Advance Journal of Food Science and Technology 11(11): 728-733, 2016
DOI:10.19026/ajfst.11.2771
ISSN: 2042-4868; e-ISSN: 2042-4876

Submitted: June 17, 2015 Accepted: July 14, 2015 Published: August 15, 2016

Research Article

Modeling Analysis of Mungbean in Regional Trials with Partial Least Squares Regression

[1]Juanqin Wang, [1]Zhiyong Zhang and [2]Xiaoli Gao
[1]College of Information Engineering,
[2]College of Agronomy, Northwest A&F University, Yangling, Shaanxi 712100, China

Abstract: The study is conducted to determine the reactions of different mungbean varieties to the test-environment in regional trials, the agronomic traits of different kinds of mungbean expressed and the impact of different agronomic traits on the output at the same site, ultimately provide improved methods and ideas for mungbean cultivation and help improve the experimental design. Based on the characteristics of experimental samples in regional trials and the data analysis, partial least squares regression is adopted to analyze the experimental data of the national regional trials of mungbean in spring from 2009 to 2011 and derived the production standard regression coefficients graph, established production model analysis of four high-yielding varieties of mungbean between the agronomic traits and the yield. The study is valuable for clarifying the characteristics of different varieties, evaluating varieties objectively. Identifying the main factors affecting yield performance, fully demonstrating the regularity of test data. All would provide improved methods and new ideas for mungbean breeding and the designing of regional pilot programs and provide the basis for the improvement of the production and cultivation techniques.

Keywords: Mungbean, partial least squares regression, regional trials, standard regression coefficient, yield model

INTRODUCTION

Mungbean is an important cultivated crop in arid and semiarid regions of our country. With the characteristics of wide adaptability, longer suitable sowing period and short growth period, strong drought resistance, mungbean plays an important role in the agricultural restructuring and in improving the production, quality and farming efficiency (Lin *et al*., 2002). Mungbeans is a kind of nutritious with high food and medicinal value and the processing technology is simple. Nation-wide regional trials of mungbean varieties is an experiment conducted at different ecological zones jointly for identification of varieties suitable area and for providing scientific evidences for mungbean cultivation (Gao *et al*., 2005). This study intends to adopt partial least squares regression analysis, through decomposing and screening, this method could extract the most powerful explanatory variables to fully demonstrate the regularity of regional trials data, that is the rule of different test environments for mungbean varieties (agronomic traits) influence, performance of agronomic traits of different varieties of mungbean under the test environment and the performance of agronomic traits of mungbean cultivars. The results are purposed to provide academic references for agricultural work and research.

LITERATURE REVIEW

In order to obtain accurate information of tested varieties from the regional trials, appropriate statistical methods are evolving. Methods employed to objectively evaluate the yield and stability of the tested varieties include rank analysis (Jin, 2000), variance analysis (Hu *et al*., 2009), grey correlation degree method, similarities, differences evaluation (Sun, 2011), biplot method (Yan, 2010; Zhang *et al*., 2010; Chen *et al*., 2009; Yan, 2001), etc. Using multiple regressions to analyze regional trials data has not been reported yet.

In regional trials, using multiple regression analysis to identify the factors influencing the yield and to establish yield model would have theoretical significance for conducting regional trials and the variety extension. However, there is multi-collinearity between economic and physiological traits with regard to the yield of mungbean. Principal component regression and partial least squares regression is an effective way to solve the multiple correlation regression problems. however, the method of partial least squares regression absorb the advantage of principal component regression of extracting, simplifying data structure, the principal component

Corresponding Author: Zhiyong Zhang, College of Information Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China

regression can not identify the noise and the lack of information of data and the partial least squares regression is more advanced than principal component regression method in that point. Using the method of the partial least squares regression, we do the several researches about traits of mungbean relative to yield in regional trials as following.

Setting up regression models of the varieties of mungbean relatives to yield by grouping similar area, in order to understand the impact of environment on the traits of mungbean. The regression model of four high yield of mungbean traits relative to yield was built in the planting area, in contrast to understand the different varieties of mungbean traits perform on production, the influence of mungbean which traits is more important on yield in this environment. To discuss and contrast similar statistical analysis method of regional trials, show the advantages and disadvantages of various methods and provide reference and ideas for the breeding and agricultural research.

