Improvement of the methodology for assessing the productivity of sunflower varieties based on the multilinear regression model with the optimal irrigation regime

Sherzod Hakimov [a)], Ikrom Tursunov, Parvina Akramova, Akmal Fayzullayev, Khilola Murodova

Bukhara state technical university, Bukhara, Uzbekistan

[a)] *Corresponding author:* sherzodhk4@gmail.com

Abstract. Effective management of agriculture in conditions of irrigation water scarcity is an important and urgent problem. In the Bukhara region, using groundwater, growing sunflower (*Helianthus annuus L.*) as a second crop brings high economic benefits. In this study, the relationship between the physiological parameters and yield of the Dushko F_1 and Dilbar varieties was studied using a multilinear regression model. The main task is to determine the possibility of predicting yields based on stem length, number of leaves, and basket diameter. According to the research results, the basket diameter has the strongest impact on yield, and the models are characterized by a high degree of accuracy ($R^2 > 0,8$) and a low degree of error ($MSE < 0,2$). These conclusions serve as an important scientific basis for controlling secondary crops and increasing the efficiency of agriculture in the context of climate change.

INTRODUCTION

In recent decades, global climate change and the shortage of natural resources have increased the need for more sustainable and innovative development of agriculture on a global scale. In regions of Uzbekistan with limited irrigation water supply, such as the Bukhara region, this requirement is becoming even more acute. Increasing the volume of agricultural production through the rational use of available resources and the cultivation of repeated crops remains a strategic task. Sunflower is a valuable repeated crop, resistant to heat and drought due to its biological properties. At the same time, since the level of yield is closely related to the physiological indicators of the plant, forecasting the yield based on these indicators is of great importance for agricultural entrepreneurs. This study was aimed at developing and evaluating a model for predicting the yield of the Dushko F_1 and Dilbar sunflower varieties using groundwater in the conditions of the Bukhara region [1]-[13].

EXPERIMENTAL RESEARCH

The research was conducted during the 2023 vegetation season in the irrigation water-scarce areas of Bukhara region. The Dushko F_1 and Dilbar sunflower varieties were irrigated in a controlled manner at pre-irrigation soil moisture levels of 70–75–70% FC. For each variety, 100 plants were selected. Their physiological indicators were determined based on the following parameters: Stem height (cm), Number of leaves (pcs), Basket diameter (cm), Total weight of seeds per plant (g). After the data were collected, they were analyzed in the Python programming environment. The Pandas library was used for data preparation and processing, while the scikit-learn library was employed for building and evaluating the multiple linear regression model. The dataset was randomly split into 80% training and 20% test sets. Model quality was assessed using the coefficient of determination (R^2), mean squared error (MSE), and visualization methods. Separate regression equations were developed for each variety, and their accuracy was verified through the test data [14]-[21].

RESEARCH RESULTS

For each variety, 100 seedlings were selected. For the Dushko F_1 variety, the height ranged from 125 cm to 173 cm, the number of leaves from 14 to 21, and the seed weight per plant showed variability from 37,82 g to 56,71 g. For the Dilbar variety, the height ranged from 144 cm to 206 cm, the number of leaves from 17 to 24, and the seed weight per plant was recorded from 27,83 g to 41,75 g. These data were used to predict yield using the multiple linear regression method. The general statistical characteristics of the data are given in the following table (Table 1).

TABLE 1. Statistical parameters of physiological indicators (stem height, number of leaves, and basket diameter) and seed weight per plant for Dushko F_1 and dilbar sunflower varieties

№	Variety	Parameter	Average value	Standard deviation	Minimum value	Maximum value
1	Dushko F_1	Height, cm	172,8	8,0	125,0	173,0
2		Number of leaves, pcs	19,7	1,2	14,0	21,0
3		Basket diameter, cm	17,8	0,6	16,5	19,0
4		Seed weight per plant, g	45,5	3,2	37.8	56,7
5	Dilbar	Height, cm	145,0	10,3	144,0	206,0
6		Number of leaves, pcs	17,1	1,2	17,0	24,0
7		Basket diameter, cm	18,7	0,7	17,5	20,0
8		Seed weight per plant, g	37,9	2,3	27,8	41,7

As can be seen from Table 1, the Dushko F_1 variety has significantly higher plant height (average 172,8 cm) and number of leaves (average 19,7) than the Dilbar variety (145,0 cm and 17,1 pieces). However, in terms of yield, the Dilbar variety surpasses Dushko F_1 (31,83 grams) with an average of 37,5 grams. These differences may be related to the phenotypic characteristics of the varieties and the genetic basis of the yield potential (Table 2).

