African Journal of Biotechnology Vol. 10(59), pp. 12584-12594, 3 October, 2011
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.1220
ISSN 1684–5315 © 2011 Academic Journals

Full Length Research Paper

Modelling of seed yield and its components in tall fescue (*Festuca arundinacea*) based on a large sample

Quanzhen Wang[1]*, Tianming Hu[1], Jian Cui[2], Xianguo Wang[3], He Zhou[3], Jianguo Han[3] and Tiejun Zhang[4]

[1]Department of Grassland Science, College of Animal Science and Technology, Northwest A and F University, Yangling 712100, Shaanxi, China.
[2]Department of Plant Science, College of Life Science, Northwest A F University, Yangling 712100, Shaanxi, China.
[3]Institute of Grassland Science, College of Animal Science and Technology, China Agricultural University, Beijing 100094, China.
[4]Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China.

Accepted 29 August, 2011

Tall fescue (*Festuca arundinacea* Schreb.) is a primary cool-season grass species that is widely used as a cold-season forage and turfgrass throughout the temperate regions of the world. The key seed yield components, namely fertile tillers m^{-2} (Y_1), spikelets fertile tiller $^{-1}$ (Y_2), florets spikelet $^{-1}$ (Y_3), seed number spikelet $^{-1}$ (Y_4), seed weight (Y_5), and the seed yield (Z) of tall fescue were determined in field experiments from 2003 to 2005. The experiments produced a large sample for analysis. The correlations among Y_1 to Y_5 and their direct and indirect effects on Z were investigated. All of the direct effects of the Y_1, Y_3, Y_4 and Y_5 components on the seed yield were significantly positive. However, the effect of Y_2 was not significant. In decreasing order, the contributions of the five components to seed yield are $Y_1 > Y_4 > Y_3 > Y_5 > Y_2$. Y_4 and Y_5 were not significantly correlated with Z. However, the components Y_1, Y_2 and Y_3 were positively correlated with Z in all the three experimental years and the intercorrelations among the components Y_1, Y_2 and Y_3 were significant. Ridge regression analysis was used to derive a steady algorithmic model that related Z to the five components; Y_1 to Y_5. This model can estimate Z precisely from the values of these components. Furthermore, an approach based on the exponents of the algorithmic model could be applied to the selection for high seed yield via direct selection for large Y_2, Y_3 and Y_5 values in a breeding program for tall fescue.

Key words. Modelling, seed yield, components, tall fescue, path and ridge analyses, large sample.

INTRODUCTION

Tall fescue (*Festuca arundinacea* Schreb.) is a primary and important cool-season forage grass species. It is grown for livestock production throughout the temperate regions of the world (Majidi et al., 2009). Because the grass thrives on impoverished soils in pastoral environments (under simultaneously occurring multiple stresses) (Belesky et al., 2010), tall fescue plays a significant role in soil conservation in arid and semi-arid regions. Tall fescue is also widely used as a cold-season turfgrass in residential and commercial landscapes. For turf-type tall fescue, previous research has focussed mainly on cultivation. The purpose of this previous research was to identify heat- and drought-tolerant selections that can produce a higher-quality turf. Specific research topics in this area have included the effects of organic fertilisers on greening quality, shoot, root growth, etc. (Cheng et al., 2010) and the genetic mechanism of brown patch resistance in tall fescue (Bokmeyer et al., 2009). Tall fescue also has the potential to serve as a sink for industrial pollutants, as reported in a study of lead uptake by the roots of turfgrass tall fescue (Qu et al., 2003).

Nevertheless, little research has been conducted regarding the algorithms of seed yield and its key components in grasses. This information is crucial to meet the demands of commercial propagation. Seed

---

*Corresponding E-mail: wangquanzhen191@163.com.

yield, a quantitative character, is largely influenced by the environment and thus has a low heritability (Bliss et al., 1973; Boelt and Gislum, 2010; Wang et al., 2010). Therefore, the response to direct selection for seed yield may be unpredictable unless environmental variation is well controlled. Thus, there is a need to examine the mathematical relationships among various characters. The investigation of such relationships involving seed yield and yield components, interior yield components, and a certain amount of interdependence is especially important. To date, although some research has focused on the seed yield and the yield components of tall fescue (Young et al., 1998a, b), no information is available on the algorithmic relationships between these characters.

Path analysis has been widely used by plant breeders to assist in identifying the traits that are useful as selection criteria in improving crop yield (Akinyele and Osekita, 2006; Bicer, 2009; Ceyhan, et al. 2008; Karasu, et al. 2009; Kaya, et al. 2010; Kokten et al., 2009; Mensah et al., 2007). However, morphological characters (that is, Y_1 to Y_5) influencing seed yield (Z), are often highly intercorrelated. This situation leads to multi-collinearity when the intercorrelated variables are regressed against yield in a multiple-regression equation (Wang et al. 2011). For such situations, the estimation of regression coefficients through ridge regression was developed by Hoerl and Kennard (1970 a, b) to ameliorate problems of multicollinearity. These problems may result in the inflation of the absolute value of the regression coefficients and may also produce incorrect signs for the regression coefficients resulting from these intercorrelated variables.

The objective of this study was to examine the mathematical relationships between seed yield and its components by using a path analysis and ridge regression modelling approach to forecast the seed yield in seed production. This approach offers a reference algorithm suitable for quantitative genetics and breeding in tall fescue and can stimulate further investigations of seed yield and its components in grasses.

MATERIALS AND METHODS

A multifactor, orthogonal design involving various field experimental management conditions (Hedayat et al., 1999) was used in this study. The seed yield components considered in this study, were fertile tillers m^{-2} (Y_1), spikelets fertile tiller^{-1} (Y_2), florets spikelet^{-1} (Y_3), seed number spikelet^{-1} (Y_4) and seed weight (Y_5) (Canode, 1980; Fairey and Hampton, 1997).The following theoretical formulas describe the relationships between the seed yield components and seed yield (or seed yield potential).

Seed yield: $Z_{SY} = Y_1 \cdot Y_2 \cdot Y_4 \cdot Y_5$

If one floret contents one seed embryo for grasses, then
Seed yield potential: $Z_{SYP} = Y_1 \cdot Y_2 \cdot Y_3 \cdot Y_5$

Research location and field conditions

A field experiment was conducted from 2003 to 2005 at the China Agricultural University Grassland Research Station located at Yinger village of Shangba Commune, in Jiuquan, Gansu province, northwestern China (latitude 39°37′ N, longitude 98°30′ E; elevation 1480 m). The initial soil at the site is Mot-Cal-Orthic Aridisols, classified as Xeric Haplocalcids in the USDA soil classification (Soil-Survey-Staff, 1996). The plots used in this experiment had been planted with ‘alfalfa’ (*Medicago sativa* L.) during the previous season.

