

Use of Factor Analysis in Grain Sorghum Breeding

Viktoria Starchak^{1,*}, Denis Stepanchenko¹ and Evgeniy Matvienko²

¹ Federal State Budgetary Scientific Institution Russian Research and Design Technological Institute of Sorghum and Corn Rossorgo, 410050, Saratov Region, Saratov, 1st Institutskiy proezd, 4, Russian Federation

² Volga Research Institute of Breeding and Seed Production named after P.N. Konstantinov, branch SamNTsRAN, 446442, Samara region, Kinel, urban-type settlement Ust-Kinelsky, 76, Shosseinaya street, Russian Federation

Abstract. The article presents the results of calculating factor loads for 2015-2018 obtained by the method of principal components of grain sorghum varieties. Field studies were laid and carried out in the fields of the Federal State Budgetary Scientific Institution RosNIISK Rossorgo. The objects of the study were grain sorghum plants of the Institute's selection, as well as samples presented by the VIR world collection. 25 economically valuable indicators in 32 samples were analyzed. The main role in selection of indicators for inclusion in the correlation model is assigned to the expert approach in assessing breeding material. Effective and combinational groupings of correlation coefficients by hypothetical factors are very important tools in solving the issue of selecting indicators in a correlation model of interrelationships of selectively valuable indicators of various plant populations. Grouping the correlation coefficients makes it possible to assess the general nature of the interaction of the indicators included in the experimental data matrix. Calculations of the loads on hypothetical factors in the experiment made it possible to identify factors whose contribution to the accumulated variance is more than 5 %.

1 Introduction

Grain sorghum is a valuable agricultural crop. Sorghum is cultivated for both feed production and food purposes. Production of feed and its quality largely determines the efficiency and level of the livestock industry development. In limited quantities, grain sorghum is grown for food purposes. The studied culture has a number of advantages over other crops, including drought resistance and heat resistance. Sorghum is an important insurance crop in case of summer drought, and most often has an advantage when used as a stubble or post-harvest crop [1]. Due to biological and physiological characteristics, sorghum crops are adapted to arid growing conditions [2,3]. The great potential and capabilities of the sorghum crop allows to get stable, high yields of grain and green mass. This places the sorghum crop among the leading grain feed and food crops.

* Corresponding author: viktoria_starchak@rambler.ru

Currently, the priority direction in breeding is creation of heterotic hybrids based on CMS. The high yields of the sorghum crop are the great advantage in the arid regions of the country for the stable production of feed grains. F1 hybrids are clearly inferior to varieties, being high intensity, due to hybrid power. Therefore, the problem of increasing economic efficiency in the context of increasing intensification of agriculture is of paramount importance.

An integrated (systemic) approach and assessment of general combining ability and specific combining ability, true and hypothetical heterosis in test crosses of vegetative, generative indicators and biochemical composition of sorghum grain, as well as factorial and cluster analysis of the model population should be considered as a necessary factor when adjusting and optimizing breeding programs for creating varieties and hybrids of grain sorghum, allowing the design of adaptive agroecosystems in conditions of insufficient moisture [4-6].

The correct selection of parental forms and analysis of the source material is one of the important criteria when creating highly heterotic hybrids. Some researchers note the ameliorating role of grain sorghum when sowing on saline soils [7].

The strategy of modern breeding of grain sorghum for the conditions of arid regions of Russia shall provide for creation of a completely new source material – early ripening, highly productive self-pollinated lines and varieties characterized by valuable morphological indicators and biological properties of plants that are highly adaptable and resistant to biological and abiotic stresses.

In the process of breeding work, considerable attention is paid to formation of a model of varieties and hybrids for specific purposes. In this regard, formation of model populations of grain sorghum and the study of the totality of interrelationships of morphophysiological parameters are of priority importance in comparison with visual observations and statement of actual values. When analyzing large data matrices, it becomes necessary to optimize the analysis of correlations.