TABLE 2. Physiological indicators and on yield of Dushko F_1 and Dilbar varieties: Analysis of genetic characteristics.

Variety name	Stem height (cm)	Number of leaves (units)	Basket diameter (cm)	Yield (kg)
Dushko F_1	178,5	21,0	27,8	3,9
Dilbar	179,2	20,8	27,6	3,8

In Table 4.7, for the Dushko F_1 variety, the average stem height was 178.5 cm, the number of leaves was 21 pieces, the basket diameter was 27.8 cm, and the yield was 3.9 kg/plant. For the Dilbar variety, these indicators were 179.2 cm, 20.8 pcs., 27.6 cm, and 3.8 kg/plant, respectively. These differences may be related to the genetic characteristics of the varieties.

Dushko F_1 regression equation:

$$Y = -2{,}12 + 0{,}031X_1 + 0{,}043X_2 + 0{,}076X_3 + \in \quad (1)$$

Here:

$$\overline{Y_1} = \frac{Y_1 + Y_2 + Y_3 + \cdots + Y_{100}}{100} \quad (2)$$

Y - dependent variable, i.e., yield (kg);

$$\overline{X_1} = \frac{X_1 + X_2 + X_3 + \cdots + X_{100}}{100} \quad (3)$$

$$\overline{X_2} = \frac{X_1 + X_2 + X_3 + \cdots + X_{100}}{100} \quad (4)$$

$$\overline{X_3} = \frac{X_1 + X_2 + X_3 + \cdots + X_{100}}{100} \quad (5)$$

X_1, X_2,X_3 - stem height (cm), number of leaves (units), basket diameter (cm);

$$\beta_1 = \frac{\sum(X_1 - \overline{X_1}) \cdot (Y_1 - \overline{Y_1})}{\sum(X_1 - \overline{X_1})^2} \quad (6)$$

$\beta_1 =$ 0,031 stem height (cm); If the stem height increases by 1 cm, and all other variables remain unchanged, the yield increases by 0.031 kg. This means that stem height has a positive effect on yield, meaning taller plants yield more.

$$\beta_2 = \frac{\sum(X_2 - \overline{X_2}) \cdot (Y_1 - \overline{Y_1})}{\sum(X_2 - \overline{X_2})^2} \quad (7)$$

$\beta_2 = 0{,}043$ the number of leaves, i.e., if the number of leaves increases by 1 unit, the yield increases by 0,043 kg. This confirms the hypothesis that the number of leaves increases the efficiency of photosynthesis.

$$\beta_3 = \frac{\sum(X_3 - \overline{X_3}) \cdot (Y_1 - \overline{Y_1})}{\sum(X_3 - \overline{X_3})^2} \quad (8)$$

$\beta_3 = 0{,}076$ when the number of leaves, i.e., the diameter of the basket, increases by 1 cm, the yield increases by 0,076 kg. This indicator has the greatest influence among all three factors, i.e., the size of the basket has the most significant impact on yield.

$$\beta_0 = Y - \beta_1 X_1 - \beta_2 X_2 - \beta_3 X_3 \quad (9)$$

$\beta_0 = -2{,}12$ constant value (intercept) is estimated at −2.12 kg of yield when all independent variables (stem, leaves, basket) are zero. This is technically the starting point of the model. Of course, in practice, such a situation (zero stem, zero leaf, and zero basket) does not exist, but the necessary mathematical point for the model.

$$SSR^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \qquad MSE = \frac{1}{n}\sum(y_i - \hat{y}_i)^2 \quad (10)$$

Here y_i − real values, $\hat{y}_i$ − predicted model n - and the number of samples represents the values.

In this case, the $R^2 = 0.85$ model explains changes in yield with an accuracy of 85%. This is a very high result, the error between the forecast MSE = 0,13 and the real value is small, therefore, the model is accurate.

TABLE 3. Regression coefficients for the Dushko F_1 variety

Indicators	**Coefficient value**
Stem height	β_1= 0,031
Number of leaves	β_2=0,043
Basket diameter	β_3=0,076
Constant value	β_0=−2,12
R^2=0,85	
MSE = 0,13	

The multilinear regression model constructed to predict the yield of the Dushko F_1 sunflower variety revealed that physiological indicators such as leaf count, stem length, and basket diameter have a significant positive effect on yield; the R^2 value of the model is 0,85, which indicates that it is capable of explaining 85% of yield differences, and the regression equation is expressed as follows (Table 3).