The 0.6 ha experimental site was tilled using a chisel plough in the fall and a disk harrow in the spring for seedbed preparation. ‘Fawn’ tall fescue was planted on 23 April, 2002 at a planting depth of 2.5 cm and at a seeding rate of 15 kg ha^{-1}. The rows were 0.45 m apart and were planted in a south to north direction. Fertiliser was initially applied in a 6 cm-deep band and 5 cm to the side of the seed furrow at a rate of 104 kg hm^{-2} N and 63 kg hm^{-2} P_2O_5. There was no seed yield in autumn of 2002. This research trial was conducted during the next three years (2003 to 2005), using five groups (A to E) of designed field management regimes ($X_{1\text{-}6}$) that were repeated yearly.

Experimental design

For the simulation of various growing conditions, the experiment used five groups (A to E) of multifactor, orthogonal experimental-designed field block designs with six experimental factors, including the time of fertilisation (X_1), the quantity of irrigation (X_2), the amount of N applied (X_3), the amount of P_2O_5 applied (X_4), the seeding density (X_5) and the amount of plant growth regulator sprayed (X_6) (Hedayat et al., 1999; Lattin et al., 2003; Yandell, 1997).

Groups A and B each consisted of a 2-D-optimum design (a 2-D-optimum matrix applied with six plots) and arranged experimental factors X_3 and X_4 with different levels, respectively. Group A included three replicates [3 × 6 = 18 plots (treatments), stochastic arrangement]. Group B had one replicate (6 plots, stochastic arrangement; not used in 2003). The design of groups C, D and E was based on an application of compound matrices. Group C was arranged according to a Quinque-factor orthogonal design (factors: X_1 to X_5, one repeat: 36 plots, stochastic arrangement).

Group D involved Bin-factor orthogonal contract plots (factors: X_2, $X_3 + X_4$, one repeat: 22 plots, stochastic arrangement). Group E consisted of a Tri-factor orthogonal rotary design (factors: X_1, X_3 and X_6, one repeat: 23 plots, stochastic arrangement). Six additional plots to which no treatment was applied were included as controls from 2003 through 2005. Therefore, a total of 111 experimental field plots (treatments) divided into the five groups defined above, plus the control, were arranged via designs of orthogonal arrays (Hedayat et al., 1999).

Each of the individual plot areas was 28 m^2 (that is, 4 × 7 m) with 1.5 m spacing between the adjacent plots. These orthogonal experiments were conducted yearly and repeated under various field management conditions for the controlled growing environments involving X_1 to X_6.

Data collection

To avoid marginal effects from anthesis to seed harvest in the experimental years of 2003, 2004 and 2005, 1 m was left at the edge of the plots. Data on the seed-yield components and the seed yields of each plot were collected in the following manner. Ten samples of a 1 m long row were randomly selected in each plot to count the number of fertile tillers. The resulting counts were then converted [divided by 0.45 (row space)] and expressed in units of fertile tillers m^{-2} (Y_1). From each plot, 30 to 51 fertile tillers, 27 to 30 spikelets and 24 to 30 spikelets were randomly selected for measuring the values of spikelets fertile tillers^{-1} (Y_2), florets spikelet^{-1}

Table 1. The sample size of Y_1-Y_5 and Z for each field experimental plot of *Festuca arundinacea* Schreb.

Year	Sample size of plots (N)	Sample size of each plot (n)					
		Fertile tillers $m^{-2} Y_1$	Spikelets fertile tillers^{-1} Y_2	Florets spikelet^{-1} Y_3	Seed numbers spikelet^{-1} Y_4	Seed weight † Y_5	Seed yield Z
		Number (m^{-2})	Number	Number	Number	mg	kg ha^{-1}
2003	105	10	51	27	24	10	4
Total sample size (n) ‡		1050	5355	2835	2520	1050	420
2004	111	10	30	30	30	10	4
Total sample size (n)		1110	3330	3330	3330	1110	444
2005	111	10	30	30	30	10	4
Total sample size (n)		1110	3330	3330	3330	1110	444
Three years totally (n)		3270	12015	9495	9180	3270	1308

†: Y1 to Y5 and Z are stand for fertile tillers m^{-2}, spikelets fertile tillers^{-1}, florets spikelet^{-1},seed numbers spikelet^{-1}, seed weight (mg) and seed yield (kg ha-1), respectively. ‡: F-values are presented along with statistical differences; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$; §, Tthe direct effects of Y1~Y5 to Z are highlighted in bold (on main diagonal cell); arrows illustrate the direction of the effects.

(Y_3) and seed numbers spikelet^{-1} (Y_4), respectively (Table 1). When the seed heads were ripe, four samples from a 1 m long row were separately threshed by hand from each plot.

The weight of the clean seed in each sample was determined. The water content of the seed was found to be 7 to 10%. This percentage was used to obtain the value of seed yield (Z, kg hm^{-2}). Ten lots of 100 seeds were randomly sampled from each plot for determining the seed weight (Y_5, mg) from the seed yield samples. Table 1 shows the sample size that was determined for each of the individual years and then specified for the experimental databases with Visio FoxPro (Version 6.0) (Crook, 2001).

Statistics and analytical method

The study was conducted for three consecutive years (2003, 2004 and 2005) in the same location. The separate analyses and the combined analysis for the three years both provided information (Chatterjee and Price, 1977). Pearson correlation analysis (both total and individual for three years) was performed. A Qbasic program was written to conduct the path coefficient analysis. Duncan's multiple range test for Z and Y_1 to Y_5 was also applied.

Several procedures have been proposed for the selection of k in ridge regression analysis, in view of the fact that the optimal value of k cannot be determined with certainty (Hoerl et al., 1975; Marquardt and Snee, 1975; Lawless and Wang, 1976; Chatterjee and Price, 1977). Hoerl and Kennard (1970a) have suggested that k may be determined from the ridge trace, with k selected to obtain a stable set of regression coefficients (Newell and Lee, 1981).

To establish a reliable model, the combined data for all of the Z and Y_1 to Y_5 in Visio FoxPro, and a total of 327 samples of Z (111 × 3 − 6 = 327) with the corresponding components (Y_1 to Y_5) over the three years studied, were transformed using the natural logarithm. This transformation produced better statistical properties and did not influence the essential mathematical relations of the variables (Bradley et al., 1977; Lattin et al., 2003; Gao et al., 2005).

