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

Zhao Chen1, Junpeng Niu1, Xinlong Cao1, Wenbo Jiang1, Jian Cui2, Quanzhen Wang1* & Quan Zhang3

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 (Y₁), spikelets/fertile tiller (Y₂), florets/spikelet (Y₃), seed numbers/spikelet (Y₄), and seed weight (Y₅) on seed yield (Z) were determined over three successive years. Correlation analysis indicated that fertile tillers (Y₁) was the most important seed yield component. And the biggest contribution of those five yield component is fertile tillers (Y₁), followed by seed numbers/spikelet (Y₄), spikelets/fertile tiller (Y₂), florets/spikelet (Y₃) and seed weight (Y₅), 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 world1. Because it thrives on impoverished soils in pastoral environments, even with multiple, simultaneous stressors2, 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 China3–6. Thus, it can appear as a dense monoculture7,8. 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 capacity9,10.

There have been many studies on the factors affecting crop yields, with the aim of improving yields as much as possible11. Seed yield is a complex trait that is the culmination of growth and developmental processes that are influenced by multiple yield components12,13. Understanding the relationship between yield and yield components and the correlations between yield components is a prerequisite for building an efficient breeding program14. 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 experiments15. 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
yield components compared to seed yield, the use of these traits to practice indirect selection for yield is justified. 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. 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. reported correlations between yield and planting density, and other yield components. Tang et al. 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. Thus, plant morphological traits such as fertile tillers m⁻² (Y₁), spikelets/fertile tillers (Y₂), florets/spikelet (Y₃), seed numbers/spikelet (Y₄), or seed weight (Y₅), 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, the experimenter selects the best design based on the information available with respect to field spatial heterogeneity. 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 (Y₁), spikelets/fertile tillers (Y₂), florets/spikelet (Y₃), seed numbers/spikelet (Y₄), and seed weight (Y₅), on seed yield (Z) of Paspopyrum 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 = Y₁ × Y₂ × Y₃ × Y₄ × Y₅. 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 Y₁ and Y₂, but was negatively affected to a lesser extent by Y₃ and Y₄ (P < 0.05 or P > 0.01) (Table 1). Y₁ had the maximum coefficient on Z (0.472, P < 0.001). The correlations among Y₁ through Y₄ showed significance (P < 0.001), while Y₁ and Y₂ were negatively correlated with Y₃ and Y₅. However, Y₁ and Y₂ showed strong positive correlations with Z in all three years (P < 0.001) (Table 2). The order of the strength of correlation coefficients in each year was 2004 > 2005 > 2003. Y₁ had the highest positive influence on Z (0.452, 0.534, and 0.657 for 2003, 2004 and 2005, respectively), and Y₅ had the strongest negative influence on Z (−0.266, 0.371 at 0.05, and −0.391 at 0.05) and Y₄ in 2003 (−0.480, 0.423 and 0.777, respectively). Therefore, Y₁ had the largest contribution to Z in 2004 (−0.605 at P < 0.001), Y₅ in 2003 (0.454 at P < 0.05) and Y₄ in 2003 (−0.371 at P < 0.05) had weak but statistically significant direct effects on Z.

| Seed yield components | Y₁   | Y₂   | Y₃   | Y₄   | Y₅   | Z(seed yield) |
|-----------------------|------|------|------|------|------|--------------|
| Y₁                    | 1.0000 | 0.118** | −0.056** | −0.071** | −0.286** | 0.472** |
| Y₂                    | 1.0000 | 0.0000 | 0.105** | 0.031** | −0.104** | 0.116** |
| Y₃                    | 1.0000 | 0.105** | 0.155** | 0.023 | 0.030* |
| Y₄                    | 1.0000 | 0.064** | −0.019** |
| Y₅                    | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

Table 1. Matrix of Pearson correlation coefficients of Y₁–Y₅, Z (Paspopyrum smithii Schreb.) averaged over 3 years. **Significant at the 0.01 probability level. *Significant at the 0.05 probability level. N = 380.