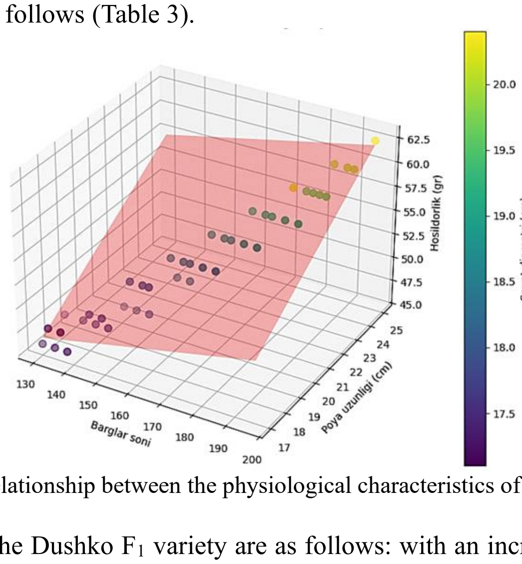

FIGURE 1. Clear statistical relationship between the physiological characteristics of the Dushko F_1 variety and yield

The regression coefficients of the Dushko F_1 variety are as follows: with an increase in stem height (β_1 = 0,031) by 1 cm, the yield increased by 0,031 kg, and the number of leaves increased.(β_2 = 0,043) the yield increases by 0,043 kg with each increase in 1 unit, and the yield increases by 0,076 kg with each increase in the basket diameter (*β_3* = 0,076) by 1 cm. The constant value (*β_0* = −2,12) represents the initial yield value when all other indicators are zero, i.e., − 2,12 kg.

In this graph, the multilinear regression model shows a clear statistical relationship between the physiological characteristics of the Dushko F_1 variety and yield. The proximity of the prediction level and real points on the graph confirms the adequacy and compatibility of the model.

Dilbar variety regression equation:

$$Y = -1{,}97 + 0{,}026X_1 + 0{,}039X_2 + 0{,}071X_3 + \in \quad (11)$$

Here: Y - dependent variable, i.e., yield (kg);

X_1, X_2,X_3 - stem height (cm), number of leaves (units), basket diameter (cm);

$\beta_1 =$ 0,026 stem height (cm), i.e., if the stem height increases by 1 cm, and all other variables remain unchanged, the yield increases by 0.031 kg. This means that stem height has a positive effect on yield, meaning taller plants yield more.

$\beta_2 =$ 0,039 number of leaves In this case, if the number of leaves increases by 1, the yield increases by 0.043 kg. This confirms the hypothesis that the number of leaves increases the efficiency of photosynthesis.

$\beta_3 =$ 0,071 - the number of leaves, and when the diameter of the basket increases by 1 cm, the yield increases by 0.076 kg. This indicator has the greatest influence among all three factors, i.e., the size of the basket has the most significant impact on yield.

$\beta_0 = -1{,}97$ - constant value (intercept): When all independent variables (stem, leaves, basket) are zero, the yield is estimated at −1.97 kg. This is technically the starting point of the model. Of course, in practice such a situation (zero stem, zero leaf, and zero basket) does not exist, but it is a necessary mathematical point for the model.

$$SSR^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2\,, \qquad MSE = \frac{1}{n}\sum(y_i - \hat{y}_i)^2 \quad (12)$$

Here,, y_i − actual values, $\hat{y}_i$ − model predicted, n - and the number of samples represents the values.. In this, *R^2*=0,82 - the model explains changes in yield with an accuracy of 82%. This is a very high result, the error between the forecast *MSE = 0,15* and the real value is small, therefore, the model is accurate.

The regression coefficients of the Dilbar variety are as follows (Table 4).

TABLE 4. Regression coefficients for the Dilbar variety

Indicators	**Coefficient value**
Stem height	β_1= 0.026
Number of leaves	β_2=0.039
Basket diameter	β_3=0.071
Constant value	β_0=−1,97
R^2=0,82	
MSE = 0,15	

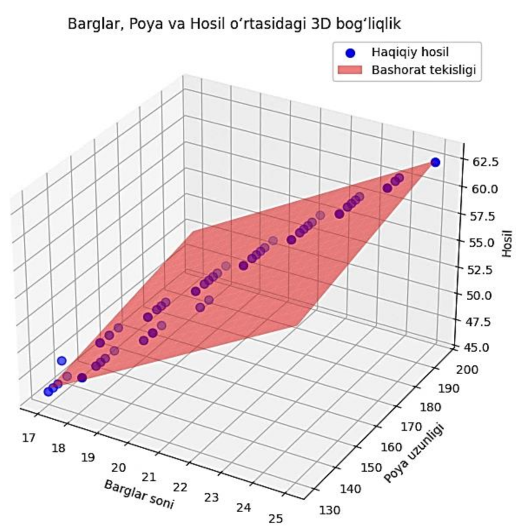

FIGURE 2. A clear statistical relationship between the physiological characteristics of the Dilbar variety and yield.

