

# Path Coefficient and Ridge Regression Analysis to Improve Seed Yield of *Psathyrostachys juncea* Nevski

WANG Quanzhen<sup>2,2</sup>, CUI Jian<sup>1</sup>, ZHOU He<sup>2</sup>, WANG Xianguo<sup>2</sup>, ZHANG Tiejun<sup>3</sup> and HAN Jianguo<sup>2\*</sup>  
<sup>1</sup>College of Animal Sci. and Techn., China Agricultural University, Beijing 100094, China, <sup>2</sup>College of Animal Sci. and Techn., Northwest A&F University, Yangling 712100, Shaanxi Province, China, <sup>3</sup>Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, PR China  
E-mail: [wangquanzhen191@163.com](mailto:wangquanzhen191@163.com)

## Abstract

Based on multi-factor orthogonal designed field experimental blocks, the yield components and their direct and indirect influences on the seed yield of *Psathyrostachys juncea* Nevski. were investigated under variable growing conditions. In each block the yield components: fertile tillers/m<sup>2</sup> (y1), spikelets/fertile tillers (y2), florets/spikelet (y3), seed numbers/spikelet (y4), seed weight (y5) and seed yield were determined by hand in 2003. The results show that in *P. juncea* . seed yield is significantly correlated with yield components y1 (0.749\*\*\*), y2 (0.159\*) and y5 (0.231\*). All of the ridge regression coefficients are >0, which means that increasing any one of the yield components (y1~y5) will increase seed yield, in accordance with biological theory. This study indicates that ridge regression is one of the most promising methods available to unravel the tangled skeins of inter-correlated factors.

## Introduction

*Psathyrostachys juncea* Nevski is a cool-season forage species well adapted to semi-arid climates (Wang et al., 2004). We are interested in investigating the relationships between the seed yield and its components to improve the seed yield of this forage grass.

The advantage of path analysis is that it permits the partitioning of the correlation coefficient into its components, one component being the path coefficient that measures the direct effect of a predictor variable upon its response variable; the second component being the indirect effect(s) of a predictor variable on the response variable through another predictor variable (Milligan et al., 1990). Path analysis has been used by plant breeders to assist in identifying traits that are useful as selection criteria to improve crop yield.

For grass crops, the correlation of economic yield components with grain yield and the partitioning of the correlation coefficient into its components of direct and indirect effects have been extensively reported: e.g. highly significant associations of grain yield were observed with

1000-grain weight and tiller number per plant (Das and Taliaferro, 2009), the number of filled grains per panicle (Wu et al., 2008) and harvest index (Sured et al., 1998).

However, morphological characters influencing yield are often highly inter-correlated, leading to multi-collinearity when the inter-correlated variables are regressed against yield in a multiple-regression equation.

In this study, based on multi-factor orthogonal field experimental design, an attempt was made to study the direct and indirect influences of some important yield components on seed yield in *P. juncea* via correlation, path coefficient and ridge regression analyses with big sample sizes under various growing conditions (field managements).

## Materials and Methods

Using 5 groups of multi-factor orthogonal field experimental designed blocks, the field experiment was set up in Jiuquan (39°37'N, 98°30'E), Gansu Province, China from 2002 to 2003. In total 112 experimental blocks with an area of 28 m<sup>2</sup> were sown in spring (Table 1). Seed yield was recorded in autumn 2002 and averaged 165kg ha<sup>-1</sup>.

According to the orthogonal experimental designs, yearly repeated, under various field managements, conditions from controlled environments, including regimes of fertilizer (experimental factor: X1, X3, and X4), irrigation system (experimental factor: X2), plant density (experimental factor: X5) and plant growth regulators (experimental factor: X6) (Table 1). In each block was measured the yield components: Fertile tillers/m<sup>2</sup> (y1), spikelets/fertile tillers (y2), florets/spikelet (y3), seed numbers/spikelet (y4), seed weight (mg) (y5). Seed yield (kg/hm<sup>2</sup>) (z) was determined by hand in field, randomly based on order, from anthesis to seed harvest in year 2003. The sample size of y1~y5 and z for field experimental block in *P. juncea* are 10, 36, 27, 54, 10 and 4, respectively. Seed weight was measured from 100-grains at a moisture content of 7~10. Seed yield was measured from hand-harvesting a subsample of 1 m<sup>2</sup>. The statistical analysis was performed using Visio FoxPro (Version 6.0) and SAS (Release 8.1, SAS Institute Inc, 1988 (Table 2). Based on Pearson Correlation Analyzed by SAS (Table 3), a QBASIC program was written for path coefficient analysis (Table 4). Furthermore, via SAS, ridge regression analysis was performed on seed yield (z) and its components (y1~y5).

