OPEN

Seed yield can be explained by altered yield components in field-grown western wheatgrass (*Pascopyrum smithii Rydb*.)

Zhao Chen[1], Junpeng Niu[1], Xinlong Cao[1], Wenbo Jiang[1], Jian Cui[2], Quanzhen Wang[1]* & Quan Zhang[3]

Western wheatgrass (*Pascopyrum smithii* Rydb.) is an important cool-season forage and turfgrass. However, due to seed dormancy and poor seedling vigor, it is difficult to develop high seed yield production systems, and assessing these components in response to seed yield. Based on multifactor orthogonally designed field experimental plots under various field management regimes, the effects of numbers of fertile tillers m^{-2} (Y_1), spikelets/fertile tiller (Y_2), florets/spikelet (Y_3), seed numbers/ spikelet (Y_4), and seed weight (Y_5) on seed yield (Z) were determined over three successive years. Correlation analysis indicated that fertile tillers (Y_1) was the most important seed yield component. And the biggest contribution of those five yield component is fertile tillers (Y_1), followed by seed numbers/ spikelet (Y_4), spikelets/fertile tiller (Y_2), florets/spikelet (Y_3) and seed weight (Y_5), respectively. By using ridge regression analysis, we have developed an accurate model of seed yield with its five components. Finally, the results of synergism and antagonism among these yield components on seed yield showed that fertile tillers and seed numbers/spikelet had an antagonistic effect on seed yield. Therefore, selection for high seed yield by direct selection for large values of fertile tillers and seed numbers/ spikelet would be the most effective breeding strategy for western wheatgrass.

Western wheatgrass (*Pascopyrum smithii* Rydb.) is a native, perennial cool-season grass, found most abundantly in the southern, mixed-grass prairie region of the Great Plains of North America and is grown for livestock production throughout the temperate regions of the world[1]. Because it thrives on impoverished soils in pastoral environments, even with multiple, simultaneous stressors[2], it is an important species for soil protection, water conservation, and vegetation protection in arid and semi-arid regions. Due to these characteristics, it is a competitive, high-yielding species, providing forage for livestock and wildlife on semi-arid rangelands in Eurasia and northwest China[3–6]. Thus, it can appear as a dense monoculture[7,8]. Seed yields in the cool-season perennial grasses are often low, and in China, the seed supply of perennial grasses has depended on imports for many years due to inadequate supplies of locally produced high quality seed. To become more self-sufficient in supplying seed for its many needs, the Chinese government has encouraged the development of increased seed production capacity[9,10].

There have been many studies on the factors affecting crop yields, with the aim of improving yields as much as possible[11]. Seed yield is a complex trait that is the culmination of growth and developmental processes that are influenced by multiple yield components[12,13]. Understanding the relationship between yield and yield components and the correlations between yield components is a prerequisite for building an efficient breeding program[14]. To date, no research has been conducted to develop models for seed yield and related yield components in western wheatgrass. Therefore, it is necessary to examine the relationships among various factors, especially seed yield, yield components, and their interdependence. To create a high seed yield breeding program, a mathematical model needs to be developed that accurately predicts forage and seed yield, and this must subsequently be validated through field experiments[15]. Owing to the lower complexity and lower environmental influence of

[1]College of Grassland Agriculture, Northwest A&F University, Yangling, 712100, Shaanxi Province, China. [2]College of Life Science, Northwest A&F University, Yangling, 712100, Shaanxi Province, China. [3]Jiuquan Daye Seed Industry Co. Ltd., Jiefang Road, #325, Suzhou Qu, 735000, Jiuquan, Gansu Province, China. *email: wangquanzhen191@163.com

1

Seed yield components	Y_1	Y_2	Y_3	Y_4	Y_5	Z(seed yield)
Y_1	1.0000	0.118**	−0.056**	−0.071**	−0.286**	0.472**
Y_2		1.0000	−0.055**	−0.051**	−0.104**	0.116**
Y_3			1.0000	0.105**	0.155**	−0.023
Y_4				1.0000	0.064**	−0.030*
Y_5					1.0000	−0.192**

Table 1. Matrix of Pearson correlation coefficients of $Y_1 \sim Y_5$, Z (*Pascopyrum smithii* Schreb.) averaged over 3 years. *Significant at the 0.05 probability level. **Significant at the 0.01 probability level. N = 380.

yield components compared to yield, the use of these traits to practice indirect selection for yield is justified[16]. Several research groups have attempted to determine the association between the characters of forage yield. To obtain a high-yield forage species, several scholars have focused their attention on the agronomic traits of yield groups, which are closely related to seed yield. Annapurna *et al.*[17] found that seed yield showed a significant positive correlation with plant height, ear diameter, number of seeds per row, and number of rows per ear. Assefa *et al.*[18] reported correlations between yield and planting density, and other yield components. Tang *et al.*[19] found that grain yield per plant showed a highly significant positive correlation with the 1000-grain weight, plant height and planting density. Traits related to the generative parts of the plant, such as pods per plant, number of seeds per pod, number of fertile spikelets per panicle, panicle length, spikelet density, number of filled seeds, number of effective tillers per plant, and 1000 seed weight, are the most frequently considered parameters[20–25]. Thus, plant morphological traits such as fertile tillers m^{-2} (Y_1), spikelets/fertile tillers (Y_2), florets/spikelet (Y_3), seed numbers/spikelet (Y_4), or seed weight (Y_5), can be valuable in defining the best criteria for selection in biological and agronomic studies.

Since seed yield is influenced by environmental conditions, and agronomic factors[26–28], the experimenter selects the best design based on the information available with respect to field spatial heterogeneity[29]. Hexi Corridor, in Gansu Province, China, is considered a key seed production centre due to plentiful mountain run-off water, groundwater and dry, warm summer conditions. It is the largest maize seed production area and one of the main vegetable and flower seed production regions in China. Thus, we chose this area to evaluate the potential seed production of cool-season grasses. The orthogonal experimental design (OED) method allows researchers to test the effectiveness of many interventions simultaneously in a single experiment (and identify their interactions) with far fewer experimental units than it would take to exhaust all possible intervention combinations using other techniques. Therefore, since our objective was to investigate relationships among seed yield traits and develop a model of seed yield and yield components, we chose the OED. The aim of this study was to confirm the direct and indirect effects of key seed yield components, including fertile tillers m^{-2} (Y_1), spikelets/fertile tillers (Y_2), florets/spikelet (Y_3), seed number/spikelet (Y_4), and seed weight (mg) (Y_5), on seed yield (Z) of *Pascopyrum smithii* based on a multifactor orthogonal design under field conditions, with various managements.

Formulae representing the theoretical relationship between the seed yield components and seed yield (or potential seed yield) are represented as: $Z_{SY} = Y_1 \times Y_2 \times Y_3 \times Y_4 \times Y_5$. This study evaluated two hypotheses: (1) all five seed yield components and the seed yield are inter-related, and each yield factor has different direct and indirect effects on seed yield, and (2) an algorithmic model of seed yield based on these five components can be developed to accurately estimate seed yield.

Results

Correlations between traits.
Pearson correlation analysis revealed that seed yield (Z) was significantly and positively influenced ($P < 0.001$) by seed yield components Y_1 and Y_2, but was negatively affected to a lesser extent by Y_4 and Y_5 ($P < 0.05$ or $P < 0.01$) (Table 1). Y_1 had the maximum coefficient on Z (0.472, $P < 0.001$). The correlations among Y_1 through Y_5 showed significance ($P < 0.001$), while Y_1 and Y_2 were negatively correlated with Y_3 and Y_5. However, Y_1 and Y_2 showed strong, positive correlations with Z in for all three years ($P < 0.001$) (Table 2). The order of the strength of correlation coefficients in each year was 2004 < 2003 < 2005. Y_1 had the strongest positive influence on Z (0.452, 0.534, and 0.657 for 2003, 2004 and 2005, respectively), and Y_5 had the strongest negative influence on Z (−0.219, −0.209 and −0.354 for 2003, 2004 and 2005, respectively, Table 2).

Path analyses of Y_1 to Y_5 with Z.
The results of the path analyses showed that all five seed yield components presented a strong, highly significant direct effect on Z in at least two of the years (Table 3). However, the direct effect of Y_1 on Z was strong and positive (highlighted in bold) for all 3 years (2003, 2004 and 2005 were at $P < 0.0001$), where the coefficients were 0.480, 0.423 and 0.777, respectively. Therefore, Y_1 had the largest contribution to Z. Y_5 in 2004 (−0.065 at $P < 0.001$), Y_3 in 2003 (0.454 at $P < 0.05$) and Y_4 in 2003 (−0.371 at $P < 0.05$) had weak but statistically significant direct effects on Z.