### Path analyses of Y₁ to Y₅ 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 Y₁ on Z was strong and positive (highlighted in bold) for all 3 years (2003, 2004 and 2005) at P < 0.001, where the coefficients were 0.480, 0.423 and 0.777, respectively. Therefore, Y₁ had the largest contribution to Z. Y₂ in 2004 (−0.065 at P < 0.001), Y₃ in 2003 (0.454 at P < 0.05) and Y₄ in 2003 (−0.371 at P < 0.05) had weak but statistically significant direct effects on Z. Regarding the contributions of Y₁–Y₄ to Z, the strongest positive indirect influence was presented by the pathway from Y₁ via Y₂ (the coefficient was 0.1740 in 2003), and the second indirect effect was demonstrated by Y₃ through Y₁ (0.1039 in 2003). In order of decreasing magnitude, Y₄ via Y₃ (0.0582 in 2005) and Y₃ via Y₅ (0.0447 in 2003) were observed. The strongest negative indirect influence on Z was Y₃ via Y₄ (−0.0799 in 2003), and the second indirect effect was demonstrated by Y₅ through Y₄ (−0.0517 in 2005). Y₅ had a negative and marginal direct influence on yield in 2004 and positive effects during the other 2 years on seed yield (Table 3). Y₃ had the smallest contribution to Z (Table 3). In summary, the order of contributions of the five seed yield components were Y₁ > Y₂ > Y₃ > Y₄ > Y₅, and the total direct effects were 1.68, 0.529, 0.195, 0.077 and −0.266, respectively.
and 2005, the curves for Y1 to Y5 showed estimated k values at 0.77, 0.64 and 0.79, respectively (Fig. 1). The ridge suggested that the value of the ridge parameter K1 should be determined using ridge traces (Fig. 1). For 2003, 2004 and 2005, the curves for Y1 to Y5 showed estimated k values at 0.77, 0.64 and 0.79, respectively (Fig. 1). The ridge suggested that the value of the ridge parameter K1 should be determined using ridge traces (Fig. 1). For 2003, 2004 and 2005, the curves for Y1 to Y5 showed estimated k values at 0.77, 0.64 and 0.79, respectively (Fig. 1). The ridge suggested that the value of the ridge parameter K1 should be determined using ridge traces (Fig. 1).

Table 2. Matrix of Pearson correlation coefficients of Y1~Y5, 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.

| Year | Y1   | Y2   | Y3   | Y4   | Y5   | Z    |
|------|------|------|------|------|------|------|
| 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 |
| 2006 | 0.061*** | 0.0003 | -0.0002 | 0.0064 |
| 2007 | 0.046 | -0.0367 | -0.0294 | -0.0367 |
| 2008 | 0.454** | -0.2920 | 0.0062 |
| 2009 | -0.006 | 0.0004 | -0.0083 |
| 2010 | 0.081 | 0.0582 | 0.0447 |
| 2011 | -0.371* | 0.0129 |
| 2012 | 0.005 | -0.0034 |
| 2013 | 0.100 | 0.0280 |
| 2014 | 0.055 |
| 2015 | -0.063** |
| 2016 | 0.087 |

Table 3. Path analysis showing direct and indirect effect of Y1−Y5 to Z (Pascopyrum smithii Schreb.). *Significant at the 0.05 probability level. **Significant at the 0.01 probability level. The direct effects of Y1−Y5 to Z are highlighted in bold (on the main diagonal cells); Arrows indicate the direction of the effect.

| Year | Indirect effect via | Total direct effect |
|------|--------------------|---------------------|
|      | → Y1→Z            | → Y2→Z             | → Y3→Z             | → Y4→Z             | → Y5→Z             |
| 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             |
| 2006 | 0.061***           | 0.0003              | -0.0002             | 0.0064              |
| 2007 | 0.046              | -0.0367             | -0.0294             | -0.0367             |
| 2008 | 0.454**            | -0.2920             | 0.0062              |
| 2009 | -0.006             | 0.0004              | -0.0083             |
| 2010 | 0.081              | 0.0582              | 0.0447              |
| 2011 | -0.371*            | 0.0129              |
| 2012 | 0.005              | -0.0034             |
| 2013 | 0.100              | 0.0280              |
| 2014 | 0.055              |
| 2015 | -0.063**           |
| 2016 | 0.087              |