EXPERIMENT

From year 2009 to 2011, there were 11 kinds of spring varieties mungbean tested in the nation-wide regional trials, namely LD05-01 (Zhenglu 9th), LD05-02 (Bao 200017-9), LD05-03 (Pinlu 2005-353-1), LD05-04 (CK) (Bailu 522 (CK)), LD05-05 (Fen mungbean 2nd), LD05-06 (Tao 9947-6), LD05-07 (Jilu 9802-19-2), LD05-08 (Bailu 8th), LD05-09 (Weilu 2116), LD05-10 (An 07-3B), LD05-11 (Sulu 04-23), respectively. The experiments were conducted at 12 piloting sites, namely Harbin of Heilongjiang province, Baicheng and Gongzhuling of Jilin province, Shenyang of Liaoning province, Zhangjiakou of Hebei province, Datela, Chifeng and Wulanhaote of Inner Mongolia, Datong and Fenyang of Shanxi province, Yulin and Yan'an of Shaanxi province. After three years of regional trials, it was found that 4 mungbean varieties, namely LD05-04 (Bailu 522), LD05-05 (Fen mungbeans 2nd), LD05-06 (Tao 9947-6) and LD05-08 (Bailu 8th) have a relatively high yield.

Experimental design: Experiments were conducted on mungbean producing areas from 2009 to 2011 in spring, using randomized block experiment with three duplications, covering an area of 10 m^2 (2 m×5 m) for each plot. The mungbean was sowed in drills with a row space of 50 cm, the planting density was in accordance with the local production habits. Strictly following the implementation rules of the National Spring Mungbean Regional Trials, all the field management, investigation and recording were standardized. The main traits explored included the growing period, plant height, number of branching on the main stem, number of nodes on the main stem, number of pods per plant, pod length, seed number per pod, kernel weight, yield etc.

Variables and the experimental data set: The main agronomic traits of Mungbean were selected as the independent variable, they were the growing period as x_1 (days), plant height as x_2(cm), number of branching on the main stem as x_3, number of nodes on the main stem as x_4 (section), number of pods per plant as x_5, pod length as x_6 (cm), seed number per pod as x_7 (grains), kernel weight as x_8 (g) and the yield was selected as the dependent variable.

The data of mungbean regional trials were collected from 2009 to 2011, namely the date about the 8 agronomical traits of 11 mungbean varieties at 12 trial sites in 3 years, the data comes from the state of mung bean (spring group) variety regional trials (http://www.mgcic.com/). Because the regional trials were reflected by the mean value of the combined treatment of varieties-environment (Hu *et al*., 2009), the yields of 11 varieties of mungbean in 12 regional trials were grouped according to geographical regions, latitude and longitude, each four regions were treated as a group (Table 1). Using 8 kinds of traits of mungbean as independent variables, yields as the dependent variables in four regions, comprehensive modeling and analysis were conducted. That is using the independent variables (x_1-x_8) and the dependent variables (y_1-y_4) for analysis. Four varieties of mungbean with higher average yield in the trial were selected for analysis, modeling and analysis for each single species were conducted, x_1-x_8 as the independent variables and y as the dependent variable.

The results of collinearity analysis were shown in Table 2. The maximal condition index was 214.36, which is greater than 100, indicating that there are serious multicollinearity among independent variables. In the table, the values of variance proportion of independent variables x_6, x_7 and x_8 are larger in the 9th line, indicating that there exists serious multicollinearity among several independent variables, x_1 and x_2 have large values in the 8th line, x_2, x_3 and x_4 have large values in the 7th line, indicating there are multiple collinearity among them.

There exists apparent collinearity among the independent variables. The number of sample data is small, so it is not suitable to use the traditional method of multivariate regression for analysis and modeling. The partial least squares regression which combines the features of principal component analysis, canonical correlation analysis and multiple linear regression functions, was used to effectively solve collinearity problem among the variables and improve modeling accuracy, robustness and practicality.

Statistical analysis: SAS9.1 statistical analysis software was used to conduct collinear analysis (reg), standardization (standard), Partial Least Squares regression (PLS) and draw a scatter diagram (gplot) (Der and Everitt, 2002; Hu, 2011; Hu and Wang, 2001; Aczel, 1989; Debiolles *et al*., 2004).