Regarding the contributions of Y_1–Y_5 to Z, the strongest positive indirect influence was presented by the pathway from Y_1 via Y_4 (the coefficient was 0.1740 in 2003), and the second indirect affect was demonstrated by Y_2 through Y_4 (0.1039 in 2003). In order of decreasing magnitude, Y_3 via Y_4 (0.0582 in 2005) and Y3 via Y5 (0.0447 in 2003) were observed. The strongest negative indirect influence on Z was Y_2 via Y_3 (−0.0799 in 2003), and the second indirect affect was demonstrated by Y_1 through Y_3 (−0.0517 in 2005). Y_3 had a negative and marginal direct influence on yield in 2004 and positive effects during the other 2 years on seed yield (Table 3), Y_3 had the smallest contribution to Z (Table 3). In summary, the order of contributions of the five seed yield components were $Y_1 > Y_3 > Y_2 > Y_5 > Y_4$, and the total direct effects were 1.68, 0.529, 0.195, 0.077 and −0.266, respectively

	year	Y_1	Y_2	Y_3	Y_4	Y_5	Z
Y_1	2003	1.0000	0.435**	−0.297**	−0.469**	−0.434**	0.534**
	2004	1.0000	0.127**	−0.171**	−0.108**	−0.327**	0.452**
	2005	1.0000	0.558**	−0.638**	−0.391**	−0.637**	0.657**
Y_2	2003		1.0000	−0.176	−0.280**	−0.341**	0.302**
	2004		1.0000	−0.045**	−0.047**	−0.099**	0.122**
	2005		1.0000	−0.453**	−0.294**	−0.422**	0.377**
Y_3	2003			1.0000	0.787**	0.113	0.010
	2004			1.0000	0.071**	0.127**	−0.089**
	2005			1.0000	0.582**	0.514**	−0.332**
Y_4	2003				1.0000	0.234*	−0.250*
	2004				1.0000	0.053**	−0.047**
	2005				1.0000	0.322**	−0.142
Y_5	2003					1.0000	−0.219*
	2004					1.0000	−0.209**
	2005					1.0000	−0.354**

Table 2. Matrix of Pearson correlation coefficients of $Y_1 \sim Y_5$, Z (*Pascopyrum smithii* Schreb.) for each year. *Significant at the 0.05 probability level. **Significant at the 0.01 probability level. N = 105, 129 and 146 for year 2003 to 2005, respectively.

		Indirect effect via				
	year	$\rightarrow Y_1 \rightarrow Z$	$\rightarrow Y_2 \rightarrow Z$	$\rightarrow Y_3 \rightarrow Z$	$\rightarrow Y_4 \rightarrow Z$	$\rightarrow Y_5 \rightarrow Z$
Y_1	2003	**0.480***	0.0383	−0.1348	0.1740	−0.0239
	2004	**0.423***	0.0077	0.0010	−0.0005	0.0213
	2005	**0.777***	0.0257	−0.0517	−0.0391	−0.0554
Y_2	2003		**0.088**	−0.0799	0.1039	−0.0188
	2004		**0.061***	0.0003	−0.0002	0.0064
	2005		**0.046**	−0.0367	−0.0294	−0.0367
Y_3	2003			**0.454**	−0.2920	0.0062
	2004			**−0.006**	0.0004	−0.0083
	2005			**0.081**	0.0582	0.0447
Y_4	2003				**−0.371***	0.0129
	2004				**0.005**	−0.0034
	2005				**0.100**	0.0280
Y_5	2003					**0.055**
	2004					**−0.065****
	2005					**0.087**
Total direct effect		**1.68**	**0.195**	**0.529**	**−0.266**	**0.077**
Total effect		1.68	0.2667	0.2272	−0.2907	0.05

Table 3. Path analysis showing direct and indirect effect of $Y_1 - Y_5$ to Z (*Pascopyrum smithii* Schreb.). *Significant at the 0.05 probability level. **Significant at the 0.01 probability level. The direct effects of $Y_1 - Y_5$ to Z are highlighted in bold (on the main diagonal cells); Arrows indicate the direction of the effect.

(Table 3). With Y_4, the influence was negative, but the overall order of effects was $Y_1 > Y_2 > Y_3 > Y_5 > Y_4$ (1.68, 0.2667, 0.2272, 0.05 and −0.2907, respectively, Table 3).

Ridge regression models of seed yield and 5 seed yield components. Seed yield (Z) was highest in 2003 followed by 2004 and 2005 (Table 4). Y_1 was highest in 2003 and produced the highest Z. It has been suggested that the value of the ridge parameter K[11] should be determined using ridge traces (Fig. 1). For 2003, 2004 and 2005, the curves for Y_1 to Y_5 showed estimated k values at 0.77, 0.64 and 0.79, respectively (Fig. 1). The ridge regression models A, B and C for 2003, 2004 and 2005, respectively (Table 4), were as follows:

A. $Z = 529.067 + 0.371 \times Y_1 + 19.944 \times Y_2 + 15.127 \times Y_3 − 55.349 \times Y_4 − 29.703 \times Y_5$
(Ridge k = 0.77; F = 8.274 Pr < 0.0001)

B. $Z = 444.094 + 0.358 \times Y_1 + 2.075 \times Y_2 − 2.205 \times Y_3 − 0.920 \times Y_4 − 4.078 \times Y_5$
(Ridge k = 0.64; F = 114.768 Pr < 0.0001)

C. $Z = 271.685 + 0.381 \times Y_1 + 14.551 \times Y_2 − 7.485 \times Y_3 + 18.647 \times Y_4 − 48.165 \times Y_5$
(Ridge k = 0.79; F = 10.797 Pr < 0.0001)

	year	N	Y_1	Y_2	Y_3	Y_4	Y_5	Z
Duncan's Multiple Range Test								
	2003	105	646.14 a	17.5805 a	12.5887 a	5.7998 a	4.8564 a	822.21 a
	2004	129	236.44 b	18.3464 c	7.3498 b	4.4944 b	4.5575 ab	492.49 b
	2005	146	401.85 c	18.7094 b	7.5531 c	4.7874 c	4.5718 b	473.47 c
	F Value		363.244	27.27	31.13	29.11	159.52	132.88
	Pr > F		<0.0001	<0.0001	<0.0001	<0.0001	<0.001	<0.0001
Ridge regression coefficients								
k	year	Intercept	Y_1	Y_2	Y_3	Y_4	Y_5	Z
0.77	2003	529.067	0.371	19.944	15.127	−55.349	−29.703	−1
0.64	2004	444.094	0.358	2.075	−2.205	−0.920	−4.078	−1
0.79	2005	271.685	0.381	14.551	−7.485	18.647	−48.165	−1

Table 4. Duncan's Multiple Range Test of the *Pascopyrum smithii* seed yield (Z) and yield components (Y_1–Y_5) for the 3 years and of the ridge regression coefficients. Means with the same letter are not significantly different at Alpha = 0.05.

Source	DF	Sum of squares	Mean square	F value	Pr > F
Model	1	43398208.445	43398208.445	1047.004	<0.0001
Error	379	183167057.556	41449.889		
Corrected total	380	226565266.002			

Table 5. Analysis of variance for dependent variable Z_{actual} with the 5 seed-yield components of the 380 samples.
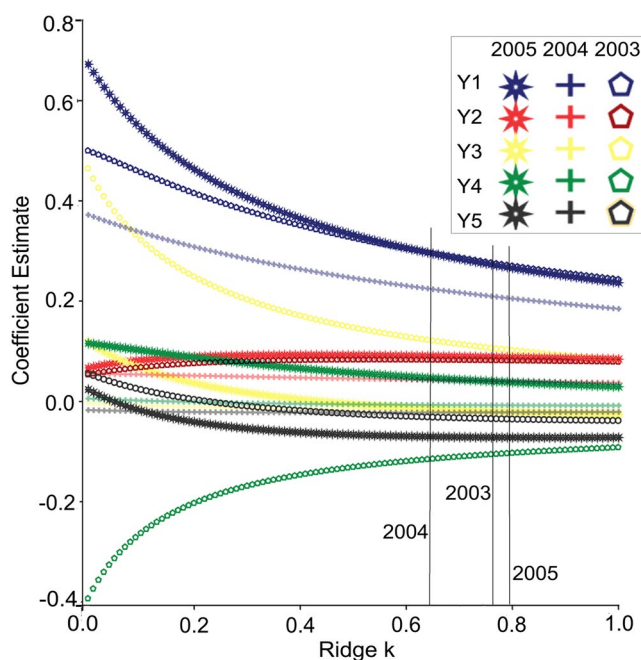

Figure 1. Ridge traces of standard partial regression coefficients for increasing values of k for five yield components for year 2003, 2004 and 2005, respectively. Y_1 to Y_5 are fertile tillers m^{-2}, spikelets per fertile tillers, florets per spikelet, seed numbers per spikelet and seed weight, respectively.