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). Y1 was highest in 2003 and produced the highest Z. It has been suggested that the value of the ridge parameter K1 should be determined using ridge traces (Fig. 1). For 2003, 2004 and 2005, the curves for Y1 to Y5 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 × Y1 + 19.944 × Y2 + 15.127 × Y3 - 55.349 × Y4 - 29.703 × Y5
   (Ridge k = 0.77; F = 8.274 Pr < 0.001)

B. Z = 444.094 + 0.358 × Y1 + 2.075 × Y2 - 2.205 × Y3 - 0.920 × Y4 - 4.078 × Y5
   (Ridge k = 0.64; F = 114.768 Pr < 0.001)

C. Z = 271.685 + 0.381 × Y1 + 14.551 × Y2 - 7.485 × Y3 + 18.647 × Y4 - 48.165 × Y5
   (Ridge k = 0.79; F = 10.797 Pr < 0.001)
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).

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.

| year | $N$  | $Y_1$   | $Y_2$   | $Y_3$   | $Y_4$   | $Y_5$   | $Z$  |
|------|------|---------|---------|---------|---------|---------|------|
| 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   | 19.944| 15.127| −55.349| −29.703| −1   |
| 0.64 | 2004 | 444.094   | 2.075 | −2.205| −0.920| −4.078| −1   |
| 0.79 | 2005 | 271.685   | 14.551| −7.485| 18.647| −48.165| −1   |

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

![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.](image)

**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.
Thus, \[ Z = S = -230.174 + 284.235 \times Y_1^{0.21} \times Y_2^{0.1} \times Y_3^{0.01} \times Y_4^{-0.01} \times Y_5^{-0.30} \] of the 3 years. It is superimposed on the 1:1 line.

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

\[ Z_{\text{actual}} = -230.174 + 1.568 \times Z_{\text{estimated}} \]  

(4)

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

\[ Z_{\text{actual}} = -230.174 + 284.235 \times Y_1^{0.21} \times Y_2^{0.1} \times Y_3^{0.01} \times Y_4^{-0.01} \times Y_5^{-0.30} \]  

(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 \( Y \)s 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).

### Table 6. Parameter estimates for \( Z_{\text{estimated}} \).

| Variable  | DF | Parameter estimate | Standard error | t value | Pr > |t| |
|:---------|:---|:------------------|:--------------|:--------|:-----|---|
| Intercept| 1  | −230.174          | 24.195        | −9.513  | <0.0001 |
| \( Z_{\text{estimated}} \) | 1  | 1.568             | 0.048         | 32.357  | <0.0001 |

### Table 7. Parameter estimates for \( Z_{\text{estimated}} \) after adjustment by linear regression.

| Variable  | DF | Parameter estimate | Standard error | t value | Pr > |t| |
|:---------|:---|:------------------|:--------------|:--------|:-----|---|
| Intercept| 1  | −0.033            | 17.164        | −0.002  | 0.998 |
| \( Z_{\text{estimated}} \) | 1  | 1.000             | 0.031         | 32.357  | <0.0001 |

### Figure 2. Scatter plot to fit regression line of actual and estimated seed yield adjusted by \( Z_{\text{actual}} = -230.174 + 284.235 \times Y_1^{0.21} \times Y_2^{0.1} \times Y_3^{0.01} \times Y_4^{-0.01} \times Y_5^{-0.30} \) of the 3 years. It is superimposed on the 1:1 line.
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 (Y1, Y2, Y3, Y4, and Y5) 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 Y1 (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, Y1 and Y4 weakly decreased along with age from 2003 to 2005, which may have been due to genetic factors, while larger difference indicate the role of environment30,31. Fertile tillers m−2 (Y1), florets/spikelet (Y3), and seed numbers/spikelet (Y4) were higher in 2003 compared to 2004 and 2005, whereas spikelets/fertile tillers (Y2) was lower in 2003 than in 2004 and 2005. Seed weight (Y5), 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 importance32. 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 interest33. Alvi et al.34 and Asghari-zakaria et al.35 used path analysis to enable the identification

---

**Figure 3.** Ridgelines of the response surface models showed the synergism and antagonism through Y4 (A) Y3 (B) and Y4 (C) to Y2 and Y3 (D) Y2 (E) and Y1 (F) to Y5.
Figure 4. Ridgelines of the response surface models showed the synergism and antagonism through $Y_1$ to $Y_5$ (A) and $Y_2$ (B), $Y_3$ (C) and $Y_1$ (D) to $Y_5$.