The main role in selection of indicators for inclusion in the correlation model is assigned to the expert approach in assessing breeding material. Effective and combinational groupings of correlation coefficients by hypothetical factors are very important tools in solving the issue of selecting indicators in the correlation model of interrelationships of morphophysiological characteristics of various plant populations. Grouping the correlation coefficients makes it possible to assess the general nature of the interaction of the indicators included in the experimental data matrix.

The limiting factor for the wide spread of this crop is only an underdeveloped livestock base and insufficient work to create highly heterotic hybrids and sorghum varieties that meet production requirements.

The study of biochemical parameters characterizing the quality of sorghum grain is a prerequisite for effective breeding work. It was revealed that the biochemical composition of grain depends on many factors: genotype, weather and environmental conditions, the number of plants per unit area, the level of technology, including the integral system of plant protection from diseases and pests [8].

2 Methods

The objects of study are 32 varieties and breeding lines of Federal State Funded Research Institution RosNIISK Rossorgo, most of which are approved for use [9] - Perspektivnyy 1, Merkuriiy, Ogonek, Kremovoye, Zenit, Start, Azart, Topaz, RSK Oniks, Vostorg, Granat, Kamelik, Assistent, M-60887, V-03-3003, L-214, Volzhskoye 4, Volzhskoye 44, Volzhskoye 615, Pishchevoye 35, Pishchevoye 614, Sarmat, Avans, Fakel, Geleofor, L.infiniti, Bogdan, L 251, Magistr, 06-2198, Kafrrskoye beloye 127, K-266 Feterita which

were sown on the experimental field of the Federal State Budgetary Scientific Institution RosNIISK Rossorgo in 2015-2018, in a wide-row method (with a row spacing of 70 cm). The plot area was 7.7 m². Repetition – three times. Placement of the plots is randomized. The counts and observations were carried out according to the CMEA broad unified classifier and the CMEA international classifier of cultivated species of the genus *Sorghum* Moench [10]. Statistical processing of study results was carried out using the AGROS 2.09 program – factor analysis, principal component analysis.

Sowing was carried out in the first or second ten days of May with a selective cassette seeder SKS-6-10, the predecessor was pure fallow. Crop care included two inter-row cultivators. In the phase of 3-4 leaves, the number of plants was manually adjusted (100 thousand plants/ha). Harvesting was done by hand.

The height of the plants was measured in the phase of milky-wax ripeness – from the root collar to the top of the panicle of the main stem, the total bushiness was measured by the number of stems extending from the root collar, and the productive bushiness was measured by the number of shoots with mature and ripening panicles. Protrusion of the pedicle on the main stem was measured before harvesting, the length and width of the panicle after flowering. The parameters of the flag and fourth leaves from the top on the main stem were also measured after the flowering of the plant. The sample is 15 plants in each repetition. The above-ground biomass yield and grain yield were calculated using the control sheaf. Sorghum ripens in September-October, depending on the temperature within the grain filling.

The biochemical composition of grain was assessed in the laboratory of biochemistry and biotechnology of the Federal State Budgetary Scientific Institution RosNIISK Rossorgo according to the following methods: protein — according to Kjeldahl (Kjeltec System 2100 device) [11], "raw" fat — according to the Sosklet method [12], ash — according to dry ashing [13], "raw" fiber — in accordance with GOST 13496.2-91. [14], starch — according to GOST 10845-98 [15], nitrogen-free extractable substances were calculated according to generally accepted methods [16].

The experimental field of the Institute of Federal State Funded Research Institution RosNIISK Rossorgo is located in the southern right-bank natural and economic microzone of the Saratov region. The climate of the Saratov region is characterized by a sharp continentality and is formed under the influence of the transfer of air masses from the northwest, north and southeast, which determines the appropriate weather conditions and the manifestation of various natural processes. In general, the climate can be characterized as arid continental, with cold and little snowy winters, short dry springs and hot, long and dry summers. A characteristic feature of the climate in this region is the predominance of cloudy and clear days throughout the year.