The highest absolute values of the ridge regression coefficients for Y_4, Y_3 and Y_2 occurred in 2003, and for Y_5 and Y_1, the highest values occurred in 2005 (Table 4). For a reliable model from the data of three successive years, the 380 samples of Z with Y_1 to Y_5 in the database were transformed using the natural logarithm: $S = \ln Z$, $C_1 = \ln Y_1$, $C_2 = \ln Y_2$, $C_3 = \ln Y_3$, $C_4 = \ln Y_4$ and $C_5 = \ln Y_5$.

Then, the ridge regression model was obtained with S and C_1-C_5 as follows: (The variance analysis and the parameter estimates are listed in Tables 5 and 6 respectively).

Variable	DF	Parameter estimate	Standard error	t value	Pr > |t|
Intercept	1	−230.174	24.195	−9.513	<0.0001
$Z_{estimated}$	1	1.568	0.048	32.357	<0.0001

Table 6. Parameter estimates for $Z_{estimated}$.

Variable	DF	Parameter estimate	Standard error	t value	Pr > |t|
Intercept	1	−0.033	17.164	−0.002	0.998
$Z_{estimated}$	1	1.000	0.031	32.357	<0.0001

Table 7. Parameter estimates for $Z_{estimated}$ after adjustment by linear regression.
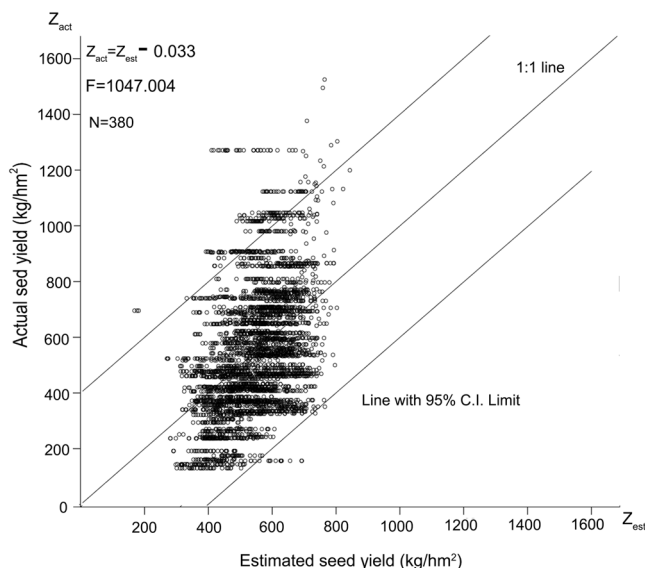
Figure 2. Scatter plot to fit regression line of actual and estimated seed yield adjusted by
$Zactual. = -230.174 + 284.235 \times Y_1^{0.21} \times Y_2^{0.1} \times Y_3^{0.01} \times Y_4^{-0.01} \times Y_5^{-0.30}$ of the 3 years. It is superimposed on the 1:1 line.

$$S = 5.219 + 0.211 \times C_1 + 0.095 \times C_2 + 0.005 \times C_3 - 0.004 \times C_4 - 0.295 \times C_5$$
$$(N = 380, F = 209.514, Pr < 0.0001) \tag{1}$$

Thus,

$$\ln Z = 5.219 + 0.211 \times \ln Y_1 + 0.095 \times \ln Y_2 + 0.005 \times \ln Y_3 - 0.004 \times \ln Y_4 - 0.295 \times \ln Y_5 \tag{2}$$

Model (2) was transformed to an exponential function:

$$Z = 181.272 \times Y_1^{0.21} \times Y_2^{0.1} \times Y_3^{0.01} \times Y_4^{-0.01} \times Y_5^{-0.30} \tag{3}$$

Equation (3) was used to estimate the seed yield of all 380 samples, and the results were denoted as $Z_{estimated}$. The actual seed yields were denoted as Z_{actual}. To test the accuracy, the values of Z_{actual} to $Z_{estimated}$ were used for linear regression (analyses of the variance is shown in Tables 5 and 6). The linear model was as follows:

$$Zactual = -230.174 + 1.568 \times Z_{estimated} \ (N = 380, \ F = 1047.004, \ Pr < 0.0001) \tag{4}$$

Then, via formula (4), the model was adjusted

$$Zactual = -230.174 + 284.235 \times Y_1^{0.21} \times Y_2^{0.1} \times Y_3^{0.01} \times Y_4^{-0.01} \times Y_5^{-0.30} \tag{5}$$

The variance test estimated that the intercept and Z were −0.033 and 1.000, respectively (Table 7), which are presented in Fig. 2, superimposed on the 1:1 line.

We determined the antagonistic and synergistic effects among the Ys on Z using pairwise models (Figs. 3 and 4). In addition, the results of synergism and antagonism among the Y_1 to Y_5 on Z are discussed (Figs. 3 and 4).
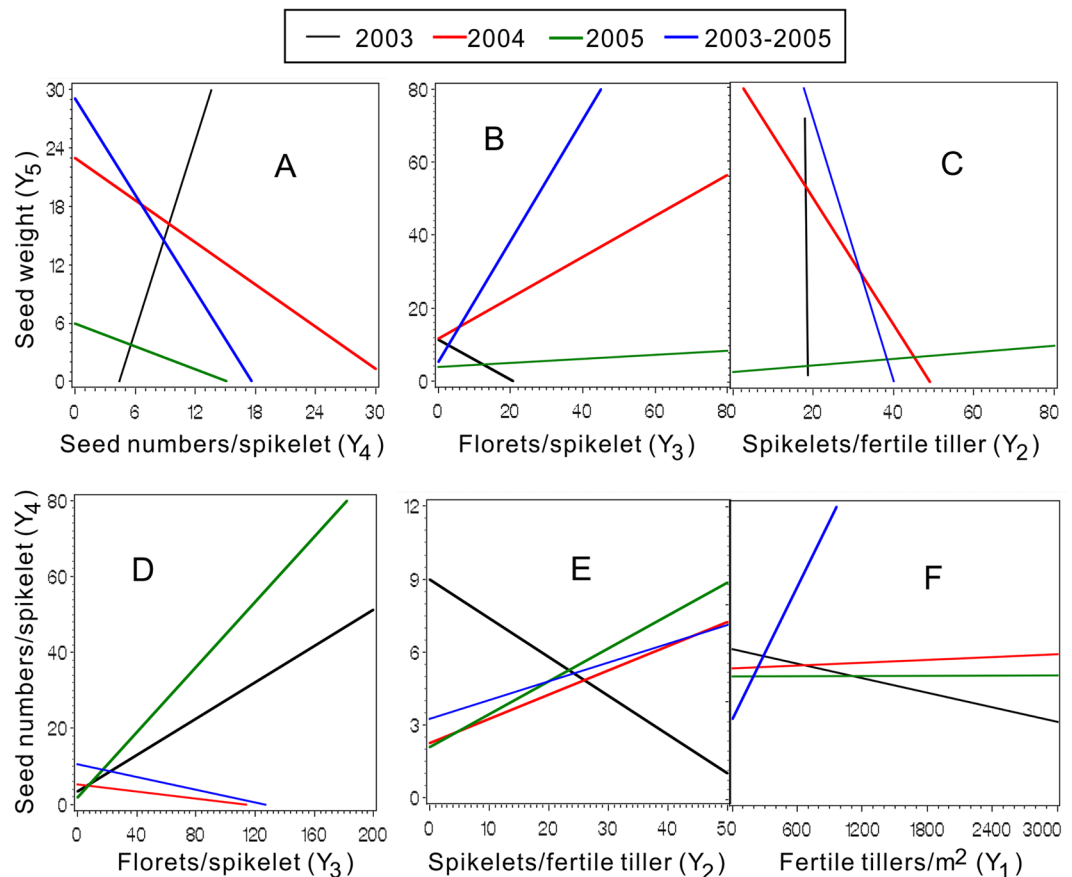
Figure 3. Ridgelines of the response surface models showed the synergism and antagonism through Y_4 (**A**) Y_3 (**B**) and Y_2 (**C**) to Y_5 and Y_3 (**D**) Y_2 (**E**) and Y_1 (**F**) to Y_4.