Let $S = \ln Z$ and let $C_i = \ln Y_i$ for $i = 1$ to 5. S and C_1 to C_5 were used for the ridge regression analyses (Chatterjee and Price, 1977). The ridge regression model was

$$S = C ß + u \qquad (1)$$

Where, S is an n×1 vector of observations on a response variable, C is an n×p matrix of observations on p explanatory variables, ß is the p×1 vector of regression coefficients and u is an n×1 vector of residuals satisfying E (❻) = $\dot{C}$, E (uu′) = δ^2 I. It is assumed that C and S have been scaled so that C′C and S′S are matrices of correlation coefficients. Here $n = 327$, $p = 5$.

Thus, $$\ln Z = \left(\sum_{i=1}^{5} \ln Y_i\right) ß + u \qquad (2).$$

The logarithmic model (2) above was transformed to yield the following exponential function:

$$Z = e^{\alpha} \cdot \prod_{i=1}^{5} (Y_i^{\beta}) \qquad (3)$$

Where, α and β are constants.

Formula (3) was used to estimate the Z of all 327 samples. This estimate was denoted as $Z_{estimated}$. The actual seed yields were denoted as Z_{actual}.

A general linear regression model was used to compare the Z_{actual} with the $Z_{estimated}$. An analysis of variance was used to assess the dependent variable Z_{actual} in terms of the parameter estimates of $Z_{estimated}$. The linear regression model is:

$$Z_{actual} = \beta + k \cdot Z_{estimated} \qquad (4).$$

Using formula (4), the model was adjusted to

$$Z = \beta + k \cdot e^{\alpha} \cdot \prod_{i=1}^{5} (Y_i^{\beta}) \qquad (5).$$

In addition, the ridge trace and appropriate scatter plots were graphed.

The analyses and graphical procedures specified in the foregoing were all performed using SAS Version 8.2 (SAS Institute Inc., 1988).

Table 2. The Pearson correlation coefficients of Y_1-Y_5 and Z of *Festuca arundinacea* Schreb. for the three years[††].

Seed yield component	Y_1[†]	Y_2[‡]	Y_3[§]	Y_4[¶]	Y_5[#]	Z (seed yield)
Y_1	1.0000	0.2201***	0.3067***	-0.2195***	-0.1070	0.7668***
Y_2		1.0000	0.3555***	-0.2569***	0.0070	0.4917***
Y_3			1.0000	0.2568***	0.0885	0.6023***
Y_4				1.0000	0.2826***	-0.1099*
Y_5					1.0000	0.0032

†, Fertile tillers m^{-2}; ‡, spikelets fertile $tillers^{-1}$; §, florets $spikelet^{-1}$; ¶, seed numbers $spikelet^{-1}$; # seed weight (mg); ††, F-values are presented along with statistical differences; * P <0.05, ** P <0.01, *** P <0.0001. The sample size is totaling in database of the three years, N=327.

RESULTS

Correlations among Y1 to Y5 and Z

Pearson correlation coefficients (Table 2) calculated for all three years show that the seed yield components Y_1, Y_2 and Y_3 had a significantly positive correlation ($P \leq 0.0001$) with Z. Y_4 was negatively correlated ($P \leq 0.01$) with Z (Table 2). All of the pair wise correlations of the components Y_1, Y_2 and Y_3 were significant ($P \leq 0.0001$). Y_4 was negatively correlated with Y_1 and Y_2 ($P \leq 0.0001$).

Path analyses of Y_1 to Y_5 with Z

The direct and indirect effects of Y_1 to Y_5 on the seed yield are presented in Table 3. In the individual years (2003 to 2005), four components (Y_1 and Y_3 to Y_5, but not Y_2) exhibited a significant, direct effect on Z (Table 3). Y_1 showed significant direct effects on Z in 2003; Y_1 and Y_4 showed significant direct effects on Z in 2004 when Y_1 and Y_3 to Y_5 significantly affected Z in 2005. However, path analysis showed that only Y_1 had a strong direct effect (Table 3) on Z over the entire three years ($P \leq 0.0001$). The coefficients of Y_1 were 0.4427, 0.6172 and 0.6616 in 2003, 2004 and 2005, respectively. Thus, Y_1 made the largest contribution to Z. In 2005, Y_3, Y_4 and Y_5 (0.1780 at $P \leq 0.05$, 0.2713 at $P \leq 0.01$ and 0.2152 at $P \leq 0.01$, respectively) had significant direct effects on Z. Y_4 in 2004 (0.1983 at $P \leq 0.05$) also had significant direct effects on Z.

The analysis of the contributions of components Y_1 through Y_5 to Z showed that the strongest indirect effect on Z was Y_3 via Y_4 (the coefficients are 0.1525, 0.1437 and 0.1872) and Y_4 via Y_3 (0.1808, 0.1172 and 0.1228), followed by Y_2 via Y_1 (0.0361, 0.1717 and 0.0480) and Y_5 via Y_4 (0.0594, 0.1310 and 0.0016).

The comparison of the effects of Y_1 through Y_5 on Z suggests that the rank-ordered contributions of the five components to the seed yield may be represented as $Y_1 > Y_4 > Y_3 > Y_5 > Y_2$. This order is the same as that found by considering the total of the direct effects.

Ridge regression models of Z with Y_1 to Y_5

The results of the Duncan multiple range tests conducted in SAS for Z and its components (Y_1 to Y_5) in the three years are presented in Table 4. Z and the components Y_1 to Y_4 all differed significantly in the three years of the study (P<0.0001).

Ridge regression and multiple regression analyses were applied to avoid the high intercorrelation and multicollinearity between the five seed yield components and the seed yield (Hoerl and Kennard, 1970; Hoerl et al., 1975; Chatterjee and Price, 1977).

The estimated values of the ridge coefficients were obtained using the stable k value. This value was selected by using the ridge trace method developed by Hoerl and Kennard (1970 a, b). Figure 1 show the standardised ridge traces calculated from the ridge traces for the three study years; 2003, 2004 and 2005. Using values of k from 0 to 1, the curves of Y_1 to Y_5 were stabilised and were asymptotically parallel to the horizontal axis for the values of k estimated at points 0.4, 0.3 and 0.2 in Figure 2 for the three respective years. The ridge regression models were obtained for the selected values of k for the years 2003, 2004 and 2005.