Table 1: Analysis of the regional data and cumulative explanation of the extracted factors

Group	Area	F	P	R^2	X	Y
The first group	Datong of Shanxi province	11.81	<0.0001	0.73	91.9%	61.8%
	Fengyang of Shanxi province	13.55	<0.0001	0.76		
	Yulin of Shaanxi province	8.62	<0.0001	0.66		
	Yan'an of Shaanxi province	1.82	0.1231	0.17		
The second group	Harbin of Heilongjiang province	4.01	0.0038	0.43	90.3%	51.2%
	Baicheng of Jiling province	6.00	0.0003	0.56		
	Gongzhuling of Jiling province	8.14	<0.0001	0.64		
	Shenyang of Liaoning province	3.16	0.0137	0.35		
The third group	Zhangjiakou of Hebai province	4.18	0.003	0.44	84.6%	46.0%
	Dalatet of Inner Mongolia	5.38	0.0006	0.52		
	Chifeng of Inner Mongolia	6.93	<0.0001	0.60		
	Wulanhaote of Inner Mongolia	1.39	0.2502	0.09		

Table 2: Results of collinearity analysisi

				Variance proportion							
No.	Eigen value	Condition index	Intercept	x1	x2	x3	x4	x5	x6	x7	x8
1	8.92	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.03	16.68	0.00	0.00	0.01	0.01	0.00	0.07	0.00	0.00	0.01
3	0.03	17.51	0.00	0.00	0.03	0.00	0.01	0.00	0.00	0.00	0.01
4	0.01	33.83	0.00	0.03	0.02	0.09	0.00	0.01	0.07	0.00	0.00
5	0.00	54.18	0.00	0.03	0.02	0.05	0.02	0.00	0.09	0.02	0.22
6	0.00	57.13	0.01	0.04	0.01	0.23	0.00	0.78	0.05	0.00	0.02
7	0.00	103.26	0.01	0.05	0.50	0.38	0.61	0.09	0.00	0.08	0.01
8	0.00	134.76	0.13	0.85	0.42	0.00	0.32	0.00	0.05	0.15	0.00
9	0.00	214.36	0.84	0.00	0.00	0.24	0.04	0.05	0.73	0.74	0.73

Microsoft Excel 2003 was employed to draw histograms and line chart for the derived data and VIP standardized regression coefficients for better comparison.

EXPERIMENTAL RESULTS AND ANALYSIS

Analysis of regional mungbean yield model: According to the experiment procedure, partial least squares regression analysis of the grouped data set is listed in Table 1. Data collected from Yan'an of Shaanxi province and from Wulanhaote of Inner Mongolia were not statistically significant, others were statistically significant. Four main factors were extracted when we did partial least squares regression analysis for each group of 33 sets of data, namely t_1, t_2, t_3 and t_4, they were sum of cumulative explanatory power for independent and dependent variables shown in Table 1.

The relational diagram between t_1 and u_1 was established, the diagram showed obvious linear form. It indicated that there is a strong correlation between the independent variable X and the dependent variable Y. All this proved that it is reasonable to establish the linear model of the Y about X by partial least squares regression method.

Plotted line graphs of the standardized regression coefficients derived from partial least squares regression analyses on three sets of data were drawn (Fig. 1 to 3). In the graph, x_1-x_8 stand for the eight traits of mungbean, the curve indicates the role of x_i in the output. It can be seen from the figure, the eight main traits of mungbeans have different influence on mungbean yield in different areas, but within the same group, the trend of the curve is basically the same.