Discussion

In this study, precipitation, temperature, and sunlight during the crop-growing period from March to early September for the three years of the study were provided by the Jiuquan Meteorological Observatory of Gansu Province, China (Fig. S1). Temperature and precipitation during the crop growing seasons from 2003–2005 were near the average values for the past ten years in this location (Table S14). Thus, this data should accurately represent the natural field conditions.

The average *Pascopyrum smithii* seed yield (Z) and the yield components (Y_1, Y_2, Y_3, Y_4 and Y_5) were very different from 2003 to 2005 (Table 8), mainly owing to climatic conditions (Supplementary Fig. 1), e.g., large precipitation differences between 2003, 2004 and 2005. Moreover, higher rainfall in June, which occurred during the seed growth period, was partly responsible for higher seed yields because it favored pollination and grain filling. As another example, the highest recorded rainfall was in March 2005 (28.2 mm), together with a low air temperature. These factors promoted vegetative growth and significantly decreased Y_1 (Table 4), which consequently resulted in the lowest values for Z. The highest values for Z appeared in 2003 because the mild climate and adequate water supply were conducive to crop tillering and yield improvement. In comparison, the higher Z in 2004 corresponded to higher temperatures and adequate water in June and July in 2004 compared to 2005. However, Y_1 and Y_3 weakly decreased along with age from 2003 to 2005, which may have been due to genetic factors, while larger difference indicate the role of environment[30,31]. Fertile tillers m^{-2} (Y_1), florets/spikelet (Y_3), and seed numbers/spikelet (Y_4) were higher in 2003 compared to 2004 and 2005, whereas spikelets/fertile tillers (Y_2) was lower in 2003 than in 2004 and 2005. Seed weight (Y_5), in contrast, was similar across the years. For each of the traits evaluated, descriptive statistics, including the coefficient of variation (CV), mean, minimum, maximum, Std-Dev and Std-Error values are summarized in Table 8. These indicators have strong differences for improving the expression of traits and provide a good opportunity to cultivate excellent species. The variabilities in seed yield and the yield components in 2003, 2004 and 2005 may have been caused by a stand age divergence, air temperature differences, interactions between soil fertility and climatic conditions, or combinations thereof.

Correlation analysis considers a mutual association with no regard to causation, whereas path analysis specifies causes and measures their relative importance[32]. Because the total correlations between the predictor variables and the response variable, which are partitioned into direct and indirect affects, and the direct and indirect effects of the yield factors can be determined through path analysis, and knowledge on the direct and indirect correlations, especially of the yield, allows breeders to use this additional information to discard or promote genotypes of interest[33]. Alvi *et al.*[34] and Asghari-zakaria *et al.*[35] used path analysis to enable the identification
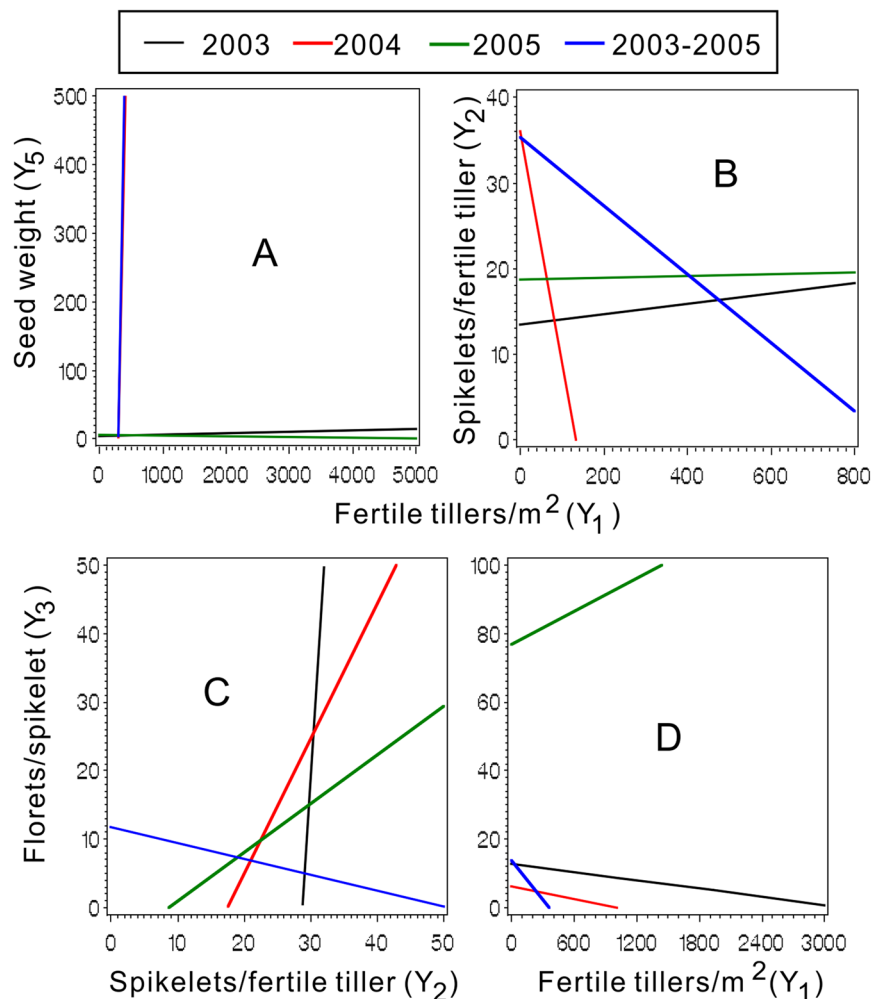
Figure 4. Ridgelines of the response surface models showed the synergism and antagonism through Y_1 to Y_5 (**A**) and Y_2 (**B**), Y_2 (**C**) and Y_1 (**D**) to Y_3.

Variable	year	N	Mean	Std-Dev	Std-Error	Minimum	Maximum	CV	Pr > |t|
Y_1	2003	105	674.4784	178.2814	17.3985	192.38	1153.33	0.27	<0.0001
	2004	129	267.7532	126.9550	1.9660	34	916	0.48	<0.0001
	2005	146	434.7001	161.2333	13.3438	167.17	830.67	0.37	<0.0001
Y_2	2003	105	17.6083	0.9908	0.0967	15.21	20.21	0.06	<0.0001
	2004	129	18.8091	4.3803	0.0678	9	40	0.23	<0.0001
	2005	146	18.7603	1.3885	0.1149	15.63	22.03	0.08	<0.0001
Y_3	2003	105	12.6991	1.6743	0.1634	8.84	17.12	0.13	<0.0001
	2004	129	7.5309	1.6297	0.0252	3	16	0.22	<0.0001
	2005	146	7.5742	0.5718	0.0473	6.4	9.03	0.08	<0.0001
Y_4	2003	105	5.8180	0.4572	0.0446	4.39	6.64	0.08	<0.0001
	2004	129	4.8165	1.6596	0.0257	1	13	0.35	<0.0001
	2005	146	4.8125	0.4960	0.0410	3.7	6.33	0.10	<0.0001
Y_5	2003	105	4.8644	0.2823	0.0275	4.336	5.802	0.06	<0.0001
	2004	129	4.5932	1.1330	0.0175	3.19	44.61	0.25	<0.0001
	2005	146	4.5838	0.3326	0.0275	3.91	5.27	0.07	<0.0001
Z	2003	105	856.1210	241.0244	23.5216	371.69	1525.07	0.28	<0.0001
	2004	129	539.4421	220.5409	3.4152	130.15	1271.088	0.41	<0.0001
	2005	146	523.4232	225.8983	18.6955	130.15	1271.088	0.43	<0.0001

Table 8. Statistics of Y_1~Y_5, Z (*Pascopyrum smithii* Schreb.) for years 2003~2005.