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.42 defines the path coefficient, or cause and effect analysis, as a standardized regression coefficient24, 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 yield45. In this study, path analysis indicates that the total direct effects of Y1, Y2, Y3, and Y4; and Y2, Y3, and Y5 have highly significant positive correlations with Z, but Z is negatively correlated with Y4 (Table 3). The explanation for why there is a weak negative correlation between the seed number (Y1) and seed yield (Z) might matter to a high density cultivation and soil nutrient limitation7. 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 spacing49. Table 3 shows that Y1 and Y2 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 Y1 (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 yearly46. However, Y3 did not significantly contribute to Z partly because Y3 was mostly under genetic control40. This discovery implies that Y3 is the sub group that should be taken into account if high quality forage is the target of the breeding system. Nevertheless, Y1 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 fescues41,42, Russian wild rye43,44, Zoysia grass45,46,47,1, gramineous plants5, and white clover40,44,45. Additionally, path analysis uncovered relationships between the components and the yield that are consistent with previously reported results55. The interrelationships among Y1 to Y5 indicated that Y4 and Y5 had remarkable negative relationships with seed yield (Z; r = 0.038* and r = −0.192*). Y1 and Y4 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, Y1, Y2, and Y3 decreased significantly, whereas Y2 and Y3 increased. This finding is consistent with the results of previous research40. This result also implies that Y4 and Y5 should and could be effectively improved if the values of Y1, Y2, and Y3 are lower than normal.

Ridge regression and multiple regression analyses were applied to avoid high inter-correlation and multico-linearity among the variables46–48. 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 reliable46. In addition, ridge regression effectively overcomes the problem of highly multi-correlated predictor variables (seed yield components)45,46. This method is the most effective and practical for the current field scientific research41. 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 Y1 through Y5. 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: Y1 > Y2 > Y3 > Y4 > Y5 (1.68, 0.529, −0.266, 0.195 and 0.077, Table 3). The total direct effects with Y4 having negative results, showed that the total influence order, in absolute values, is Y1 > Y2 > Y3 > Y4 > Y5 (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 Y1 and Y2, among the ten relation schema subgraphs (Figs. 3 and 4), as did the lines for 2005 (green), except for Y3 and Y4. The subgraphs indicated that Y1 and Y2 (Fig. 3A), Y1 and Y3 (Fig. 3C), Y1 and Y4 (Fig. 3D), Y1 and Y5 (Fig. 4B) and Y1 and Y3 (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 Y2 and Y3, Y4 and Y5 and Y1 and Y2 had synergistic effects on Z (Fig. 3B,E,F) at k = 0 in the ridgelines. The more Y2, less Y3 dynamic (Fig. 3C), which was also evident for Y3 and Y1 (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 Y3 in a certain range of growth. In this range, Y3 has the optimal value. The increase in Y1-derived Y2 and the Y3 decrease (Fig. 4B,D) were consistent with the soil nutrient limitation. For 2003, 2004, 2005 and 2003–2005, the interactions between Y1, Y2 and Y3 gradually changed from synergetic to antagonistic (Fig. 4C). As important seed yield traits of the plant, Y2 and Y3 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 (Y2) and fertile tillers m⁻² (Y3), and negative correlations were found with seed number/spikelet (Y4). The inter-correlation among these components were significant in 2004 and 2005.
be focused on changes in seed yield based on different climatic conditions and site locations. This model can be used to accurately estimate the seed yield with five yield components. Further studies should focus on changes in seed yield based on different climatic conditions and site locations.