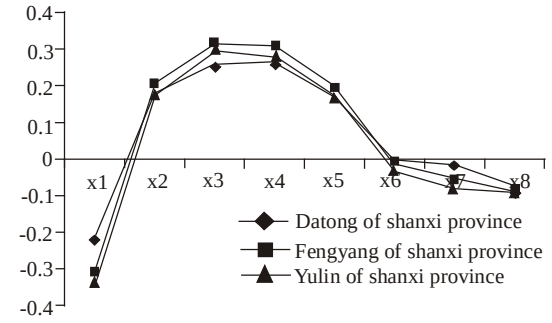

Fig. 1: Line chart of the standardized regression coefficients of group one

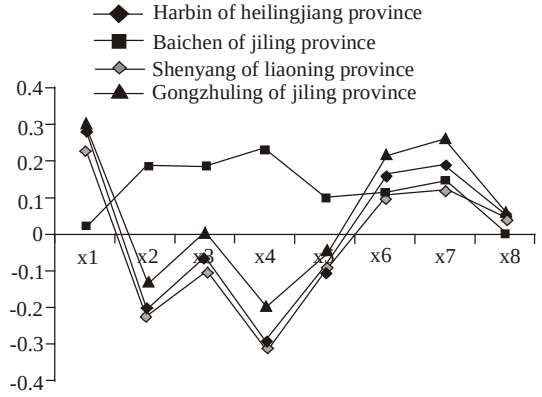

Fig. 2: Line chart of the standardized regression coefficients of group two

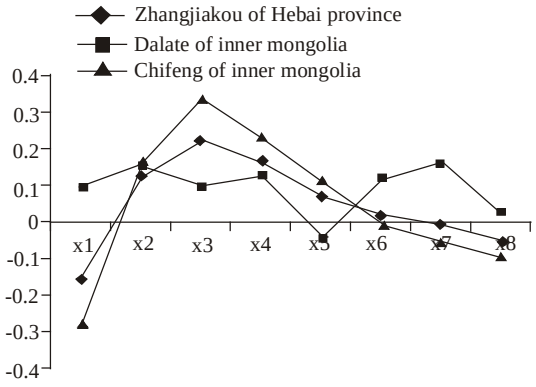

Fig. 3: Line chart of the standardized regression coefficients of group three

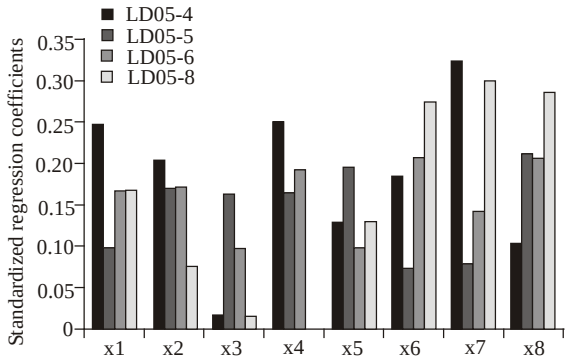

Fig. 4: Histogram about standard regression coefficient of four kinds of mungbean

Figure 1 shows that plant height (x_2), number of branching on the main stem (x_3), number of nodes on the main stem (x_4) and number of pods per plant (x_5) all have significant and positive impact on mungbean yield in Fenyang and Datong of Shanxi province and in Yulin of Shaanxi province, but growing period (x_1) demonstrates negative affect on the mungbean yield. The impact of pod length (x_6), seed number per pod (x_7) and kernel weight (x_8) on the yield is not significant. This indicates that the plant height at Datong and Fenyang of Shanxi province and at Yulin of Shaanxi province is comparatively larger, podding is mainly on the main stem and stem branches, so production depends mainly on the number of pods per plant. It means that extending the growing period for yields is meaningless.

It can be seen from Fig. 2 that the growing period (x_1), pod length (x_6) and seed number per pod (x_7) all have significant and positive impact on mungbean yield in Haerbin, Shenyang and Gongzhuling, plant height (x_2), number of nodes on the main stem (x_4) have negative impact, indicating that the northeast region has lower temperatures, mungbean has a longer growing period, pods are longer and bearing more grains and thus obtain high yield. Baicheng of Jilin province is more special, eight traits of mungbean all have positive effects on yield and plant height (x_2), number of branching on the main stem (x_3), number of nodes on the main stem (x_4), number of pods per plant (x_5), pod length (x_6) and seed number per pod (x_7) have greater impact on yield and the remained two traits demonstrate weaker impact on the yield.

Plant height (x_2), number of branching on the main stem (x_3), number of nodes on the main stem (x_4) and number of pods per plant (x_5) have important implications for mungbean production at Zhangjiakou of Hebei province and Chifeng of Inner Mongolia, but the growing period (x_1) has a negative impact, pod length (x_6), seed number per pod (x_7) and kernel weight (x_8) have weak effect. Mungbean production at Dalate in Inner Mongolia demonstrated special features. The number of pods per plant (x_5) has negative impact and the impact of the kernel weight (x_8) is weak, the other traits all have significant and positive effects on the yield (Fig. 3).