of traits that are useful for an evaluation standard in increasing crop yield. Cruz et al.[36] defines the path coefficient, or cause and effect analysis, as a standardized regression coefficient[37], because path analysis is composed of an expansion of multiple regressions when complex interrelationships are involved. Thus, path analysis has been found to be a useful technique of statistical analysis specially designed to quantify direct and indirect trait association with yield[38]. In this study, path analysis indicates that the total direct effects of Y_1, Y_2, Y_3 and Y_5 have highly significant positive correlations with Z, but Z is negatively correlated with Y_4 (Table 3). The explanation for why there is a weak negative correlation between the seed number (Y_4) and seed yield (Z) might matter to a high density cultivation and soil nutrient limitation[7]. This finding is consistent with the literature, e.g., the yield of mechanically harvested rapeseed (*Brassica napus* L.) can be increased by optimum plant density and row spacing[39]. Table 3 shows that Y_1 and Y_2 have the largest correlation coefficient and contribution rate, which is consistent with the natural law of plant growth and biological theory. Nevertheless, the correlation of Z with Y_3 (Table 1), which was not significant, is probably due to the effects of aging and climate during the individual years and to the field management systems that were repeated yearly[40]. However, Y_3 did not significantly contribute to Z partly because Y_3 was mostly under genetic control[40]. This discovery implies that Y_3 is the sub group that should be taken into account if high quality forage is the target of the breeding system. Nevertheless, Y_1 was the most critical and available group that significantly contributed to Z ($P < 0.001$): the coefficients were 0.480, 0.423, and 0.777 for 2003, 2004 and 2005, respectively. This finding is consistent with previous studies in fescues[7,41], Russian wild rye[42,43], Zoysia grass[1], gramineous plants[5], and white clover[40,44,45]. Additionally, path analysis uncovered relationships between the components and the yield that are consistent with previously reported results[45]. The interrelationships among Y_1 to Y_5 indicated that Y_4 and Y_5 had remarkable negative relationships with seed yield (Z; r = −0.03* and r = −0.192**). Y_1 and Y_2, as indicated above, had significant and positive correlations with Z. Because of the existence of positive and negative correlations between the seed yield components, simple linear relationships between two components by a correlation analysis cannot be successfully predicted by the measurements. However, with standardized variables, a path analysis can determine the relative importance of direct and indirect effects on seed yield. The results of this study further emphasize 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 is consistent with the results of previous research[40]. 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.

Ridge regression and multiple-regression analyses were applied to avoid high inter-correlation and multico-linearity among the variables[46–48]. The significant correlation coefficients ($P < 0.001$ and 0.01), path analyses, and ridge regressions of the multifactor orthogonal experimental design and large sample statistical analysis in the field experiments show that the models are reliable[48]. In addition, ridge regression effectively overcomes the problem of highly multi-correlated predictor variables (seed yield components)[48,49]. This method is the most effective and practical for the current field scientific research[43]. Unfortunately, owing to the aging of the plants, designed field management and climate conditions, the coefficients of the ridge regression models in individual years are variable, ranging from −55.349 to 19.944 (Table 4). An original exponential model was found to estimate Z via Y_1 through Y_5. First, the final algorithm model [exponential Eq. (4)] was deduced using data for 380 plots from various growth regimes in the three successive years. Moreover, all three ridge regression models [The ridge regression models A, B and C] for the individual years were significant ($P < 0.001$), and they all had coefficients matching the contributions of the five Ys to Z. This result is explained through the relationship between the path analysis and ridge regression analysis, and the test methods and results are more credible. In addition, the contributions in absolute value of the five seed yield components to the seed yield are in the following order: $Y_1 > Y_3 > Y_4 > Y_2 > Y_5$ (1.68, 0.529, −0.266, 0.195 and 0.077, Table 3). The total direct effects with Y_4 having negative results, showed that the total influence order, in absolute values, is $Y_1 > Y_4 > Y_2 > Y_3 > Y_5$ (1.68, −0.2907, 0.2667, 0.2272 and 0.05, Table 3). However, the two pairs of results are from the same database. The ridge analysis values analytically combine the effects of all Ys, especially the effects of aging and climate, to address the variation in Z for the three years, whereas the path analysis includes separate analytic effects of the individual three years.

The antagonistic and synergistic effects between Ys on Z were investigated using pairwise quadratic regression models (Table 9, Figs. 3 and 4). The lines for 2004 (red) and 2003–05 (blue) showed uniform orientations, except for Y_4 & Y_2, among the ten relation schema subgraphs (Figs. 3 and 4), as did the lines for 2005 (green), except for Y_5 & Y_4. The subgraphs indicated that Y_4 & Y_5 (Fig. 3A), Y_2 & Y_5 (Fig. 3C), Y_3 & Y_4 (Fig. 3D), Y_1 & Y_5 (Fig. 4B) and Y_1 & Y_3 (Fig. 4D) had antagonistic effects on Z, as the ridgelines were at k < 0 (Table S4)[49]. Conversely, the red and blue lines that also had the same directions indicated that Y_3 & Y_5, Y_2 & Y_4 and Y_1 & Y_4 had synergetic effects on Z (Fig. 3B,E,F) at k > 0 in the ridgelines. The more Y_2, less Y_5 dynamic (Fig. 3C), which was also evident for Y_3 & Y_4 (Fig. 3D), were mostly caused by feedforward compensation at the biological level and by the soil nutrient limitation. In Fig. 4A, the blue and red lines almost overlap and are nearly perpendicular to the horizontal axis, whereas the green and black (2003) lines are also partly overlapped and nearly perpendicular to the longitudinal axis. This may be due to genetic constraints on Y_5 in a certain range of growth. In this range, Y_1 has the optimal value. The increase in Y_1-derived Y_2 and the Y_3 decrease (Fig. 4B,D) were consistent with the soil nutrient limitation. For 2003, 2004, 2005 and 2003–2005, the interactions between Y_2 & Y_3 gradually changed from synergetic to antagonistic (Fig. 4C). As important seed yield traits of the plant, Y_2 & Y_3 are regulated by genetics. Further investigations will be needed to verify that these changes are probably due to aging of the grass.

Conclusions

Algorithmic models were developed to describe the seed yield and yield components needed to improve the seed yield of *Pascopyrum smithii*. Significant positive correlations were observed between seed yield (Z) and spikelets/fertile tillers (Y_2) and fertile tillers m^{-2} (Y_1), and negative correlations were found with seed number/spikelet (Y_4). The inter-correlation among these components were significant in 2004 and 2005.

		Y_a^2	Y_b^2	Y_a*Y_b	Y_a	Y_b	Constant	F Value	$P > F$
2003	Y_1Y_2	0.001	23.052	−0.276	4.505	−622.169	4490.5	10.30	<0.0001
N = 146	Y_1Y_3	0.0003	4.719	0.035	−0.054	−119.522	1182.2	10.63	<0.0001
	Y_1Y_4	0.0003	139.517	0.157	−0.597	−1719.56	5747.4	10.09	<0.0001
	Y_1Y_5	0.001	−120.43	0.387	−2.25	960.74	1120.8	9.91	<0.0001
	Y_2Y_3	−1.795	0.778	1.913	112.84	−45.34	−551.4	1.96	**0.092**
	Y_2Y_4	−4.117	164.24	36.83	−9.73	−2642.62	8321.6	3.61	0.0048
	Y_2Y_5	22.54	95.228	189.01	−1667.9	−4346.47	25926.8	3.16	0.0110
	Y_3Y_4	26.74	446.78	−210.7	633.1	−2887.0	5704.3	6.09	<0.0001
	Y_3Y_5	1.35	−164.8	−115.7	526.52	2969.8	−9435.1	2.37	0.0455
	Y_4Y_5	170.1	−79.72	−381.8	−232.4	2935.4	−5155.2	4.05	0.0022
2004	Y_1Y_2	−0.001	−0.322	−0.14	1.545	19.74	76.6	177.85	<0.0001
N = 4221	Y_1Y_3	−0.001	1.280	0.012	1.268	−14.19	345.5	173.16	<0.0001
	Y_1Y_4	−0.001	−2.985	−0.001	1.332	33.41	236.4	172.92	<0.0001
	Y_1Y_5	−0.001	0.145	0.388	−0.583	−114.70	863.0	187.95	<0.0001
	Y_2Y_3	−0.545	−0.064	0.864	20.767	−13.58	363.1	15.89	<0.0001
	Y_2Y_4	−0.531	−3.36	−0.435	28.73	42.79	145.0	13.14	<0.0001
	Y_2Y_5	−0.45	2.228	2.483	11.153	−155.42	969.0	69.62	<0.0001
	Y_3Y_4	0.063	−3.082	−0.153	0.997	32.46	486.3	6.54	<0.0001
	Y_3Y_5	0.580	2.554	−3.852	12.41	−91.55	931.4	64.87	<0.0001
	Y_4Y_5	−3.292	2.408	−1.797	42.35	−109.45	932.2	64.15	<0.0001
2005	Y_1Y_2	0.001	−20.61	0.049	−0.093	773.29	−6964.8	16.11	<0.0001
N = 105	Y_1Y_3	−0.001	−1.135	0.035	1.456	178.76	−1293.1	19.95	<0.0001
	Y_1Y_4	−0.001	−148.57	0.003	1.562	1498.01	−3674.6	18.84	<0.0001
	Y_1Y_5	−0.001	−312.09	−0.776	5.363	3265.24	−8375.3	15.32	<0.0001
	Y_2Y_3	−18.17	−33.81	39.38	448.06	−205.19	−3504.8	8.40	<0.0001
	Y_2Y_4	−25.70	−185.18	−1.496	1024.87	1797.63	−13779.0	5.56	<0.0001
	Y_2Y_5	−22.65	−238.08	−3.591	911.35	2198.11	−13273.1	7.56	<0.0001
	Y_3Y_4	−82.24	−148.47	−64.47	1500.05	1885.23	−9322.8	4.18	0.0014
	Y_3Y_5	−41.82	−326.69	−51.10	900.02	3212.32	−9896.0	6.52	<0.0001
	Y_4Y_5	−82.85	−327.26	−303.08	2198.81	4280.04	−14104.2	6.00	<0.0001
2003–05	Y_1Y_2	0.0005	−0.301	−0.023	1.073	21.52	109.6	214.62	<0.0001
N = 4426	Y_1Y_3	−0.001	0.823	0.059	0.485	−21.42	479.1	227.53	<0.0001
	Y_1Y_4	0.0004	−2.454	0.045	0.490	16.60	374.4	212.55	<0.0001
	Y_1Y_5	−0.0003	0.224	0.557	−1.886	−166.30	1161.5	237.93	<0.0001
	Y_2Y_3	−0.648	4.445	0.602	27.02	−69.91	483.1	26.47	<0.0001
	Y_2Y_4	−0.661	−4.380	−0.699	35.11	61.48	42.3	13.73	<0.0001
	Y_2Y_5	−0.613	1.581	2.194	19.10	−118.23	759.8	57.30	<0.0001
	Y_3Y_4	4.169	−3.831	1.110	60.67	31.19	691.9	16.31	<0.0001
	Y_3Y_5	4.995	2.425	−2.908	−51.60	−92.83	1133.8	73.72	<0.0001
	Y_4Y_5	−4.448	1.841	−1.059	54.13	−83.82	795.4	51.98	<0.0001