For sorghum, frosts are most dangerous in the third decade of May and in the first decade of June, when the seedlings can be completely damaged when the temperature drops to -3 °C, and at +6 °C, a strong suppression of growth and development processes is observed. Also, a decrease in temperature in the autumn period does great damage. Cold August causes a decrease in productivity, September — a decrease in field germination of seeds.

The prevailing southeastern winds bring dry and hot air masses of deserts, and in winter — the cold air of Siberia. The westerly winds bring cyclones. Humid Atlantic air causes a drop in air temperature and precipitation.

The soil is represented by southern chernozem, slightly leached, medium loamy in texture. This type of soil is distinguished by the content of boron, zinc and copper. The upper horizons of all types of soils in the Volga region are well supplied with manganese. However, in the chernozems of southern and dark chestnut soils with signs of solonchicity, its content is higher. The content of humus in the 0-40 cm layer in these chernozems is 3.5-

4.2 %, the exchange capacity is 17-31 mg/eq. per 100 g of soil. The number of nutrient forms available varies widely. The content of hydrolyzable nitrogen is 10 mg, phosphorus – 7.2 mg, exchangeable potassium – 21 mg, calcium – 8 mg per 100 g of soil. The hydrothermal coefficient during the years of the study was: 2015 – 0.41, 2016 – 0.64; 2017 – 0.91, 2018 – 0.69.

3 Results

Based on the analysis of the general characteristics of the variability of the parameters of the model population of grain sorghum, it was established that the following indicators shall be attributed to highly varying indicators ($V > 20.0\%$): the area of the flag leaf, the area of the largest leaf, the protrusion of the panicle stem, total bushiness, productive bushiness, panicle width, number of grains from 1 panicle, grain weight from 1 panicle (Table 1). The rest of the indicators refer to average varying signs ($20.0\% > V > 10.0\%$). The signs with a weak value of variation ($V < 10.0\%$) include biochemical parameters, namely, the content of protein, fat, ash, nitrogen-free extractable substances, starch in the grain.

Table 1. General characteristics of the variability of the parameters of the model population of grain sorghum, 2015-2018.

| № | Parameter | \bar{x} | S^2 | $S_{\bar{x}}$ | V, % |
|----|--------------------------------------|-----------|-----------|---------------|-------|
| 1 | Plant height after 30 days, cm | 50.48 | 37.96 | 6.16 | 12.20 |
| 2 | Plant height when ripe, cm | 127.26 | 412.23 | 20.31 | 15.96 |
| 3 | Flag leaf length, cm | 26.25 | 17.64 | 4.20 | 16.00 |
| 4 | Flag leaf width, cm | 3.75 | 0.25 | 0.50 | 13.32 |
| 5 | Flag leaf area, cm ² | 75.40 | 342.28 | 18.50 | 24.54 |
| 6 | Length of the largest leaf, cm | 45.58 | 47.09 | 6.86 | 15.06 |
| 7 | Largest leaf width, cm | 4.96 | 0.82 | 0.91 | 18.29 |
| 8 | Largest leaf area, cm ² | 175.29 | 3,083.52 | 55.53 | 31.68 |
| 9 | Upper internode thickness, cm | 0.66 | 0.01 | 0.12 | 17.47 |
| 10 | Lower internode thickness, cm | 1.41 | 0.08 | 0.28 | 19.94 |
| 11 | Panicle leg extension, cm | 16.82 | 35.31 | 5.94 | 35.33 |
| 12 | Total bushiness, pcs | 1.52 | 0.12 | 0.35 | 22.97 |
| 13 | Productive bushiness, pcs | 1.65 | 0.31 | 0.56 | 34.00 |
| 14 | Panicle length, cm | 21.54 | 10.57 | 3.25 | 15.10 |
| 15 | Panicle width, cm | 10.04 | 7.55 | 2.75 | 27.38 |
| 16 | Number of grains from 1 panicle, pcs | 738.11 | 63,743.23 | 252.47 | 34.20 |
| 17 | Weight of 1,000 grains, g | 30.00 | 15.11 | 3.89 | 12.96 |
| 18 | Grain weight from 1 panicle, g | 21.46 | 56.61 | 7.52 | 35.06 |
| 19 | Grain yield, t/ha | 4.74 | 0.35 | 0.59 | 12.56 |
| 20 | Protein content in grain, % | 11.43 | 0.59 | 0.77 | 6.75 |
| 21 | Fat content in grain, % | 3.98 | 0.15 | 0.39 | 9.87 |
| 22 | Ash content in grain, % | 1.64 | 0.02 | 0.14 | 8.66 |
| 23 | Fiber content in grain, % | 2.19 | 0.10 | 0.32 | 14.39 |
| 24 | Nitrogen-free extractable | 80.78 | 1.28 | 1.13 | 1.40 |