Let A, B and C denote the ridge regression models for the years 2003, 2004 and 2005, respectively.

$$A: Z = -1134.86 + 1.173Y_1 + 2.956Y_2 + 44.67Y_3 + 183.621Y_4 + 462.909Y_5 \quad (6)$$

(Ridge k = 0.4; F = 8.08 Pr<0.0001)

$$B: Z = -1692.19 + 1.916Y_1 + 8.446Y_2 + 96.098Y_3 + 143.259Y_4 + 161.143Y_5 \quad (7)$$

(Ridge k = 0.3; F = 34.59 Pr<0.0001)

$$C: Z = -1434.76 + 4.892Y_1 + 1.064Y_2 + 108.526Y_3 + 109.146Y_4 + 164.853Y_5 \quad (8)$$

(Ridge k = 0.2; F = 44.64 Pr<0.0001)

Table 3. Path analysis showing direct and indirect effects of Y_1-Y_5 on Z for *Festuca arundinacea* Schreb.‡

Parameter		**Indirect effect via§**				
	year	**$\rightarrow Y_1^{\dagger} \rightarrow Z$**	**$\rightarrow Y_2 \rightarrow Z$**	**$\rightarrow Y_3 \rightarrow Z$**	**$\rightarrow Y_4 \rightarrow Z$**	**$\rightarrow Y_5 \rightarrow Z$**
Y_1	2003	0.4427***	0.0001	-0.0970	-0.0389	-0.0111
	2004	0.6172***	0.0112	0.0361	0.0485	0.0016
	2005	0.6616***	0.0003	0.0229	0.0309	-0.0374
Y_2	2003	0.0361	0.0013	0.0220	0.0334	0.0498
	2004	0.1717	0.0404	0.0332	0.0310	0.0002
	2005	0.0480	0.0039	0.0182	0.0365	0.0172
Y_3	2003	-0.1838	0.0001	0.2336	0.1525	0.0509
	2004	0.1377	0.0083	0.1617	0.1437	0.0030
	2005	0.0850	0.0004	0.1780*	0.1872	0.0028
Y_4	2003	-0.0873	0.0002	0.1808	0.1970	0.0537
	2004	0.1510	0.0063	0.1172	0.1983*	0.0033
	2005	0.0752	0.0005	0.1228	0.2713**	0.0013
Y_5	2003	-0.0276	0.0003	0.0666	0.0594	0.1785
	2004	0.1981	0.0015	0.0968	0.1310	0.0050
	2005	-0.1151	0.0003	0.0023	0.0016	0.2152**
Total direct effect		1.7215	0.0456	0.5733	0.6666	0.4077
Total effect		2.2141	0.0751	1.1952	1.4834	0.5430

†, Y1 to Y5 and Z are stand for fertile tillers m^{-2}, spikelets fertile tillers^{-1}, florets spikelet^{-1},seed numbers spikelet^{-1}, seed weight (mg) and seed yield (kg ha^{-1}), respectively; ‡, F-values are presented along with statistical differences; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$; §, The direct effects of Y1~Y5 to Z are highlighted in bold (on main diagonal cell); arrows illustrate the direction of the effects.

Table 4. Duncan's multiple range test for seed yield (Z) and its components (Y_1-Y_5) of *Festuca arundinacea* Schreb. for three years.

Year	**N**	**$Y_1^{\dagger}$ (number m^{-2})**	**Y_2 number**	**Y_3 number**	**Y_4 number**	**Y_5 (mg)**	**Z (kg ha^{-1})**
2003	105	$402.66^{a‡}$	60.83^{a}	7.2950^{a}	4.1895^{a}	3.0485^{a}	2023.75^{a}
2004	111	271.53^{b}	44.99^{c}	6.0592^{b}	4.7463^{b}	3.0944^{ab}	972.11^{b}
2005	111	102.11c	54.88^{b}	5.7655^{c}	4.6071^{c}	3.1360^{b}	685.53^{c}
F value		333.92	407.08	74.48	36.53	5.10	330.93
Pr > F		<.0001	<.0001	<.0001	<.0001	0.0066	<.0001

†, Y_1 to Y_5 and Z are stand for fertile tillers m^{-2}, spikelets fertile tillers^{-1}, florets spikelet^{-1},seed numbers spikelet^{-1}, seed weight (mg) and seed yield (kg ha^{-1}), respectively. ‡: Means with the same letter are not significantly different at Alpha= 0.05.

All of the ridge coefficients were positive and the coefficient of same component is in same quantitative ranks, whereas their values varied during the three years examined. The highest ridge regression coefficients were Y_4 and Y_5 in 2003, Y_2 in 2004, and Y_1 and Y_3 in 2005.

The steady algorithmic model of Z with Y_1 to Y_5

In order to get a more general model to express the relationship between Y_1 to Y_5 and Z, we combined the three years' data. The 327 samples of Z with Y_1 to Y_5 in the database over the three years studied were transformed using the natural logarithm.

Let $S = \ln Z$, $C_1 = \ln Y_1$, $C_2 = \ln Y_2$, $C_3 = \ln Y_3$, $C_4 = \ln Y_4$ and $C_5 = \ln Y_5$.

S and C_1 to C_5 were used in the ridge regression analyses. The resulting ridge regression model was

$$S = -1.604 + 0.4227 \cdot C_1 + 0.9788 \cdot C_2 + 0.8909 \cdot C_3 + 0.0742 \cdot C_4 + 0.5924 \cdot C_5 \quad (9)$$

(N = 327, F = 237.55, Pr<.0001. The variance analysis and the parameter estimates are given in Tables 5 and 6). In terms of the original variables, $\ln Z = -1.604 + 0.4227 \cdot \ln Y_1 + 0.9788 \cdot \ln Y_2 + 0.8909 \cdot \ln Y_3 + 0.0742 \cdot \ln Y_4 + 0.5924 \cdot \ln Y_5$. The above logarithmic model was transformed to