Table 9. Coefficients of pair wised models among the seed yields components.

The model of seed yield with its 5 components, based on a large sample size from an orthogonal experimental design in *Pascopyrum smithii*, was:

$$Z = -230.174 + 284.235 \times Y_1^{0.21} \times Y_2^{0.1} \times Y_3^{0.01} \times Y_4^{-0.01} \times Y_5^{-0.30}$$

This model can be used to accurately estimate the seed yield with five yield components.

The total direct effects of fertile tillers, florets/spikelet, spikelets/fertile tiller and seed weight on seed yield were positive, and fertile tillers was the largest contributor. The contributions in decreasing order were fertile tillers(Y_1) > seed number/spikelet ($Y4$) > spikelets/fertile tiller (Y_2) > florets/spikelet (Y_3) > seed weight(Y_5). Fertile tillers (Y_1) was one of the most important factors that played a key role in seed production. Therefore, selection for high seed yield through direct selection for large fertile tillers (Y_1), florets/spikelet (Y_3) and seed number/spikelet ($Y4$) would be effective and reliable for breeding, based on the experimental data of three years of continuous field experiments. Finally, this research laid the foundation for the basic theory of restoration ecology in arid regions and for the promotion of plant carbon cycles to reduce the greenhouse effect. Further studies should be focused on changes in seed yield based on different climatic conditions and site locations.

Materials and Methods

Experimental site description.   The cultivar "Rosana" of *P. smithii* which is commonly planted was introduced from the United States in 2002[the 948 project (202009) of the Ministry of Agriculture of China]. Field experiments were conducted over three years (2003–2005) at the China Agricultural University Grassland Research Station located at the Hexi Corridor, in Jiuquan, Gansu province, northwestern China (39°37′ N latitude and 98°30′ E longitude; altitude 1480 m). Soil at the site is classified as a Mot-Cal-Orthic Aridisol in the Chinese system and as a Xeric Haplocalcid in the USDA soil classification system[50]. The plots used in this experiment were planted with alfalfa (*Medicago sativa* L.) in the preceding season. The 6000 m^2 experimental site was tilled using a chisel plow in the fall and disk-harrowed in the spring for seedbed preparation. *Pascopyrum smithii* seeds were planted on April 23, 2002 at a depth of 2.5 cm. The seeding rate was 500 seeds m^{-2}, and the space between rows was 0.45 m. Initial fertilizer was applied in a band that was 6 cm deep and 5 cm to the side of the seed furrows at a rate of 104 kg ha^{-1} N and 63 kg ha^{-1} P_2O_5. There was no seed yield in autumn 2002. Initial chemical characteristics of the soil (0–20 cm) were: pH = 8.39; NH_4^+ = 32.32 mg kg^{-1}; NO_3^- = 20.09 mg kg^{-1}; alkali hydrolysable nitrogen = 118.30 mg kg^{-1}; available phosphorus = 36.56 mg kg^{-1}; available potassium = 130.30 mg kg^{-1}; total nitrogen = 0.764 g kg^{-1}; total phosphorus = 0.814 g kg^{-1}; total potassium = 12.52 mg kg^{-1}; organic matter = 10.32 g kg^{-1} (Table S1).

Experimental design.   The Orthogonal Experimental Design (OED) method is typically used to study the comparative effectiveness of multiple intervention components simultaneously; OED with both Orthogonal Array (OA) and Factor Analysis (FA) makes it possible to discover the optimum combinations with only several tests[51,52] (Table S2). Because OED is characterized by equilibrium dispersion, we were able to design experiments to find the best combination of treatments with a minimum of tests. Additionally, OED allowed us to transform complex multi-factor data to single factor analysis. We used a multi-factorial orthogonal design for field plots based on the six groups (Table S2)[53], giving a total of 380 experimental plots (each with an area of 28 m^2), under various field management treatments (Table S2) within a total field area of 4100 m^2. Plots were irrigated five times during the growing season at the following growth stages: vegetative phase jointing, stem formation, ear formation and flowering, respectively. Fertilization was carried out before sowing and again before spring regrowth the following year.

Data collection.   Ten samples along 1 m of each row were randomly selected to measure the five seed yield components from anthesis to seed harvest from 2003 to 2005. Plants that were 1 m or less from the edge of the plot we not sampled. Seed yield components and seed yield data for each plot were collected as follows: fertile tillers m^{-2} (Y_1) were measured from ten randomly selected 1-m row samples, and 30 to 36 fertile tillers and 27 to 54 spikelets were randomly selected for measuring spikelets/fertile tillers (Y_2), florets/spikelet (Y_3) and seed numbers/spikelet (Y_4). When the seed heads were ripe, 4 samples from 1 m of the row length were separately threshed by hand, the yield of clean seed for each sample was weighed, and the seed moisture content was confirmed as 7 to 10% for converting into seed yield (kg hm^{-2}) (Z). Ten lots of 100 seeds each were collected to determine mg seed weight (Y_5). The total number of samples (n) used to measure Y_1 to Y_5 and Z were 3800, 13605, 11085, 10770, 3800, and 1520, across all 3 years (Table S3). The sample size for individual years is shown in Supplementary Table 3, and the experimental databases were established with Visio FoxPro (Version 6.0).

Statistics and analytical methods.   Path coefficient analysis helps to determine the direct effect of traits and their indirect contributions via other characters[38,54]. Correlation and path analysis were performed to determine the relationship among the yield and yield contributing characters. Thus, separate and combined analyses for the three years provided useful information[48].In addition, ridge regression is a useful parameter estimation method for addressing the collinearity problem frequently arising in multiple linear regression. Ridge regression provides a means of addressing the problem of collinearity without removing variables from the original set of independent variables. Ridge regression analysis[55] and Duncan's multiple range tests for seed yield (Z) and yield components (Y_1–Y_5) were performed. The data were transformed using logarithmic and power transformations to avoid the effects of highly inter-correlated data, which would lead to multico-linearity between Y_1–Y_5 and Z. To establish a reliable model, all of the Z and Y_1–Y_5 data in Visio FoxPro, representing a total of 380 samples (plots) in the three years (i.e. 105 + 129 + 146), were taken as a natural logarithm because mathematically they did not influence the essential relationships between the variables. Analyses of variance and Pearson correlation analyses were performed using SPSS Version 19.0.