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

| Year | Y_1Y_2 | Y_1Y_3 | Y_1Y_4 | Y_1Y_5 | Y_2Y_4 | Y_2Y_5 | Y_3Y_4 | Y_3Y_5 | Y_4Y_5 |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 2003 | 0.001  | 0.003  | 0.003  | 0.003  | -0.276 | 4.505  | -622.169 | 4490.5 | 10.30  | <0.0001 |
| N=146 | 0.0003 | 4.719  | 0.035  | -0.054 | -119.522 | 1182.2 | 10.63   | <0.0001 |
| 2004 | -2.952 | 60.69  | -2.35  | 960.74 | 1120.8  | 9.91  | <0.0001 |
| N=4221 | 0.094 | 1.96  | 0.902  | 0.597  | -1719.56 | 5747.4 | 10.09  | <0.0001 |
| 2005 | -0.001 | 0.322 | -0.14  | 1.545  | 19.74  | 76.6  | 177.85  | <0.0001 |
| N=105 | -0.001 | 0.012 | 0.012  | 0.012  | 0.1419 | 345.5 | 173.16  | <0.0001 |
| 2006 | -0.001 | 2.985 | 0.001  | 1.332  | 33.41  | 236.4 | 172.92  | <0.0001 |
| N=902 | 0.001  | 0.145 | 0.388  | -0.583 | -114.70 | 863.0 | 187.95  | <0.0001 |
| 2007 | -0.545 | -0.064 | 0.864  | 20.767 | -13.58 | 363.1 | 15.89   | <0.0001 |
| N=796 | 0.031  | -3.36  | -0.435 | 28.73  | 42.79  | 145.0 | 13.14   | <0.0001 |
| 2008 | -0.45  | 2.228  | 2.483  | 11.153 | -155.42 | 969.0 | 69.62   | <0.0001 |
| N=622 | 0.063  | -3.082 | -0.153 | 0.997  | 32.46  | 486.3 | 6.54    | <0.0001 |
| 2009 | 0.580  | 2.554  | -3.852 | 12.41  | -91.55 | 931.4 | 64.87   | <0.0001 |
| N=516 | -3.292 | 2.408  | -1.797 | 42.35  | -109.45 | 932.2 | 64.15   | <0.0001 |
| 2010 | 0.001  | -20.61 | 0.049  | -0.093 | 773.29 | -694.8 | 16.11   | <0.0001 |
| N=4426 | 0.001 | -1.135 | 0.035  | 1.456  | 178.76 | -1293.1 | 19.95   | <0.0001 |
| 2011 | -0.001 | -148.57| 0.003  | 1.752  | 1498.91 | -3674.6 | 18.84   | <0.0001 |
| N=610 | -0.001 | -312.39| -0.776 | 5.363  | 3265.24 | -8357.3 | 15.32   | <0.0001 |
| 2012 | -18.17 | 33.81  | 39.42  | 486.06 | -205.19 | -3504.8 | 8.40    | <0.0001 |
| N=754 | -25.70 | -185.18| -1.496 | 1024.87| 1797.63 | -15779.0 | 5.56    | <0.0001 |
| 2013 | -22.65 | -238.08| -3.591 | 911.35 | 2198.11 | -13273.1 | 7.56    | <0.0001 |
| N=877 | -82.24 | -148.47| -64.47 | 1500.05 | 1885.23 | -9322.8 | 4.18    | 0.0014 |
| 2014 | -41.82 | -326.69| -51.10 | 900.02 | 3212.32 | -9896.0 | 6.52    | <0.0001 |
| N=825 | -82.85 | -327.26| -303.08| 2198.81 | 4280.04 | -14104.2 | 6.00    | <0.0001 |
| 2015 | 0.0005 | -0.301 | -0.023 | 1.073  | 21.52  | 109.6 | 214.62  | <0.0001 |
| N=4426 | -0.001 | 0.823  | 0.059  | 0.485  | -21.42 | 479.1 | 227.53  | <0.0001 |
| 2016 | 0.0004 | -2.454 | 0.045  | 0.490  | 16.60  | 374.4 | 212.55  | <0.0001 |
| N=610 | -0.0003 | 0.224 | 0.557  | -1.886 | -166.30 | 1161.5 | 237.93  | <0.0001 |
| 2017 | -0.648 | 4.445  | 0.602  | 27.02  | -69.91 | 483.1 | 26.47   | <0.0001 |
| N=754 | -0.661 | -4.380 | -0.699 | 35.11  | 41.48  | 42.3 | 13.73   | <0.0001 |
| 2018 | -0.613 | 1.581  | 2.194  | 19.10  | -118.23 | 759.8 | 57.30   | <0.0001 |
| N=877 | 4.169 | -3.831 | 1.110  | 60.67  | 31.19  | 691.9 | 16.31   | <0.0001 |
| 2019 | 4.995 | 2.425  | -2.908 | -51.60 | -92.83 | 1133.8 | 73.72   | <0.0001 |
| N=825 | -4.448 | 1.841  | -1.059 | 54.13  | -83.82 | 795.4 | 51.98   | <0.0001 |