| | | | | | |
|----|--------------------------------|-------|------|------|------|
| | substances content in grain, % | | | | |
| 25 | Starch content in grain, % | 74.29 | 2.23 | 1.49 | 2.01 |

Note: \bar{x} – average value, $S_{\bar{x}}$ – average error; S^2 – variance; $V, \%$ – coefficient of variation.

4 Discussion

Correlation analysis of these empirical features made it possible to calculate 300 coefficients. The critical value of the correlation coefficient at the 5 % level was 0.349. The matrix contains 66 significant positive and 51 significant negative values of the correlation coefficients. The grain yield significantly correlated with such characteristics as: plant height after 30 days from germination ($r = -0.364$), flag leaf length ($r = 0.406$), flag leaf width ($r = 0.505$), flag leaf area ($r = 0.545$), length of the largest leaf ($r = 0.403$), width of the largest leaf ($r = 0.355$), area of the largest leaf ($r = 0.369$), productive bushiness ($r = -0.432$), panicle width ($r = 0.396$), number of grains per 1 panicle ($r = 0.460$), grain weight from 1 panicle ($r = 0.506$), protein content in grain ($r = -0.467$), ash content in grain ($r = -0.474$), fiber content in grain ($r = -0.366$), nitrogen-free extractable substances content in grain ($r = 0.542$). The grain weight from 1 panicle correlated with the following features: width of the flag leaf ($r = 0.395$), area of the flag leaf ($r = 0.391$), length of the largest leaf ($r = 0.708$), width of the largest leaf ($r = 0.750$), area of the largest leaf ($r = 0.756$), thickness of the upper internode ($r = 0.545$), thickness of the lower internode ($r = 0.625$), protrusion of the panicle stem ($r = -0.381$), total bushiness ($r = -0.760$), productive bushiness ($r = -0.714$), number of grains from 1 panicle ($r = 0.926$), grain yield ($r = 0.506$), protein content in grain ($r = -0.512$), nitrogen-free extractable substances content in grain ($r = 0.447$). The number of grains per panicle is in close correlation with the width of the flag leaf ($r = 0.423$), the area of the flag leaf ($r = 0.413$), the length of the largest leaf ($r = 0.640$), the width of the largest leaf ($r = 0.701$), and the area of the largest leaf ($r = 0.682$), thickness of the upper internode ($r = 0.520$), thickness of the lower internode ($r = 0.570$), protrusion of the panicle pedicle ($r = -0.409$), general bushiness ($r = -0.758$), productive bushiness ($r = -0.717$), panicle length ($r = 0.350$), grain weight per panicle ($r = 0.926$), grain yield ($r = 0.460$), protein content in grain ($r = -0.596$), nitrogen-free extractable substances content in grain ($r = 0.513$).

The total contribution to the accumulated variance of 6 hypothetical factors was 77.1 %, the first three accounted for 58.85 % (Table 2).