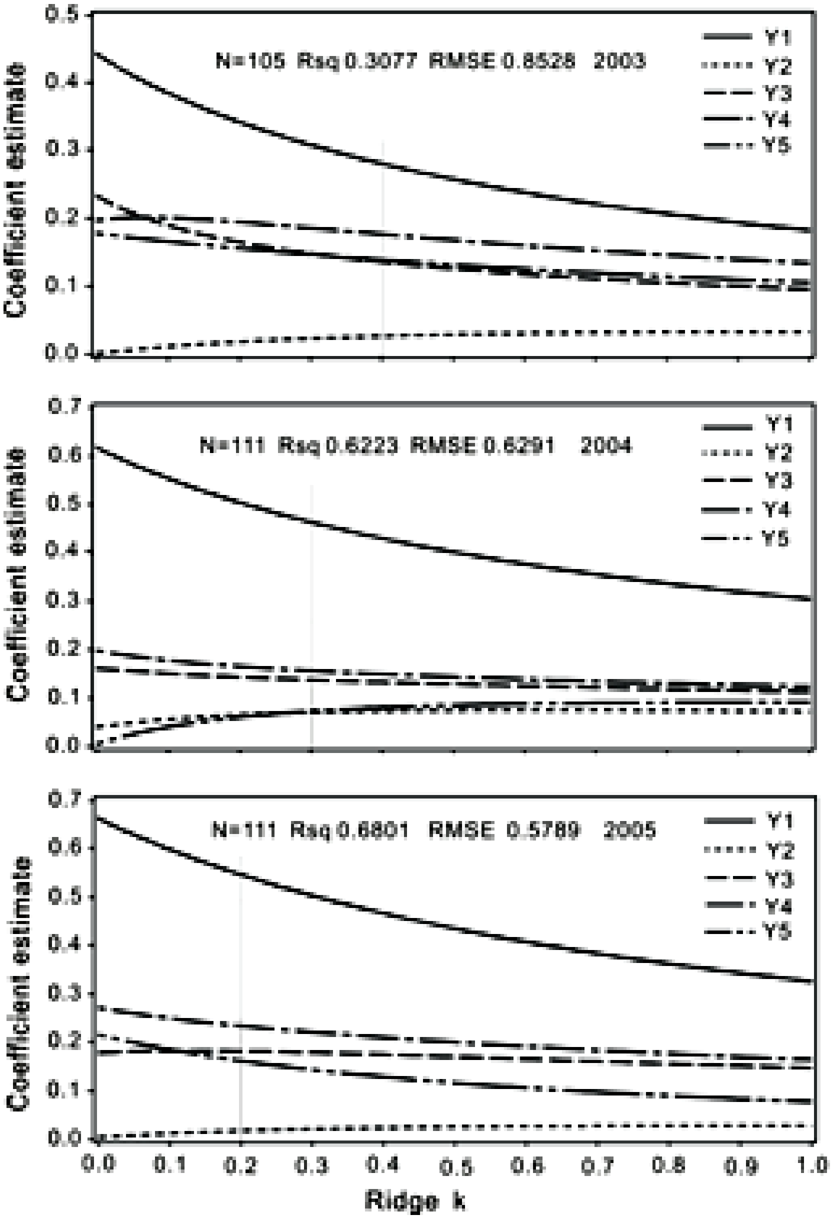

Figure 1. Ridge traces of the standard partial regression coefficients for the increasing values of k for the five yield components of tall fescue in Jiuquan, Gansu province, China, for the years 2003, 2004 and 2005. Y_1 to Y_5 and Z denote fertile tillers m^{-2}, spikelets fertile tillers^{-1}, florets spikelet^{-1}, seed numbers spikelet^{-1}, seed weight (mg) and seed yield (kg ha^{-1}), respectively.

an exponential function as follows:

$$Z = e^{-1.6}\cdot Y_1^{0.42}\cdot Y_2^{0.98}\cdot Y_3^{0.89}\cdot Y_4^{0.07}\cdot Y_5^{0.59} \quad (10).$$

Formula (10) was used to estimate the seed yield of all the 327 samples. These estimates were denoted by $Z_{estimated}$. The observed seed yields were denoted by Z_{actual}.

A general linear regression model was used to compare the values of Z_{actual} with the values of $Z_{estimated}$. An analysis of variance was used to assess the dependent variable Z_{actual} and the parameter estimates of $Z_{estimated}$ (Tables 7 and 8). The linear regression model is graphed in Figure 2.

The regression model obtained in this analysis is as follows:

$$Z_{actual} = -106.49 + 1.17\cdot Z_{estimated} \text{ (N = 327, F = 1036.95, Pr<.0001)} \quad (11).$$

Using formula (11), the model was adjusted to:

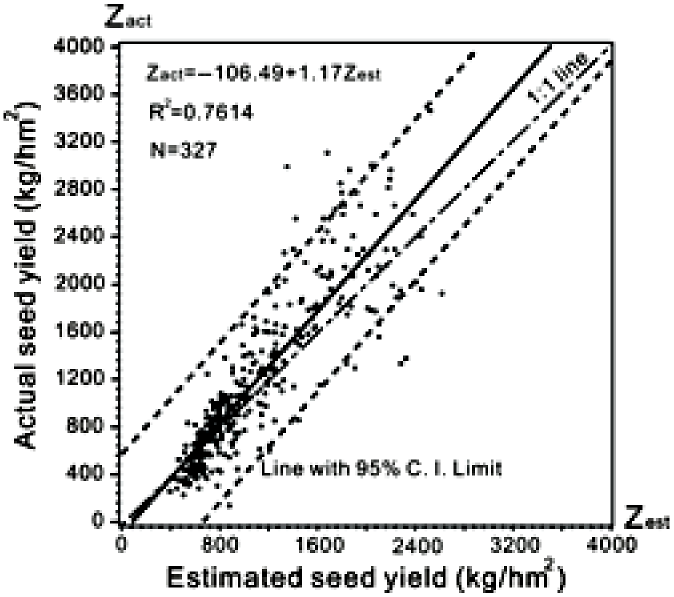

Figure 2. Scatterplot used to fit the regression of the actual seed yield on the estimated seed yield for the combined three years. Z_{est} was estimated by the model $Z=e^{-1.6}Y_1^{0.42}Y_2^{0.98}Y_3^{0.89}Y_4^{0.07}Y_5^{0.59}$ for tall fescue.

Table 5. Analysis of variance for the dependent variable of Z_{actual} with the five seed yield components of a total of 327 samples.

Source	DF	Sum of squares	Mean square	F value	Pr > F
Model	5	96.4791	19.2952	237.55	<.0001
Error	321	26.0740	0.0812		
Corrected total	326	122.5501			

Table 6. The parameter estimates of the five seed yield components of a total of 327 samples.

Variable	DF	Parameter estimate	Standard error	t	Value
Intercept	1	-1.6040	0.6093	-7.02	<.0001
y1	1	0.4227	0.0249	22.94	<.0001
y2	1	0.9788	0.1276	10.55	<.0001
y3	1	0.8909	0.1286	5.56	<.0001
y4	1	0.0742	0.1627	2.38	0.0178
y5	1	0.5924	0.2459	3.45	0.0006

$$Z=-106.95+1.17\cdot e^{-1.6}\cdot Y_1^{0.42}\cdot Y_2^{0.98}\cdot Y_3^{0.89}\cdot Y_4^{0.07}\cdot Y_5^{0.59}$$
$$=-106.95+0.24\cdot Y_1^{0.42}\cdot Y_2^{0.98}\cdot Y_3^{0.89}\cdot Y_4^{0.07}\cdot Y_5^{0.59} \quad (12).$$

According to the variance test, the parameter estimates of the intercept and $Z_{estimated}$ were 0.0019 and 1.0000, respectively (Table 9). The regression line (Figure 3) was very close to the 1:1 line.

