Ş. (2007) Investigation of Some Seed Quality Components in Winter Rapeseed Grown in Çanakkale Province. Akdeniz University Journal of the Faculty of Agriculture., 2007, 20(1), 87-92.
5. Gülmezoğlu M.B., (1999), “A Novel Approach to Isolated Word Recognition”, IEEE Trans. Speech and Audio Processing, vol. 7, No.6, pp. 620-628.
6. Gülmezoğlu, M.B., Dzhafarov, V., Barkana, A. (2001) The Common Vector approach and its Relation to Principal Component Analysis. IEEE Trans. Speech and Audio Processing 9(6), 655-662.
7. Gülmezoğlu, M.B., Gülmezoğlu, N. (2015) Classification of bread wheat varieties and their yield characters with the common vector approach. International Conference on Chemical, Environmental and Biological Sciences (CEBS-2015), March, 18-19, 2015, Dubai, BAE.
8. Jankulovska, M., Ivanovska, S., Marjanovic-Jeromela, A., Bolaric, S., Jankuloski, L., Dimov, Z, Bosev, D., Kuzmanovska, B. (2014) Multivariate Analysis Of Uantitative Traits Can Effectively Classify Rapeseed Germplasm. Genetika, Vol. 46, No.2, 545-559.
9. Neuman M. and Bushuk, W. (1987) Discrimination of wheat class and variety by digital image analysis of whole grain samples. Journal of Cereal Science 6(2) 125-132.
10. Sharghi, S., Shirani, A.M.R., Noormohammadi, G. Zahedi, H. (2011) Yield and yield components of six canola (Brassica napus L.) cultivars affected by planting date and water deficit stress. African Journal of Biotechnology. 10(46): 9309-9313.
11. Shuaib, M, Jamal, M., Akbar, H., Khan, I., Khalid, R. (2010) Evaluation of wheat by polyacrylamide gel electrophoresis. African Journal of Biotechnology 9(2) 243-247.
12. Utku, H., Köksel, H. (1998) Use of statistical filters in the classification of wheats by image analysis. Journal of Food Engineering 36(4) 385-394.
13. Wang, D., Dowell F. E., Lacey, R. E. (1999) Single wheat kernel color classification using neural networks. Transactions of the ASABE 42(1) 233-240.
14. Zayas, I. Y. Martin, C. R. Steele J. L., Katsevich, A. (1996) Wheat classification using image analysis and crush-force parameters. Transactions of the ASABE 39(6) 2199-2204. eligible