## Results

The results show that in *P. juncea* seed yield is significantly correlated with yield components y1 (0.749\*\*\*), y2 (0.159\*) and y5 (0.231\*). The order of correlation coefficients is: y1>y5>y2>y3>y4 (table 3). The order of direct effects of yield components (y1~y5) on seed yield is y1>y5>y3>y2>y4 (Table 4) by path coefficient analysis, the biggest direct effects to seed yield are the yield component y1 (0.774) and, the biggest indirect effects to seed yield are y3

through  $y_4$  (0.133). Increasing  $y_1$  is the most productive on the seed yield in the components followed by the  $y_5$  and  $y_3$ .

In ridge regression analysis, although the optimal value of  $k$  cannot be determined with certainty, several procedures have been proposed for the selection of  $k$  (Hoerl and Kennard 1970 a, b) have suggested that  $k$  is determined from the ridge trace, with  $k$  selected so that a stable set of regression coefficients was obtained (Newell and Lee, 1981). In this study, according to the ridge trace, for various values of  $k$ , with the values of  $k$  estimated as 0.5, using the method of Hoerl and Kennard (1970a, b), the standard ridge regression models are lined in the figures. The resulting ridge regression coefficients are -892.63, 2.19, 4.61, 15.46, 3.20 and 263.96 for intercept,  $y_1$ ,  $y_2$ ,  $y_3$ ,  $y_4$  and  $y_5$  respectively. Following the ridge regression models is:  $Z = -892.63 + 2.19y_1 + 4.61y_2 + 15.46y_3 + 3.20y_4 + 263.96y_5$  ( $F=33.11$ ,  $Pr<0.0001$ ).

## Discussion

The various field experimental managements ( $X_1 \sim X_6$ ) created a very wide range of seed yields and yield components (Table 2), with a maximum seed yield of 1969.65 kg/hm<sup>2</sup>, and a minimum of 358.41 kg/hm<sup>2</sup> (little irrigation, no fertilizer and low plant density) (table 2).

With multi-factor orthogonal experimental designs and big sample statistics analysis in field experiment, the significant correlation coefficients (at  $P=0.0001$  and 0.01) show the results are reliable. Ridge regression effectively overcome the problem of highly multi-correlated predictor variables like yield components (Hoerl and Kennard, 1970a,b).

In this study all of ridge regression coefficients are  $>0$ , that means increasing any one of the yield component ( $y_1 \sim y_5$ ) will increase seed yield, in accordance with biological theory. In the result of the ridge regression model, increasing every 1 unit of yield components ( $y_1 \sim y_5$ ), respectively will increase seed yield ( $Z$ ) by 2.19, 4.61, 15.46, 3.20, 263.96 kg/hm<sup>2</sup>, respectively; The component  $y_2$  may develop in a oversaturated situation or from reasons of climate; By comparisons, it is in average 90.2 spikelets/fertile tillers in this study, whereas *P. juncea* in the second year of growth the spikelets/fertile tillers ( $y_2$ ) are 20~36, 27.5~34.5 and 53.6~60.2 in Xinjiang Weiwuer Autonomous Region (44°31'N, 87°5'E), northwest of China (Zhang et al, 2002), in Yuershan Farm (41°44'N, 140°16'E), Hebei province, east of China (Mao et al, 2000), and in Pulandian Farm in Dalian city (39°47'N, 121°54'E), Liaoning province, northeast of China (Fang et al, 2001),, respectively. So, it is needed to clarify the dynamic effect of weather factors such as temperature, rainfall and hours of sunshine, etc. on yield components of *Psathyrostaehys juncea* Nevski. in the period of the grass development. However, the order of the yield components:  $y_1 > y_5 > y_2 > y_3 > y_4$  of correlation and  $y_1 > y_5 > y_3 > y_2 > y_4$  of path coefficients, imply that  $y_1$  and  $y_5$  is most important yield components for improving seed yield, and  $y_4$  the least.