If S = lnZ and C_i = lnY$_i$ (i = 1 to 5), then S and C_1 to C_5 were used for the ridge regression analyses, and the ridge regression model was:

$$S = C \times \beta + u \tag{6}$$

where S is the n × 1 vector of observations of a response variable, C is the n × p matrix of observations on p explanatory variables, β is the p × 1 vector of regression coefficients, and u is the n × 1 vector of residuals satisfying E(ū) = C', E(uu') = δ^{2}I.

It is assumed that C and S have been scaled such that C' C and S' S are matrices of correlation coefficients. In this equation, n = 380, and p = 5. Thus,

$$LnZ = \left(\sum_{i=1}^{5} Ln\ Y_i \right) \times \beta + u \tag{7}$$

The above logarithmic model (7) was transformed to the following exponential function:

$$Z = e^a \times \prod_{i=1}^{5} (Y_i^\beta),$$

(8)

where α and β are constants.

Formula (8) was used to estimate the Z of all 380 samples, which is denoted as $Z_{estimated}$; the actual seed yield is denoted as Z_{actual}.

A general linear regression model was used to assess the Z_{actual} compared to the $Z_{estimated}$, and an analysis of variance was used to assess the dependent variable Z_{actual} and the parameter estimates of $Z_{estimated}$. The linear regression model is:

$$Z_{actual} = \beta + K \times Z_{estimated}$$

(9)

So, via formula (9), the model was adjusted to:

$$Z = \beta + k \times e^a \times \prod_{i=1}^{5} (Y_i^\beta).$$

(10)

In addition, the ridge trace and appropriate scatter plots were graphed. The analyses and graphical procedures specified above were all performed using SAS Version 8.2 (Inc. 1988).

Quadratic two-variable regression models between Z and Y_1 to Y_5 were used as follows:

$$Z = \sum_{i=1}^{2} \left(\beta_{i \times j+1} Y_j^i \right) + u(i = 1, 2; j = 1, 2),$$

(11)

Where β is a constant.

The equivalent effects of Yi and Yj were determined using:

$$\frac{\partial Z}{\partial Y_i} = \frac{\partial Z}{\partial Y_j}$$

(12)

which produced

$$Y_j = k \times Y_i \pm b,$$

(13)

where b is a constant. The presented ridgelines (13) (Figs. 3 and 4) correspond to the response surface models (11) to show the synergetic and antagonistic effects. The analyses and graphics were all performed using the SAS (v8.2) software[49].

References

1. Ma, C. H. *et al*. Effects of nitrogen fertilizer on seed yields and yield components of *Zoysia japonica* established by seeding and transplant. *Agri Sci China* **3**, 553–560, http://www.cnki.com.cn/Article/CJFDTotal-ZGNX200407010.htm (2004).
2. Belesky, D. P., Ruckle, J. M. & Halvorson, J. J. Carbon isotope discrimination as an index of tall fescue–endophyte association response to light availability and defoliation. *Environ Exp Bot.* **67**, 515–521, https://doi.org/10.1016/j.envexpbot.2009.09.009 (2010).
3. Berdahl, J. D. & Ries, R. E. Development and vigor of diploid and tetraploid Russian wildrye seedlings. *J Range Manage* **6**, 80–84, https://doi.org/10.2307/4002709 (1997).
4. Jefferson, P. & Muri, R. Competition, light quality and seedling growth of Russian wildrye grass (Psathyrostachys juncea). *Acta Agrestia Hungarica* **55**, 49–60, https://doi.org/10.1556/AAgr.55.2007.1.6 (2007).
5. Hamel, C. *et al*. Arbuscular mycorrhizal fungi and nematodes are involved in negative feedback on a dual culture of alfalfa and Russian wildrye. *Appl Soil Ecol.* **40**, 30–36, https://doi.org/10.1016/j.apsoil.2008.03.004 (2008).
6. Ali, S. *et al*. Planting patterns and deficit irrigation strategies to improve wheat production and water use efficiency under simulated rainfall conditions. *Front Plant Sci.* **8**, 1408–1416, https://doi.org/10.3389/fpls.2017.01408 (2017).
7. Wang, Q., Zhou, H. & Han, J. Analysis on a model for water and fertilizer coupling effects on *Psathyrostaehys juncea* seed yield. *Acta Prata Sinica.* **14**, 41–50, https://doi.org/10.3321/j.issn:1004-5759.2005.06.006 (2005b).
8. Ott, J. P., Butler, J. L., Rong, Y. & Xu, L. Greater bud outgrowth of *Bromus inermis* than *Pascopyrum smithii* under multiple environmental conditions. *J Plant Ecol.* **10**(3), 518–527, https://doi.org/10.1093/jpe/rtw045 (2016).
9. Wang, M. *et al*. Influence of row spacing and P and N applications on seed yield components and seed yield of *Siberian Wildrye* (L.). *Crop Sci.* **57**, 2205–2213, https://doi.org/10.2135/cropsci2016.08.0713 (2017).
10. Han, Y. *et al*. Effect of row spacing on seed yield and yield components of five cool-season grasses. *Crop Sci.* **53**, 2623–2631, https://doi.org/10.2135/cropsci2013.04.0222 (2013).
11. Ali, S., Khan, H., Shah, I., Butt, M. M., & Suhail, M. A comparison of some new and old robust ridge regression estimators. *Commun Stat-Simul C.* 1–19, https://doi.org/10.1080/03610918.2019.1597119 (2019).
12. Li, X. & Yang, Y. A novel perspective on seed yield of broad bean (*Vicia faba* L.): differences resulting from pod characteristics. *Sci Rep.* **4**, 6859, https://doi.org/10.1038/srep06859 (2014).
13. Iannucci, A. & Martiniello, P. Analysis of seed yield components in four Mediterranean annual clovers. *Field Crops Res.* **55**, 235–243, https://doi.org/10.1016/S0378-4290(97)00091-9 (1998).
14. Song, M. *et al*. Genetic analysis of yield and yield-related traits in short-season cotton (*Gossypium hirsutum* L.). *Euphytica* **204**, 135–147, https://doi.org/10.1007/s10681-014-1348-1 (2015).
15. Khan, S. *et al*. Optimization of nitrogen rate and planting density for improving yield, nitrogen use efficiency, and lodging resistance in Oilseed Rape. *Front Plant Sci.* **8**, 532–541, https://doi.org/10.3389/fpls.2017.00532 (2017).

16. Fırıncıoğlu, H. K., Ünal, S., Erbektaş, E. & Doğruyol, L. Relationships between seed yield and yield components in common vetch (*Vicia sativa* ssp. sativa) populations sown in spring and autumn in central Turkey. *Field Crops Res.* **116**, 30–37, https://doi.org/10.1016/j.fcr.2009.11.005 (2010).

17. Môro, G. V., Santos, M. F. & Júnior, C. Use of genomic and phenotypic selection in lines for prediction of testcrosses in maize I: grain yield and its components. *Euphytica* **9**, 213–220, https://doi.org/10.1007/s10681-017-1915-3 (2017).

18. Annapurna, D., Khan, H. & Mohammad, S. Genotypic-phenotypic correlations and path coefficient analysis between seed yield and other associated characters in tall genotypes of maize. *Field Crops Res.* **16**, 205–209, https://doi.org/10.4238/gmr16029696 (1998).

19. Assefa, Y. *et al*. Analysis of long term study indicates both agronomic optimal plant density and increase maize yield per plant contributed to yield gain. *Sci Rep.* **8**, 4937, https://doi.org/10.1038/s41598-018-23362-x (2018).

20. Tang, L. *et al*. Density resistance evaluation of maize varieties through new "Density–Yield Model" and quantification of varietal response to gradual planting density pressure. *Sci Rep* **8**(1), 17281, https://doi.org/10.1038/s41598-018-35275-w (2018).

21. Sarawgi, A., Rastogi, N. & Soni, D. Correlation and path analysis in rice accessions from Madhya Pradesh. *Field Crops Res.* **52**, 161–167, https://doi.org/10.1016/S0378-4290(96)01061-1 (1997).

22. Güler, M., Adak, M. S. & Ulukan, H. Determining relationships among yield and some yield components using path coefficient analysis in chickpea (*Cicer arietinum* L.). *Eur J Agrone* **14**, 161–166, https://doi.org/10.1016/S1161-0301(00)00086-1 (2001).

23. Tahir, M., Sadaqat, H. A. & Bashir, S. Correlation and path coefficient analysis of morphological traits in sunflower (*Helianthus annuus* L.) populations. *Int J Agri Biol.* **4**, 341–343, https://doi.org/10.1016/j.ancard.2009.10.002 (2002).