DISCUSSION

The results of the analysis failed to confirm the first hypothesis. This hypothesis stated that all of the five seed-yield components and the seed yield were inter-correlated and that all of the five seed-yield components contributed positively to seed yield.

Table 7. Analysis of variance for the dependent variable Z_{actual} with the estimated seed yield.

Source	DF	Sum of squares	Mean square	F value	Pr > F
Model	1	120601166	120601166	1036.95	<.0001
Error	325	37798702	116304		
Corrected total	326	158399868			

Table 8. The parameter estimates of $Z_{estimated}$.

Variable	DF	Parameter estimate	Standard error	t value	Pr > |t|
Intercept	1	-106.49	39.1487	2.74	0.0065
$Z_{estimated}$	1	1.1722	0.0291	32.21	<.0001

Table 9. The parameter estimates of $Z_{estimated}$ after adjustment by the linear regression.

Variable	DF	Parameter estimate	Standard error	t value	Pr > |t|
Intercept	1	-0.0019	42.1125	-0.00	1.0000
$Z_{estimated}$	1	1.0000	0.03105	32.20	<.0001

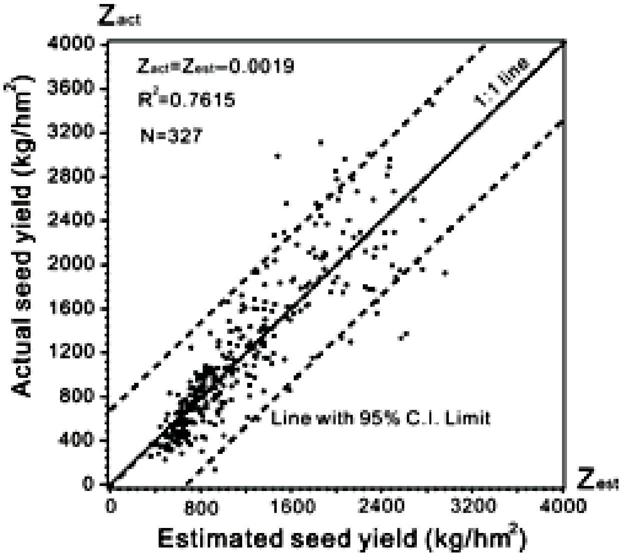

Figure 3. Scatterplot used to fit the regression of the actual seed yield of tall fescue on the estimated seed yield. The regression was adjusted by $Z_{act} = -106.49+1.17\cdot Z_{est}$ for the 3 years. This line nearly coincides with the 1:1 line.

Conversely, the results of the analysis supported the second hypothesis, which states that a steady algorithm model could estimate seed yield by using the values of the components.

Y_1 to Y_5 contribute to Z

The results of this study indicate that the total direct effects of Y1, Y3, Y4 and Y5 contributed significantly and positively to Z (Table 3). This finding is consistent with reports in the literature. For example, the component most associated with yield plant-1 was the number of mature seeds panicle-1 in Panicum coloratum L. (Barrios et al., 2010). In this study, Y2 did not contribute significantly to Z. However, the exponent of Y2 (0.98) is the largest value appearing in the algorithm model. It can generic confirm that Y2 is primarily under genetic control in grasses (Fairey and Hampton, 1997; Hampton and Fairey, 1998; Boelt and Gislum, 2010). This finding implies that Y2 is the component that should first be considered if high seed yield in grasses is the goal of the breeding program.

Nevertheless, Y1 was the most important and effective component associated with Z, as shown by its significantly ($P<0.0001$) highest contribution in this trial (path coefficients: 0.4427, 0.6172 and 0.6616 in 2003, 2004 and 2005, respectively). This finding agrees with previous reports in grasses, for example, in fescues (Young et al., 1998 b; Wang, 2005; Boelt and Gislum, 2010), Russian wild rye (Sun et al., 2005; Wang et al., 2010), zoysiagrass (Ma et al., 2004), bermudagrass (Wu et al., 2008), crested wheatgrass (Jafari et al., 2007; Taghizadeh et al., 2008), perennial ryegrass(Deleuran et al., 2009) and other grasses (Canode, 1980; Hampton and Fairey, 1998).

A steady algorithmic model describing Z in terms of Y_1 to Y_5

We developed an original exponential model for estimating Z from the values of Y1 through Y5. To our knowledge, this model is the first of its kind to be developed for this purpose. The model was statistically reliable. Its performance was verified by the fact that the regression line fit to the adjusted scatter plot of the actual and estimated Z (Table 9); almost coincided with the 1:1 line (Chatterjee and Price, 1977; Lattin et al., 2003) (Figure 3). First, the final algorithm model [exponential equation (12)] was deduced from the data of 327 samples from various growth regimes in three successive years. Moreover, all three ridge regression models (varying coefficients will be interpreted in this paper) for the individual years were significant ($P<0.0001$) and all had positive coefficients matching with the contributions of the five Ys to the Z. This result can tentatively explain the good correspondence found between the path coefficient analyses and the ridge regressions. In addition, ridge regression effectively avoided the problems potentially caused by the high intercorrelations of the predictor variables Y1 to Y5 (Hoerl and Kennard, 1970a; Hoerl and Kennard, 1970b; Chatterjee and Price, 1977).

However, it is interesting that the exponents of the model decreased in the rank order Y_2 (0.98) > Y_3 (0.89) > Y_5 (0.59) Y_1 (0.42) > Y_4 (0.07), whereas the contributions of the same variables exhibited the rank order Y_1 (2.22) > Y_4 (1.48) > Y_3 (1.20) > Y_5 (0.54) > Y_2 (0.08) (Table 3). Although, these two sets of calculated values were computed from the same database, the ridge analysis values analytically combined the effects of all the Ys, especially the effects of aging and climate, to address the variation in Z for the three years, whereas the path analysis included the separate analytic effects of the individual three years. The former analysis is mathematically more generic than the latter (Lawless and Wang, 1976; Chatterjee and Price, 1977; Gregory, 1978; Lattin et al., 2003). Obviously, in the present trial, the genetic controls were more generic than the environmental controls for Y_1 to Y_5. Therefore, we tentatively propose that Y_2, Y_3 and Y_5 were orderly more genetic and less environmental control than Y_1 and Y_4 and vice versa. These considerations might suggest that improvement of Y_1 and Y_4 should be the primary focus of breeding programs aimed at improving the seed production of tall fescue. This suggestion is consistent with previous literature on the topic (Young et al., 1989c).