Table 2. Factor loads of the model population of grain sorghum, 2015 - 2018.

| № | Indicator | Factor | | | | | |
|----|--|--------|-------|-------|-------|-------|-------|
| | | Z-1 | Z-2 | Z-3 | Z-4 | Z-5 | Z-6 |
| 1 | Plant height 30 days after germination, cm | -0.43 | 0.36 | -0.08 | -0.39 | 0.04 | 0.24 |
| 2 | Plant height when ripe, cm | 0.35 | 0.20 | -0.57 | -0.58 | -0.07 | -0.08 |
| 3 | Flag leaf length, cm | 0.53 | -0.32 | -0.59 | 0.30 | 0.13 | 0.09 |
| 4 | Flag leaf width, cm | 0.64 | -0.04 | -0.16 | 0.44 | 0.22 | 0.29 |
| 5 | Flag leaf area, cm ² | 0.69 | -0.20 | -0.40 | 0.42 | 0.20 | 0.21 |
| 6 | Length of the largest leaf, cm | 0.83 | 0.15 | -0.39 | 0.02 | -0.10 | -0.10 |
| 7 | Largest leaf width, cm | 0.79 | 0.52 | -0.06 | 0.07 | 0.09 | -0.01 |
| 8 | Largest leaf area, cm ² | 0.82 | 0.43 | -0.23 | 0.05 | 0.06 | -0.09 |
| 9 | Upper internode thickness, cm | 0.66 | 0.30 | 0.44 | 0.16 | -0.04 | -0.01 |
| 10 | Lower internode thickness, cm | 0.69 | 0.48 | 0.32 | 0.19 | -0.08 | -0.12 |

| | | | | | | | |
|-------------------------|--|-------|-------|-------|-------|-------|-------|
| 11 | Panicle leg extension, cm | -0.32 | -0.44 | -0.51 | -0.30 | 0.05 | -0.36 |
| 12 | Total bushiness, pcs | -0.84 | -0.06 | 0.03 | 0.29 | 0.33 | 0.01 |
| 13 | Productive bushiness, pcs | -0.84 | 0.12 | -0.07 | 0.30 | 0.26 | 0.09 |
| 14 | Panicle length, cm | 0.57 | -0.43 | -0.31 | -0.05 | -0.44 | 0.10 |
| 15 | Panicle width, cm | 0.09 | -0.65 | 0.04 | 0.12 | -0.44 | 0.23 |
| 16 | Number of grains from 1 panicle, pcs | 0.83 | 0.16 | 0.15 | -0.24 | 0.02 | 0.27 |
| 17 | Weight of 1,000 grains, g | 0.08 | 0.27 | 0.02 | 0.28 | -0.18 | -0.69 |
| 18 | Grain weight from 1 panicle, g | 0.84 | 0.26 | 0.13 | -0.15 | -0.07 | 0.01 |
| 19 | Grain yield, t/ha | 0.63 | -0.34 | 0.15 | 0.26 | 0.16 | -0.04 |
| 20 | Protein content in grain, % | -0.54 | 0.31 | -0.43 | 0.42 | -0.24 | -0.22 |
| 21 | Fat content in grain, % | -0.29 | 0.54 | -0.21 | 0.13 | -0.13 | 0.27 |
| 22 | Ash content in grain, % | -0.50 | 0.47 | -0.22 | -0.11 | 0.29 | 0.15 |
| 23 | Fiber content in grain, % | -0.45 | 0.11 | 0.19 | 0.19 | -0.57 | 0.26 |
| 24 | Nitrogen-free extractable substances content in grain, % | 0.62 | -0.47 | 0.33 | -0.34 | 0.35 | -0.07 |
| 25 | Starch content in grain, % | 0.14 | -0.32 | 0.25 | 0.22 | 0.07 | -0.27 |
| Dispersion | | 9.26 | 3.14 | 2.28 | 1.94 | 1.38 | 1.27 |
| Dispersion, % | | 37.03 | 12.57 | 9.12 | 7.77 | 5.52 | 5.09 |
| Accumulated variance, % | | 37.03 | 49.59 | 58.71 | 66.48 | 72.00 | 77.09 |

Note: Z 1... Z 6 is a hypothetical factor.