## **Conclusion**

Combined path coefficients and ridge regression model analyses, show that  $y_1$  is most important yield components for improving seed yield,  $y_5$  and  $y_3$  comes next, and  $y_4$  the least. All of ridge regression coefficients are  $> 0$ , that theoretically means increasing any one of yield component ( $Y_1 \sim Y_5$ ) will increase seed yield. Additionally, evidence from this study and others indicates that ridge regression is one of the most promising of the methods available to unravel the tangled skeins of inter co-related factors. Inter-correlated variables are very prevalent in the arrays of measured yield components and other agronomic characters used by herbage breeders trying to enhance the seed yield efficiency of crops.

## **Acknowledgements**

The National Basic Research and Development Programme (973 project, 2007CB106805) and the Key Laboratory of Grassland Agro-ecosystems, Ministry of Agriculture of P R China, funded this work. We are grateful to my skilful technical assistants, Mr. Zhang Bing, Miss Yan Xuehua, of Daye Institute of Forage & Grass Products in Jiuquan, Gansu Branch of Chengdu Daye International Interest Co. Ltd

## **References**

- Das, M.K., and C.M. Taliaferro (2009). Genetic variability and interrelationships of seed yield and yield components in switchgrass. *Euphytica* 167:95-105.
- Fang L N, Han J G, et al. (2001). Seed Yield Response of Smooth Brome to Nitrogen, Plant Growth Regulator and Environmental Influences, *Grassland of China*, 23(4):31~37
- Ghebremariam, W.K., G.F. Ortmann, and I.V. Nsahlai. (2006). A production function analysis of commercial dairy farms in the Highlands of Eritrea using ridge regression. *Agrekon* 45:225-242.
- Hoerl A.E., R.W.Kennard (1970a). Ridge regression: Applications to non-orthogonal problems. *Technometrics* 12:69-82.
- Hoerl A.E., R.W.Kennard. (1970b). Ridge regression: biased estimation for non-orthogonal problem. *Technometrics* 12:55-67.
- Mao P S, Han J G, Wang P et al, (2000), Effects of Fertilizer Application on Seed Yield of Smooth Brome grass and Siberian Wildrye, *Acta Agrestia Sinica*, 8:273-278
- Milligan S.B., K.A.Gravois, K.P.Bischoff, F.A.Martin. (1990). Crop effects on genetic relationships among sugarcane traits. *Crop Science* 30:927-931.

Newell G.J., Lee B. (1981) Ridge regression: an alternative to multiple linear regression for highly correlated data [in food technology]. *Journal of Food Science (USA)*, 46:968-969.

SAS Institute Inc. (1988). *SAS/STAT User's Guide*. North Carolina : SAS Institute Inc..

Sured H., Z.K.Korkut, O.Bilgin (1998). Correlation and path analysis for yield and yield components in rice in a 8-parent half diallet set of crosses. *Oryza*, 35:15-18.

Thimmegowda, M.N., V.K.K. Kumar, A.G. Bandi, and Sharanappa. (2008). Correlation, regression and path coefficient analysis of seed yield with other attributes as influenced by irrigation, land and water management practices in sunflower. *Research on Crops* 9:367-370.

Wang, Z.Y., J. Bell, and D. Lehmann. (2004). Transgenic Russian wildrye (*Psathyrostachys juncea*) plants obtained by biolistic transformation of embryogenic suspension cells. *Plant Cell Reports* 22:903-909.

Wu, Y.Q., C.M. Taliaferro, et al. (2008). Correlation analyses of seed yield and its components in bermudagrass. *Multifunctional grasslands in a changing world, Volume II: XXI International Grassland Congress and VIII International Rangeland Congress, Hohhot, China, 29 June-5 July 2008:562*.

Zhang X S, Dai L Y and Wang Z J, (2002), Smooth Bromegrass,. *Xinjiang Animal Husbandry* 4 : 28~29.