The intercorrelation among Y_1 to Y_5 and Z

In a study conducted in Corvallis, Oregon (United States), Young (1998c) found that the Zs of all four experimental tall fescue cultivars tested (including Fawn), were closely correlated with $Y_1 \times Y_2 \times Y_4$ .We found that Z was significantly positively correlated both with Y_1 and Y_2 but negatively correlated with Y_4. Neither Y_1, Y_2 nor Y_3 were significantly correlated with Y_5 (Table 2), but it was negatively correlated with Y_4 over the entire three years (Table 2). Variation may be the reason for this apparent discrepancy (Jafari et al., 2006). Our result in this experiment appears to be in theoretical accordance with current biological theory. Except for the correlation of Y_3 with Y_4 and Y_1 and the correlation of Y_3 and Y_4 with Z, the significant correlations were variable. This result was probably a consequence of the effects of aging of the plant and the climate of the individual year. The management regimes were repeated each year in the experiment and therefore would not have produced this result (Fairey and Hampton, 1997; Hampton and Fairey, 1998). The results of this study further emphasises that as the plants aged during the successive experimental years, Y_1, Y_2 and Y_3 decreased significantly, whereas Y_4 and Y_5 increased. This finding agrees with the results of previous research (Fairey and Lefkovitch, 1999). This result also implies that Y_4 and Y_5 should and could be effectively improved if the values of Y_1, Y_2 and Y_3 are lower than normal. The justification for this argument is that Y_1 through Y_5 represented successive phenological periods in the production cycle of the grass seed.

Significantly varying coefficients of ridge regressions

The coefficients of the ridge regression models for the individual years were variable and ranged from 1.064 to 462.909. The main apparent causes of this variation were co-effects of the aging of the plants, variable climatic conditions and variation among the designs for the experimental management of the fields. These causes added to the effects of high intercorrelation among the components and led to multicollinearity in the regression analysis that linked Y_1 to Y_5 with Z. For this very reason, the data from all three years were summed, log-transformed and subjected to ridge regression analysis to reveal the essential algorithmic relations underlying the data (Hoerl and Kennard, 1970; Hoerl et al., 1975; Bradley et al., 1977; Chatterjee and Price, 1977; Lattin et al., 2003; Gao et al., 2005).

Conclusions

Ridge regression analysis of a large sample produced by an orthogonal experimental design yielded the following algorithmic model:

$$Z = -106.49 + 0.24 \cdot Y_1^{0.42} \cdot Y_2^{0.98} \cdot Y_3^{0.89} \cdot Y_4^{0.07} \cdot Y_5^{0.59}.$$

The study found that Z can be accurately estimated from Y_1, Y_2, Y_3, Y_4 and Y_5. The combined direct effects of Y_1, Y_3, Y_4 and Y_5 with regard to Z were positive. Y_2 represented an exception to this pattern of positive relationships. Of the components examined, Y_1 exhibited the largest contribution to Z. In rank order, the contributions of the five key components to Z were as follows: $Y_1 > Y_4 > Y_3 > Y_5 > Y_2$. The components Y_1, Y_2 and Y_3 were positively correlated with Z, whereas Y_4 exhibited a weakly negative correlation. The intercorrelations of the components Y_1, Y_2 and Y_3 were significant. Y_1, the major component, exhibited the most important and substantial effect of any of the five components on grass seed production. However, in view of the values of the exponents of the algorithmic model, it appears that selection for high seed yield through direct selection for large Y_2, Y_3 and Y_5 values would be more effective than selection on Y_4 and Y_1 in a breeding program involving this grass.

Future studies may consider the climate (such as rainfall and temperature) in the seed production stage and different site locations to facilitate the determination and testing of models of seed yield as a function of seed yield components in grasses.

ACKNOWLEDGEMENTS

The National Science and Technology Pillar Program of China (2011BAD17B05), The National Basic Research and Development Program (973 project, 2007CB106805) and 948 Research Project (No.202099) Ministry of Agriculture of P R China, funded this work. We are grateful to Dr Luo Shuhang, Dr Liu Fuyuan, and Dr Zhongyong, presidents of the Daye International Interest Co. Ltd., and to our skilled technical assistants, Zhang Bing, Yan Xuehua, Han Juhoung, Zhang Xijun, Wang Shouguo and Zhang Guoqi, animal husbandry engineers of the Daye Institute of Forage and Grass Products in Jiuquan, and Gansu Branch of Chengdu Daye International Interest Co. Ltd. for their assistance in this research.

REFERENCES

Akinyele BO, Osekita OS (2006). Correlation and path coefficient analyses of seed yield attributes in okra (*Abelmoschus esculentus* (L.) Moench). Afr. J. Biotechnol. 5: 1330-1336.

Barrios C, Armando L, Berone G, Tomas A (2010). Seed yield components and yield per plant in populations of *Panicum coloratum* L. var. makarikariensis Goossens, Proc. Int. Herbage Seed Conf. Dallas TX. p. 7.

Belesky DP, Ruckle JM, Halvorson JJ (2010). Carbon isotope discrimination as an index of tall fescue endophyte association response to light availability and defoliation. DOI: 10.1016/j.envexpbot.2009.09.009. Environ. Exp. Bot. 67: 515-521.

Bicer BT (2009). The effect of seed size on yield and yield components of chickpea and lentil. Afr. J. Biotechn. 8: 1482-1487.

Bliss FA, Barker LN, Hall TC (1973). Genetic and environmental variation of seed yield, yield components, and seed protein quantity and quality of cowpea. Crop Sci. 13: 656-660.

Boelt B, Gislum R (2010). Seed yield components and their potential interaction in grasses to what extend does seed weight influence yield?, Proc. Int. Herbage Seed Conf. Dallas, TX. 7: 109-112.

Bradley SP, Hax AC, Magnanti TL (1977). Applied mathematical programming. Addison- Wesley Pub. Company Inc. CA.

Canode CL (1980). Grass seed production in the intermountain Pacific Northwest, USA. p. in P.D. Hebblethwaite. Seed production. Proc. Easter School Agric. Sci. Univ. Nottingham. Butterworths, London, 28: 189-202.

Ceyhan E, Avci MA, Karadas S (2008). Line X tester analysis in pea (Pisum sativum L.): Identification of superior parents for seed yield and its components. Afr. J. Biotechnol. 7: 2810-2817.