Through the exploration of the relation between the production yield and the agronomical traits of the selected 11 mungbean varieties in different places, the correlation models between the expression of the agronomic traits of mungbean and the local temperature, regional precipitation and soil types would be festablished, thus to provide reference for mungbean variety selection and cultivation. In practice, proper mungbean variety should be selected in accordance with the mungbean variety features, local temperature and soil conditions, ultimately they make full use of the variety features and increase the production.

Production model analysis of four high-yielding mungbean varieties: Based on the data collected from 2009 to 2011 in the regional trials mungbean, partial least squares regression analysis for each species of LD05-04, LD05-05, LD05-06 and LD05-08 and the histogram of standardized regression coefficients was shown in Fig. 4.

Partial least squares regression analysis for each of the four varieties of mungbean was conducted and the derived t_1/u_1 diagram is linear. The total of cumulative explanatory power of the two principal factors extracted from the independent and dependent variables has reached 100% and the histogram of the standardized regression coefficients of four varieties and the VIP_j values was shown in Fig. 5.

From Fig. 4 and 5, we can see that in production model LD05-04 and LD05-08, the seed number per pod (x_7), growing period (x_1), pod length (x_6), plant height (x_2), kernel weight (x_8) and number of pods per plant (x_5) have significant impact on the yield, the number of branching on the main stem (x_3) has less impact on the yield. The number of nodes on the main stem (x_4) has comparatively meaningful impact on the yield in LD05-04, while it has very small impact on the yield in model LD05-08 that even could be neglected. However, the

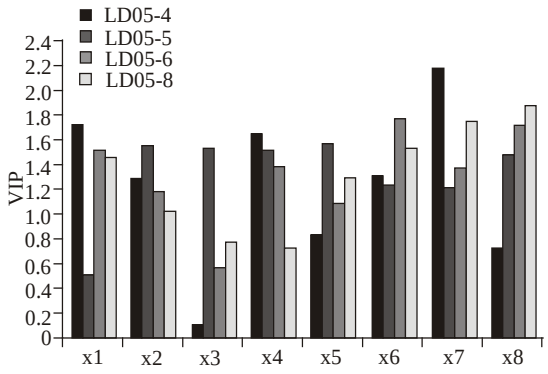

Fig. 5: VIP_j histogram about independent variables for four kinds of mungbean

impacts of all the eight traits are even in model LD05-05 and model LD05-06.

It showed that different mungbean varieties have different plant types with diversified podding habits. Some varieties are mainly dominated by the podding on the main stem pod, the podding on the branches is less and thus the number of branching of its main stem has insignificant influence on the yield. For other varieties, the number of podding on the stem is small but larger on the branches, so the number of branching on the stem will have significant impact on the yield. Therefore, in order to get more objective evaluation of mungbean varieties, the characteristics of different mungbean varieties should be considered when to design pilot programs in regional trials. In practice, to make full use of the variety potential, the growth and podding characteristics of different mungbean varieties should be considered.

DISCUSSION

Principal component regression and partial least squares regression analysis may analyze the complex data which contain many varieties, the environment and different years, they can find the interdependencies of two groups of multiple correlation variables, research with a set of variables which were named as independent variables or predictors to predict another set of variables which were named as the dependent variable or response variable. When it conduct the analysis model, the principle of the principal component regression uses the result of principal components extracted by the principal component analysis and dependent variable to model, the dependent variable are not considered when it extracted the principal components, it found a representative principal component for the independent variables, This may cause the result of that the principal component elected was independent of the dependent variable, or the principal component may reflect the important information of independent variable, but the relationship with the dependent variable was very small, meanwhile, the principal components which have the large correlation with the dependent variable while they have small proportion of the independent variable, may be deleted (Wang *et al*., 2008).

PLS is an optimal combination of the multiple linear regression, principal component analysis and canonical correlation analysis. When the number of independent variables and the dependent variable is large and there exists multiple correlations between the independent and dependent variables, what's more, if the number of observations (sample size) is small, the modeling established through partial least squares regression is proved to have advantages over other traditional regression analysis methods. Guided by the principles of the canonical correlation analysis, it uses the method of gradual extraction for the factors, thus to make sure the maximal correlation between the independent variable and dependent variable (Wang *et al*., 2008; Wold *et al*., 1983), simplifying the data structure through data analysis and make the relationship among variables be more observable. All these features make partial least squares regression become a good way for the data analysis of crop regional trials.