In this graph, the multilinear regression model shows a clear statistical relationship between the physiological characteristics of the Dilbar variety and yield. The proximity of the prediction level and real points on the graph confirms the adequacy and compatibility of the model. In this case, with an increase in stem height ($\beta_1 = 0{,}026$) by 1 cm, the yield increased by 0,026 kg, with an increase in the number of leaves ($\beta_2 = 0{,}039$) by 1 unit, the yield increased by 0,039 kg, and the diameter of the basket ($\beta_3 = 0.071$) by 1 unit.

With an increase of 1 cm, the yield increases by 0,071 kg. The constant value ($\beta_0 = -1{,}97$) indicates the initial yield value at zero for all other indicators, i.e., −2,12 kg. In both varieties, basket diameter was noted as the most important factor (see Table 4).

As can be seen from the results, the diameter of the basket is the most strongly influencing factor on yield. This circumstance confirms the importance of the seed basket in seed formation. Stem height and the number of leaves also have a positive effect, playing an important role in plant growth and photosynthesis. The results of the Dushko F_1 variety were slightly higher than that of Dilbar, which may be due to its genetic resistance. High values of the model's R^2 (0,82-0,85) indicate its reliability.

CONCLUSIONS

This study confirmed the possibility of predicting yields based on the physiological indicators of sunflower varieties. The diameter of the basket manifested itself as the main factor directly affecting yield, which confirms its role in the seed formation process. Stem height and the number of leaves positively influenced the yield through increased photosynthesis efficiency and biomass accumulation. Due to physiological characteristics and genetic stability, the Dushko F_1 variety showed higher yield indicators compared to the Dilbar variety. At the same time, for further improvement of the model, it is necessary to take into account environmental factors, the chemical composition of the soil, the properties of the root system, changes in light intensity and temperature. At the same time, the accuracy of yield forecasting can be increased by 20-30% through the widespread use of artificial intelligence and machine learning algorithms.

The research results showed that it is possible to effectively predict the yield of the Dushko F_1 and Dilbar sunflower varieties using groundwater under the conditions of the Bukhara region using a multilinear regression model. The diameter of the basket, stem height, and number of leaves were determined as the main physiological indicators of yield. The obtained results serve as a scientific and practical basis for the maximum use of repeated crops, optimization of agricultural technologies, and adaptation to climate change. In the future, through the introduction of a comprehensive database and artificial intelligence technologies, it is possible to make yield forecasts more reliable and effective

REFERENCES

1. Lukov M. Agrotechnology of the sunflower Mash'al F_1 hybrid. // Journal of Agriculture of Uzbekistan. -2019. Issue №2, P. 36
2. Mamataliyev A.B. Irrigation procedures for oilseed crops. // Republican scientific and practical conference on the topic "Environmental problems of rational use of water and land resources in irrigated agriculture" (November 24-25). Tashkent. 2017. P. 82-84
3. Khalmuratova B.U. Development of an optimal irrigation procedure for main crops in conditions of water shortage (on the example of the Republic of Karakalpakstan) Doctor of Philosophy (PhD) in Agricultural Sciences. Dissertation. Author's abstract. Tashkent. 2021. -P. 13-20.
4. Khilola Amankulova, Nizom Farmonov, Uzbekkhon Mukhtorov & László Mucsi (2023) Sunflower crop yield prediction by advanced statistical modeling using satellitederived vegetation indices and crop phenology, Geocarto International, 38:1, 2197509, DOI: 10.1080/10106049.2023.2197509. https://doi.org/10.1080/10106049.2023.2197509
5. Milyutkin V.A., Tsirulev A.P., Tolpekin S.A. Issledovaniya novykh vidov azotoserosoderjashchikh jidkih udobreniy i sposobov vneseniya pri vozdelyvani podsolnechnika// Aktualnye problemy rationalnogo ispolzovaniya zemelnyx resursov: sb. St. po mater Vseros III. (National.) nauch. -practitioner. conf. / pod borshch. ed. S.F. Sukhanova. Kurgan, 2019. S. 116-122.
6. Sh. R. Akhmedov, Kh.T. Tukhtaeva, Z.U. Amanova, I.N. Tursunov, Sh.H. Hakimov, M.M. Rajabova, M.B. Bahriddinov, Sh. Egamurodov, and S. Mirzayev. (2023). Scientific basis of the effect of groundwater sources on