The contribution to the 1st hypothetical factor (37.03 %) is largely determined by the following indicators ($r > |0.7|$): positive – the length of the largest leaf, the width of the largest leaf, the area of the largest leaf, the number of grains from 1 panicle, weight grains from 1 panicle; negative – general bushiness, productive bushiness.

The 2nd hypothetical factor (12.57 %) is largely determined by the contribution of the panicle width ($r = -0.645$), as well as the total effect of other indicators.

The 3rd hypothetical factor accounts for 9.12 % in the accumulated variance, which is determined by the total average effect of a number of indicators: plant height at ripening, flag leaf length, flag leaf area, length of the largest leaf, panicle stem protrusion, protein content in the grain.

The contribution of the 4th factor (7.77 %) is determined by the following indicators: plant height after 30 days, plant height at maturity, flag leaf width, flag leaf area, protein content in the grain.

The greatest contribution to the 5th factor (5.52 %) is made by the following indicators: panicle length, panicle width, and fiber content in the grain. The effect of the indicator "mass of 1,000 grains", as well as the total contribution of other indicators, determine the value of the 6th hypothetical factor (5.09 %).

5 Conclusion

Therefore, in the model population of grain sorghum, a different in magnitude variation of the studied indicators was established, and the most significant correlations were established.

Factor analysis of the matrix of correlation coefficients by the method of principal components made it possible to identify the indicators of the first 6 hypothetical factors that make the greatest contribution (77.1 %) to the accumulated variance.

References

1. A.P. Tsarev, E.V. Morozov, Agrobiological basis for the cultivation and use of sorghum crops in the Volga region (2011)
2. O.P. Kibalnik, Vavilov Journal of Genetics and Breeding, **21**, 6 (2017)
3. O.P. Kibalnik, BIO Web of Conferences (2021)
4. J.L. Moran, W.L. Rooney, Crop Sci., **43**, 3 (2003)
5. B.V.S. Reddy, S. Ramesh, P.S. Reddy, B. Ramaiah, Biology, Euphytica (2007)
6. C. Aruna, P.K. Shrotria, S.K. Pahuja, A.V. Umakanth, B.V. Bhat, A.V. Devender, Crop Pasture Sci., **63**, 12 (2013)
7. G.N. Hasanov, M.R. Musaev, A.O. Mamalaeva, Corn and Sorghum, **4** (2007)
8. V.I. Starchak, Agricultural Scientific Journal, **6** (2020)
9. State Register of Breeding Achievements Admitted to Use "Plant Varieties", **1** (2021)
10. Wide unified CMEA classifier and CMEA international classifier of cultivated species of the genus Sorghum Moench, VNIIR them. N.I. Vavilova (1982)
11. GOST 10846-91, Grain and products of its processing, Protein determination method (Standards Publishing House, Moscow, 1992)
12. GOST 13496.15-97, Feed, compound feed, compound feed raw materials, Methods for Determination of Crude Fat Content (Standards Publishing House, 1998)
13. GOST 26226-95, Feed, compound feed, compound feed raw materials. Methods for Determination of Raw Ash (Moscow: Standards Publishing House, Moscow, 1996)
14. GOST 13496.2-91, Feed, compound feed, compound feed raw materials, Method for Determination of Crude Fiber (Standards Publishing House, Moscow, 1992)
15. GOST 10845-98, Grain and products of its processing, Method for determination of starch (Interstate standard, Moscow, 1998)
16. A.I. Ermakov, Methods of Biochemical Research of Plants (L. Kolos, 1987)