There exists observable collinearity among the eight traits of mungbean from regional trials data, moreover, the number of data collected is small. The use of partial least squares regression analysis can effectively remove the noise in the test data and gradually extract the main factors of data and ensure maximal relevance between mungbean agronomic traits and its yield, quantitatively analyze the relationship between 8 agronomic traits of mungbean and the yield, produce effective data models of the regional trials, provide good guidance for mungbean production and variety testing.

CONCLUSION

In practice, the selection of mungbean variety should be in accordance with the local geographic features and diverse planting approaches and rational fertilization methods should be employed with regard to the agronomical features of different varieties, make full use of the advantages of each variety to increase yield. In addition, the use of partial least squares regression to establish the regression equation with the yield as the dependent variable and other economic traits and physiological traits as independent variables, can improve the ability of researchers to understand and manage test data. All would provide improved methods and new ideas for mungbean breeding and the designing of regional pilot programs and provide the basis for the improvement of the production and cultivation techniques.

ACKNOWLEDGMENT

This study was supported by the Fundamental Research Funds under Grant No. QN2013052 and The Ministry of Agriculture "State of Minor Grain and Bean Crops Variety Regional Trials" Project and Doctor Scientic Research Startup Project (No.2013BSJJ106).

REFERENCES

Aczel, A.D., 1989. Complete Businese Statistics. Irwin, Boston, pp: 1-57.

Chen, S.L., Y.R. Li, Z.S. Cheng and J.S. Liu, 2009. GGE biplot analysis of effects of planting density on growth and yield components of high oil peanut. Acta Agron. Sinica, 35(7): 1328-1335. (In Chinese)

Debiolles, A., L. Qukhellou and P. Aknin, 2004. Combined use of partial least squares regression and neural network for diagnosis tasks. Proceeding of the 17th International Conference on Pattern Recognition, 4: 573-576.

Der, G. and B.S. Everitt, 2002. A Handbook of Statistical Analyses Using SAS. Chapman &Hall/CRC, Boca Ration, pp: 240-265.

Gao, X.L., J.M. Sun, J.F. Gao, B.L. Feng, Y. Chai and P.T. Zhang, 2005. Comprehensive analysis and evaluation of mungbean varieties in regional trials. Acta Agr. Boreali-Occidentalis Sin., 14(1): 167-171. (In Chinese)

Hu, L.P., 2011. SAS Statistical Analysis Procedure. Publishing House of Electronics Industry, Beijing.

Hu, X.P. and C.F. Wang, 2001. An Introduction to SAS Base and Statistics Procedure. Xi'an Map Press, China. (In Chinese)

Hu, X.Y., H.L. You, X.F. Song and J.P. Li, 2009. Comparison of different models for crop stability analysis. J. Triticeae Crops, 29(1): 110-117. (In Chinese)

Jin, W.L., 2000. The rank analysis model of evaluating crop varieties yield stability in regional trials. Acta Agron. Sinica, 26(6): 925-930. (In Chinese)

Lin, R.F., Y. Chai, Q. Liao and S.X. Sun, 2002. Minor Grain Crops in China. China Agricultural Science and Technology Press, Beijing, pp: 192-209. (In Chinese)

Sun, J.M., 2011. Development and implementation of regional trial data collection and analysis system for minor grain crops. M.A. Thesis, Northwest University of Science and Technology, pp: 5. (In Chinese)

Wang, H.W., J. Wang and H.J. Huang, 2008. Modeling strategy of principle component regression. J. Beijing Univ. Aeronaut. Astronaut., 34(6): 661-664.

Wold, S., C. Albano and M. Dunn, 1983. Pattern Regression Finding and Using Regularities in Multivariate Data [M]. Analysis Applied Science Publication, London.

Yan, W., 2001. GGEbiplot-a windows application for graphical analysis of multienvironment trial data and other types of two-way data. Agron. J., 93: 1111-1118.

Yan, W.K., 2010. Optimal use of biplots in analysis of multi-location variety test data. Acta Agron. Sinica, 36(11): 1805-1819. (In Chinese)

Zhang, Z.F., X.F. Fu, J.Q. Liu and H.S. Yang, 2010. Yield stability and testing-site representativeness in national regional trials for oat lines based on GGE-biplot analysis. Acta Agron. Sinica, 36(8): 1377-1385. (In Chinese)