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

$$Z = -230.174 + 284.235 \times Y_1^{0.21} \times Y_2^{0.3} \times Y_3^{0.01} \times Y_4^{0.01} \times Y_5^{0.30}$$

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 (Y_4) > 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 (Y_4) 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. The plots used in this experiment were planted with alfalfa (*Medicago sativa* L.) in the preceding season. *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 P₂O₅. There was no seed yield in autumn 2002. Initial chemical characteristics of the soil (0–20 cm) were: pH = 8.39; NH₄⁺ = 32.32 mg kg⁻¹; NO₃⁻ = 20.09 mg kg⁻¹; alkali hydrolysable nitrogen = 118.30 mg kg⁻¹; available phosphor = 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. Thus, to simulate various growing conditions, we used six groups (A to E) (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), giving a total of 380 experimental plots (each with an area of 28 m²), under various field management treatments (Table S2) within a total field area of 4100 m². 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 (*Y₁*), 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 (*Y₂*), florets/spikelet (*Y₃*) and seed numbers/spikelet (*Y₄*). 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 ± 1% for converting into seed yield (kg ha⁻¹) (Z). Ten lots of 100 seeds each were collected to determine mg seed weight (*Y₅*). The total number of samples (n) used to measure *Y₁* to *Y₅* and Z were 3800, 13605, 11085, 10770, 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. 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. 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 and Duncan's multiple range tests for seed yield (Z) and yield components (*Y₁*–*Y₅*) were performed. The data were transformed using logarithmic and power transformations to avoid the effects of highly inter-correlated data, which would lead to multi-co-linearity between *Y₁*–*Y₅* and Z. To establish a reliable model, all of the Z and *Y₁*–*Y₅* 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 $S = \ln Z$ and $C_i = \ln Y_i$ (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:

$$ S = C \times \beta + u $$

(6)

where $S$ is the $n \times 1$ vector of observations of a response variable, $C$ is the $n \times p$ matrix of observations on p explanatory variables, $\beta$ is the $p \times 1$ vector of regression coefficients, and $u$ is the $n \times 1$ vector of residuals satisfying $E(u) = 0$; $E(uu') = \Sigma I$.

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,

$$ \ln Z = \left[ \sum_{i=1}^{5} \ln Y_i \right] \times \beta + u $$

(7)

The above logarithmic model (7) was transformed to the following exponential function:
\[ Z = e^\alpha \times \prod_{i=1}^{5} (Y_i^\beta), \]  \hspace{1cm} (8)

where \( \alpha \) and \( \beta \) are constants.

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

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

\[ Z_{\text{actual}} = \beta + K \times Z_{\text{estimated}} \]  \hspace{1cm} (9)

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

\[ Z = \beta + k \times e^\alpha \times \prod_{i=1}^{5} (Y_i^\beta). \]  \hspace{1cm} (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:

\[ Z = \sum_{i=1}^{2} (\beta_{x,y} Y_i^j) + u(i = 1, 2; j = 1, 2), \]  \hspace{1cm} (11)

Where \( \beta \) is a constant.

The equivalent effects of \( Y_i \) and \( Y_j \) were determined using:

\[ \frac{\partial Z}{\partial Y_i} = \frac{\partial Z}{\partial Y_j} \]  \hspace{1cm} (12)

which produced

\[ Y_j = k \times Y_i + b, \]  \hspace{1cm} (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.

Received: 6 August 2019; Accepted: 15 November 2019;
Published online: 29 November 2019
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**

We would like to thank team colleagues for their assistance with the field sampling. National Key R&D Program of China (2017YFE0111000) and the National Natural Science Foundation of China (31472138) funded this work.

**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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019