annual plant growth under current natural conditions. International Conference on Advanced Agriculture for a Sustainable Future (pp. 1755-1315). Gothenburg. Sweden: IOP Publishing.
7. Amankulova Khilola, Farmonov Nizom, Akramova Parvina, Tursunov, Ikrom, Mucsi Laszlo. (2023). Comparison of PlanetScope, Sentinel-2, and landsat 8 data in soybean yield estimation within-field variability with random forest regression. Helion. 9. e17432. 10.1016/j.helion. 2023.e17432.
8. Zagorulko A.V. Nauchnoe obsnovanie optimizatsi teknologiy vozdelyvaniya i povyshenia produktivitsi ozimoy pshenitsy, kukuruzy i podsolnechnika na vyshchelochennom chernozeme Zapadnogo Predkavkazya: avtoref. dis. d.-ra s.-ch. nauk/ Kub. GAU. - Krasnodar, 2005. - 58 p.
9. Laszlo Zsombik. Effect of weather parameters on the date of emergence and the elapsed time to blooming of sunflower// Cereal Res. Commun -2005. - #1. p. 33. -P. 337-340.
10. Bazmi, R. R., & Panichayupakaranant, P. (2023). The Role of Roots, Stems, and Leaves in Plant Function: Structural and Physiological Perspectives for Optimized Plant Growth. Australian Herbal Insight, 6(1), 1-5. https://doi.org/10.25163/ahi.619956
11. Tariq, M., Ahmad, S., Fahad, S., Abbas, G., Hussain, S., Fatima, Z., Nasim, W., Mubeen, M., Rehman, M. H. ur, Khan, M. A., Adnan, M., Wilkerson, C. J., & Hoogenboom, G. (2018). The impact of climate warming and crop management on phenology of sunflower-based cropping systems in Punjab, Pakistan. Agricultural and Forest Meteorology, 256-257, 270-282. https://doi.org/10.1016/j.agrformet.2018.03.015.
12. Gupta, S. K. (2023). Plant growth and development: Physiological and genetic perspectives. Springer. https://doi.org/
13. Monteith, J. L. (1977). Climate and the efficiency of crop production in Britain. Philosophical Transactions of the Royal Society B: Biological Sciences, 281(980), 277-294. doi:10.1098/rstb.1977.0140
14. Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for calculating crop water requirements (FAO Irrigation and Drainage Paper No. 56). Rome: FAO. Retrieved from http://www.fao.org/3/x0490e/x0490e00.htm
15. Richards, L.A. (1931) Capillary Conduction of Liquids through Porous Mediums. Journal of Applied Physics, 1, 318-333. http://dx.doi.org/10.1063/1.1745010
16. Hatfield, J. L., & Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. Weather and Climate Extremes, 10(Part A), 4-10. https://doi.org/10.1016/j.wace.2015.08.001
17. Narin, O. G., Sekertekin, A., Saygin, A., Balik Sanli, F., & Gullu, M. (2021). Yield estimation of sunflower plant with CNN and ANN using Sentinel-2. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-4/W5-2021, 385-389. https://doi.org/10.5194/isprs-archives-XLVI-4-W5-2021-385-2021
18. Gökler, S. (2024). Predicting sunflower yield using machine learning methods with hyperparameter tuning. Selcuk Journal of Agriculture and Food Sciences, 38(3), 445–462.
19. Hillel, D. (2004). Soil physics and environmental sustainability. Journal of Soil Science, 55(3), 123–130.
20. Patil, Amol & Deosarkar, D.B. & Khatri, Narendra & Ubale, Ankush. (2023). A comprehensive investigation of Genotype-Environment interaction effects on seed cotton yield contributing traits in Gossypium hirsutum L. Using multivariate analysis and artificial neural network. Computers and Electronics in Agriculture. 211. 107966. 10.1016/j.compag.2023.107966.
21. Pokhrel A, Virk S, Snider JL, Vellidis G, Hand LC, Sintim HY, Parkash V, Chalise DP, Lee JM and Byers C (2023) Estimating yield-contributing physiological parameters of cotton using UAV-based imagery. Front. Plant Sci. 14:1248152. doi:10.3389/fpls.2023.1248152