Chatterjee S, Price B (1977). Regression analysis by example. John Wiley & Sons, Inc, New York.

Cheng Z, Salminen SO, Grewal PS (2010). Effect of organic fertilisers on the greening quality, shoot and root growth, and shoot nutrient and alkaloid contents of turf-type endophytic tall fescue, Festuca arundinacea. Annal. Appl. Biolog. 156: 25-37.

Crook S (2001). Visual FoxPro Client-Server Handbook. Redware Research Ltd., Hove, England.

Deleuran LC, Gislum R, Boelt B (2009). Cultivar and row distance interactions in perennial ryegrass. Acta Agriculturae Scandinavica Section DOI: 10.1080/09064710802176642. B-Soil. Plant Sci. 59: 335-341.

Fairey DT, Hampton JG (1997). Forage seed production. CAB Internatonal. Wallingford, Oxon. OX10 8DE, UK.

Gao S, Li Y, Jin H (2005). Application of ridge regression models in economic increasing factors analysis. Statist. Decision Making, 5: 142-144.

Gregory S (1978). Statistical methods and the geographer. Fourth Edition ed. Longman Group Limited London. Printed in Great Britain by Richard Clay (The Chaucer Press) Ltd., London.

Hampton JG, Fairey DT (1998). Components of seed yield in grasses and legumes. Forage seed production, Temperate species. 1: 45-69.

Hedayat AS, Sloane NJA, Stufken J (1999). Orthogonal arrays: Theory and applications. Published by Springer-Verlag New York.

Hoerl AE, Kennard RW (1970a). Ridge regression: biased estimation for non orthogonal problem. Technometrics. 12: 55-67.

Hoerl AE, Kennard RW (1970b). Ridge regression: Applications to non orthogonal problems. Technometrics, 12: 69-82.

Hoerl AE, Kennard RW, Kent FB (1975). Ridge regression: some simulations. Commun. stat. 4: 105-123.

Jafari AA, Setavarz H, Alizadeh MA (2006). Genetic variation for and correlations among seed yield and seed components in tall fescue. J. New Seeds, 8(4): 47-65

Jafari AA, Seyedmohammadi AR, Abdi N (2007). Study of variation for seed yield and seed components in 31 genotypes of *Agropyron desertorum* through factor analysis. Iranian J. Rangelands and Forests Plant Breed. Gene Res. 15(21): 1220- 1221.

Karasu A, Oz M, Goksoy AT, Turan ZM (2009). Genotype by environment interactions, stability, and heritability of seed yield and certain agronomical traits in soybean [*Glycine max* (L.) Merr.]. Afr. J. Biotechnol. 8: 580-590.

Kaya M, Sanli A, Tonguc M (2010) Effect of sowing dates and seed treatments on yield, some yield parameters and protein content of chickpea (*Cicer arietinum* L.). Afr. J. Biotechnol. 9: 3833-3839.

Kokten K, Toklu F, Atis I, Hatipoglu R (2009) Effects of seeding rate on forage yield and quality of vetch (*Vicia sativa* L.) triticale (*Triticosecale Wittm*.) mixtures under east mediterranean rainfed conditions. Afr. J. Biotechnol. 8: 5367-5372.

Lattin JM, Carroll JD, Green PE (2003). Analyzing multivariate data. Brooks/Cole, an imprint of Thomson Learning, Pacific Grove, CA: Duxbury.

Lawless JF, Wang P (1976). A simulation study of ridge and other regression estimators. Commun. Stat. Ser. A. 5: 307-323.

Ma C, Han J, Sun J, Zhang Q,Lu G (2004). Effects of nitrogen fertilizer on seed yields and yield components of Zoysia japonica established by seeding and transplant. Agric. Sci. China. 3: 553-560.

Marquardt D, Snee R (1975). Ridge regression in practics. Am. Statist. 29: 3-14.

Mensah JK, Obadoni BO, Akomeah PA, Ikhajiagbe B, Ajibolu J (2007). The effects of sodium azide and colchicine treatments on morphological and yield traits of sesame seed (*Sesame indicum* L.). Afr. J. Biotechnol. 6: 534-538

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

Qu RL, Li D, Du R, Qu R (2003). Lead uptake by roots of four turfgrass species in hydroponic cultures. Hort. Sci. 38: 623-626.

SAS-Institute-Inc. (1988). SAS/STAT User's Guide. SAS Institute Inc. Cary, NC.

Soil-Survey-Staff. (1996). Keys to soil taxonomy (7th. ed.). Natural Resources Conservation Service, Washington, DC.

Sun T, Han J, Zhao S, Yue W (2005). Effects of fertilizer application on seed yield and yield components of *Psathyrostachys juncea*. Grassland of China. 27: 16-21.

Taghizadeh R, Jafari A, Choukan R, Asghari A (2008). Evaluation for seed yield and seed components among accessions of crested wheatgrass (*Agropyron desertorum* and *Agropyron crustatum*). Multifunctional grasslands in a changing world, Volume II: XXI Int. Grassland Congr. and VIII Int. Rangeland Congr., Hohhot, China, 29 June-5 July. p. 361.

Wang Q (2005). Coupling effects of water and fertilizer on seed yield and the yield performance of six grasses species, Ph.D. diss. China Agricult. Univ. Beijing.

Wang Q, Cui J, Zhou H, Wang X, Zhang T, Han J (2010). Path coefficient and ridge regression analysis to improve seed yield of Psathyrostachys juncea Nevski, Proc. 7th Int. Herbage Seed Conf. Dallas TX. USA.

Wang Q, Zhang T, Cui J, Wang X, Zhou H, Han J, Gislum R (2011). Path and Ridge Regression Analysis of Seed Yield and Seed Yield Components of Russian Wildrye (*Psathyrostachys juncea* Nevski) under Field Conditions. PLoS One. 6: 18-245.

Wu YQ, Taliaferro CM, Martin DL, Anderson JA, Anderson MP (2008). Correlation analyses of seed yield and its components in bermudagrass. Multifunctional grasslands in a changing world, Volume II: XXI Int. Grassland Congr. and VIII Int. Rangeland Congr., Hohhot, China, 29 June-5 July. p. 562.

Yandell BS (1997). Practical data analysis for designed experiments. Chapman & Hall, London.

Young WC, Youngberg HW, Silberstein TB (1998a). Management studies on seed production of turf type tall fescue: I. Seed yield. Agron. J. 90: 474-477.

Young WC, Youngberg HW, Silberstein TB (1998b). Management studies on seed production of turf type tall fescue: II. Seed yield components. Agron. J. 90: 478-483.