**Table.1** Field Experimental design and factors (*P. juncea*)

| Field experimental design groups        | Factors               | Repeat | blocks |
|-----------------------------------------|-----------------------|--------|--------|
| 1.2-D-optimum orthogonal design(1)*     | 2( $X_3, X_4$ )       | 3      | 18     |
| 2.2-D-optimum orthogonal design(2)*     | 2( $X_3, X_4$ )       | 1      | 6      |
| 3.Quinque-factor orthogonal design      | 5( $X_1 \sim X_5$ )   | 1      | 36     |
| 4.Bin-factor orthogonal contract blocks | 2( $X_2, X_3 + X_4$ ) | 1      | 22     |
| 5.Tri-factor orthogonal design          | 3( $X_1, X_3, X_6$ )  | 1      | 23     |
| CK                                      | --                    | 7      | 7      |
| total                                   | 6( $X_1 \sim X_6$ )   | --     | 112    |

Note: \*: Applied N and  $P_2O_5$  differently between design (1) and (2);  $X_1 \sim X_6$ : is time of fertilizing, quantity of irrigation, applied nitrogen, applied  $P_2O_5$ , planted density, amount of spray plant regulator Paclobutrazol(PP333), respectively.

**Table.2** Simple Statistics of  $Y_1 \sim Y_5$ ,  $Z$  (*P. juncea*)

| Variable | N   | Mean  | Std-Dev | Std-Error | Minimum | Maximum | Pr> t  |
|----------|-----|-------|---------|-----------|---------|---------|--------|
| y1       | 112 | 205.6 | 70.7    | 6.9       | 76.1    | 415.1   | <.0001 |
| y2       | 112 | 90.2  | 2.7     | 0.265     | 79.7    | 96.960  | <.0001 |
| y3       | 112 | 4.6   | 0.7     | 0.074     | 3.01    | 6.25    | <.0001 |
| y4       | 112 | 2.1   | 0.3     | 0.033     | 1.500   | 3.05    | <.0001 |
| y5       | 112 | 3.5   | 0.18    | 0.018     | 2.9     | 3.8     | <.0001 |
| z        | 112 | 964.4 | 336.93  | 32.9      | 358.4   | 1969.6  | <.0001 |

**Note :**  $Y_1$  : Fertile tillers/m<sup>2</sup> ;  $Y_2$  : Spikelets/Fertile tillers ;  $Y_3$  : Florets/spikelet ;  $Y_4$  : Seed numbers/spikelet ;  $Y_5$  : Seed weight(mg) ;  $Z$  : Seed yield(kg/hm<sup>2</sup>).

**Table.3** Pearson correlation coefficients of  $Y_1 \sim Y_5$ ,  $Z$  (*P. juncea*)

|    | y1    | y2      | y3     | y4       | y5     | z        |
|----|-------|---------|--------|----------|--------|----------|
| y1 | 1.000 | 0.309** | 0.106  | 0.131    | -0.008 | 0.749*** |
| y2 |       | 1.000   | -0.071 | -0.128   | -0.021 | 0.195*   |
| y3 |       |         | 1.000  | 0.927*** | 0.158  | 0.127    |
| y4 |       |         |        | 1.000    | 0.122  | 0.110    |
| y5 |       |         |        |          | 1.000  | 0.231*   |

\*\*\*, \*\* and \* significant at  $P=0.0001$ , 0.01 and 0.05, respectively

**Table.4** Path analysis showing direct and indirect effect of  $Y_1 \sim Y_5$ ,  $Z$  (*P. juncea*)

|    | y1→z         | y2→z          | y3→z         | y4→z          | y5→z         |
|----|--------------|---------------|--------------|---------------|--------------|
| y1 | <b>0.774</b> | -0.011        | -0.012       | -0.036        | -0.006       |
| y2 | 0.058        | <b>-0.052</b> | 0.002        | -0.028        | -0.001       |
| y3 | -0.051       | -0.014        | <b>0.207</b> | 0.133         | 0.041        |
| y4 | -0.063       | 0.009         | 0.057        | <b>-0.219</b> | 0.025        |
| y5 | 0.033        | -0.013        | 0.019        | 0.029         | <b>0.230</b> |