24. Mahmood, T. Path coefficient analysis of yield component in tomato (*Lycopersicon esculentum*). *Pak J Bot.* **40**, 627–635, https://doi.org/10.1094/MPMI-21-4-0490 (2008).

25. Arshad, M., Khan, M. A., Jadoon, S. & Mohmand, A. S. Factor analysis in sunflower (*Helianthus annuus* L.) to investigate desirable hybrids. *Pak J Bot.* **42**, 4393–4402, https://doi.org/10.1094/MPMI-07-10-0151 (2010).

26. Khan, A., Azam, F. & Ali, A. Relationship of morphological traits and grain yield in recombinant inbred wheat lines grown under drought conditions. *Pak J Bot.* **42**, 259–267, https://doi.org/10.1127/1438-9134/2010/0136-0325 (2010).

27. Ma, N. *et al*. Ideotype population exploration: growth, photosynthesis, and yield components at different planting densities in winter Oilseed Rape (*Brassica napus* L.). *PloS one* **9**, 25–36,, https://doi.org/10.1371/journal.pone.0114232 (2014).

28. Zhong, C. *et al*. Nitrogen metabolism in adaptation of photosynthesis to water stress in rice grown under different nitrogen levels. *Front Plant Sci.* **8**, 1079–1088, https://doi.org/10.3389/fpls.2017.01079 (2017).

29. Zhang, H. *et al*. Rational water and nitrogen management improves root growth, increases yield and maintains water use efficiency of cotton under mulch drip irrigation. *Front Plant Sci.* **8**, 912–920,, https://doi.org/10.3389/fpls.2017.00912 (2017b).

30. Bondalapati, K. D., Jenkins, J. N., Mccarty, J. C. & Wu, J. Field experimental design comparisons to detect field effects associated with agronomic traits in upland cotton. *Euphytica* **206**, 747–757, https://doi.org/10.1007/s10681-015-1512-2 (2015).

31. Rana, J. C. *et al*. Characterisation of 4274 accessions of common bean (*Phaseolus vulgaris* L.) germplasm conserved in the Indian gene bank for phenological, morphological and agricultural traits. *Euphytica* **205**, 441–457, https://doi.org/10.1007/s10681-015-1406-3 (2015).

32. Ahumada-Orellana, L. E., Ortega-Farías, S., Searles, P. S. & Retamales, J. B. Yield and water productivity responses to irrigation cut-off strategies after fruit set using stem water potential thresholds in a super-high density Olive Orchard. *Front Plant Sci.* **8**, 1280–1291, https://doi.org/10.3389/fpls.2017.01280 (2017).

33. Dewey, D. R. & Lu, K. A correlation and path-coefficient analysis of components of crested wheatgrass seed production. *Agron J.* **51**, 515–518, https://doi.org/10.2134/agronj1959.00021962005100090002x (1959).

34. Topal, M., Ozturk, E. & Polat, T. Path analysis of seed yield components using different correlation coefficients in safflower (*Carthamus tinctorius* L.). *J Anim Plant Sci.* **20**, 220–224, http://www.thejaps.org.pk/docs/20-04-2010/10-040-Revised.pdf (2010).

35. Alvi, M. B. *et al*. Character association and path coefficient analysis of grain yield and yield components maize (*Zea mays* L.). *Pak J Biol Sci.* **6**, 136–138, https://doi.org/10.3923/pjbs.2003.136.138 (2003).

36. Asghari-Zakaria, R., Fathi, M. & Hasan-Panah, D. Sequential path analysis of yield components in potato. *Potato Res.* **49**, 273–279, https://doi.org/10.1007/s11540-007-9022-9 (2006).

37. Cruz, C. D. Programa GENES: Aplicativo computacional em estatística aplicada à genética (GENES-Software for Experimental Statistics in Genetics). *Genet Mol Biol.* **21**, 148–156,, https://doi.org/10.1590/S1415-47571998000100022 (1998).

38. Zhang, H., Berger, J. D. & Herrmann, C. Yield stability and adaptability of canola (*Brassica napus* L.) in multiple environment trials. *Euphytica* **213**, 115–123,, https://doi.org/10.1007/s10681-017-1948-7 (2017a).

39. Saleem, M. *et al*. Interrelationships and path analysis of yield attributes in Chick Pea (*Cicer arietinum* L.). *Int J Agric. Biol.* **3**, 404–406, https://doi.org/10.11648/j.jps.20140202.13 (2002).

40. Kuai, J. *et al*. The yield of mechanically harvested rapeseed (*Brassica napus* L.) can be increased by optimum plant density and row spacing. *Sci Rep.* **5**, 18835, https://doi.org/10.1038/srep18835 (2015).

41. Fairey, N. A. & Lefkovitch, L. P. Crop density and seed production of tall fescue (*Festuca arundinacea* Schreber) yield and plant development. *Can J Pan Sci.* **79**, 535-541, Available online at, https://www.nrcresearchpress.com/doi/pdf/10.4141/P98-112 (1999).

42. Wang, Q. *et al*. Effects of interaction between nitrogen and plant population on seed yield of *Dactylis glomerata*. *Acta Prata Sinica* **22**, 38–44, https://doi.org/10.3969/j.issn.1009-3575.2005.02.003 (2005a).

43. Sun, T., Han, J., Zhao, S. & Yue, W. Effects of fertilizer application on seed yield and yield components of *Psathyrostachys juncea*. *Acta Agrestia Sinica* **27**, 16–21, https://doi.org/1007 4333(2005)03 0015 06 (2005).

44. Wang, Z. *et al*. Segregation analysis of rice seed germination under cold stress using major gene plus polygene mixed inheritance model. *Seed Sci Technol.* **38**, 104–113, https://doi.org/10.15258/sst.2010.38.1.11 (2010).

45. Meints, P. D. *et al*. Stubble management effects on three creeping red fescue cultivars grown for seed production. *Agron J.* **93**, 1276–1281, https://doi.org/10.2134/agronj2001.1276 (2001).

46. Lopes, R. R. & Franke, L. B. Path analysis in white clover seed yield components. *Rev Bras Zoote* **38**, 1865–1869, https://doi.org/10.1590/S1516-35982009001000002 (2009).

47. Hoerl, A. E. & Kennard, R. W. Ridge regression: biased estimation for non-orthogonal problem. *Technometrics* **12**, 55–67, https://doi.org/10.2307/1271436 (1970b).

48. Newell, G. J. Ridge regression: an alternative to multiple linear regression for highly correlated data. *J Food Sci.* **46**, 968–969, https://doi.org/10.1111/j.1365-2621.1981.tb15400.x (1981).

49. Chatterjee, S. & Hadi, A. S. Regression analysis by example. John Wiley & Sons. New York, https://doi.org/10.1080/02664763.2013.817041 (2015).

50. Poelman, M. E., Pilmanis, A. M. & Hufford, K. M. Testing the cultivar vigor hypothesis: comparisons of the competitive ability of wild and cultivated populations of *Pascopyrum smithii* along a restoration chronosequence. *Restor Ecol.* **27**(1), 92–101, https://doi.org/10.1111/rec.12822 (2019).

51. Liu, Y. *et al*. Synergistic and antagonistic effects of salinity and pH on germination in switchgrass (*Panicum virgatum* L.). *PloS one* **9**, e85282, https://doi.org/10.1371/journal.pone.0085282 (2014).

52. Rondeaux, G., Steven, M. & Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens Environ.* **55**, 95–107, https://doi.org/10.1016/0034-4257(95)00186-7 (1996).

53. Montgomery, D. C. Design and Analysis of Experiments, Student Solutions Manual, 8th Edition, https://doi.org/10.2113/gseegeosci.xxviii.3.303 (2012).

54. Zhang, G. & Li, Y. Orthogonal experimental design method used in particle swarm optimization for multimodal problems. In: Sixth International Conference on Advanced Computational Intelligence **5**, 183–188, https://doi.org/10.1109/ICACI.2013.6748498 (2014).
55. Schwing, R. C. & Mcdonald, G. C. Measures of association of some air pollutants, natural ionizing radiation and cigarette smoking with mortality rates. *Sci Total Environ.* **5**, 139–146, https://doi.org/10.1016/0048-9697(76)90052-8 (1976).

Acknowledgements

Author contributions

Q.W. designed and supervised the research project. Q.Z. performed the experiments and collected the data. J.C. helped in data collection. Z.C. and J.N. analyzed the data and wrote the manuscript. X.C. and W.J. revised and edited the manuscript and also provided advice on the experiment. Q.W. read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-54586-0.

Correspondence and requests for materials should be addressed to Q.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.