UDC 528.8
DOI https://doi.org/10.52171/herald.342

Forecasting Cotton Productivity in the Mil-Mugan Economic Region using Climate Variables and Remote Sensing Data

B.H. Babashli

Azerbaijan National Aviation Academy (Baku, Azerbaijan)

For correspondence:
Bakhtiyar Babashli e-mail: bakhtiyar.babashli@gmail.com

Abstract

Satellite imagery is considered as one of the most valuable and functional data sources for evaluating and forecasting production in agricultural systems. This study aims to forecast productivity in cotton fields within the Mil-Mugan economic region by incorporating vegetation indices and climatic variables. NDVI indicators were derived from Sentinel-2 data, temperature (T) was obtained from ERA5-Land data, and precipitation (R) was collected from the CHIRPS database. The data were analyzed using Google Earth Engine, ArcGIS Pro, and Python environments, consolidating them into an integrated temporal-spatial dimension. The correlation and regression analysis assessed the link among NDVI, temperature, precipitation, and productivity, leading to the development of a multivariate model. The findings derived from the model indicate that NDVI is the primary factor influencing productivity. The model developed for one region was implemented in the other three locations, and the correlation between the actual and predicted indicators was assessed using R^2 and RMSE metrics. The results indicate that the model's forecast accuracy fluctuates within the 90% range, considered appropriate for predicting regional cotton yield. The study's findings are significant for enhancing operational monitoring and planning of agrotechnical decisions with satellite data.

Keywords: cotton, productivity, satellite imagery, climate data, regression model.

Submitted *22 December 2025*
Accepted *10 September 2026*
Published *25 September 2026*

For citation:
B.H. Babashli
[Forecasting Cotton Productivity in the Mil-Mugan Economic Region using Climate Variables and Remote Sensing Data]
Herald of the Azerbaijan Engineering Academy, 2026, vol. 18 (3), pp. 122-131
ISSN (p): 2076-0515, ISSN (e): 2789-8245
© 2026 B.H. Babashli. This article is distributed under the terms of the https://creativecommons.org/licenses/by-nc/4.0/ (CC BY-NC 4.0)

Mil-Muğan iqtisadi rayonunda iqlim dəyişənləri və məsafədən müşahidə göstəricilərinin inteqrasiyası ilə pambıq məhsuldarlığının proqnozlaşdırılması

B.H. Babaşlı
Azərbaycan Milli Aviasiya Akademiyası (Bakı, Azərbaycan)

Xülasə

Peyk təsvirləri kənd təsərrüfatı sistemlərində məhsuldarlığın qiymətləndirilməsi və proqnozlaşdırılması üçün ən etibarlı və operativ məlumat mənbələrindən biri hesab olunur. Bu tədqiqatın məqsədi Mil-Muğan iqtisadi rayonunda yerləşən pambıq sahələrinin məhsuldarlığı vegetasiya indeksi və iqlim dəyişənlərinin inteqrasiyası ilə proqnozlaşdırılmasıdır. Bu məqsədlə Sentinel-2 məlumatları əsasında NDVI göstəriciləri, ERA5-Land məlumatlarından temperatur (T) və CHIRPS bazasından yağıntı (R) məlumatları əldə edilmişdir. Məlumatlar Google Earth Engine, ArcGIS Pro və Python mühitində emal edilmiş və vahid zaman-məkan ölçüsünə gətirilmişdir. NDVI, temperatur və yağıntı ilə məhsuldarlıq arasındakı əlaqə korrelyasiya və reqressiya analizi ilə qiymətləndirilmiş və çoxdəyişənli model qurulmuşdur. Model üzrə əldə edilən nəticələr göstərir ki, NDVI məhsuldarlığın əsas determinatıdır. Bir rayon üzrə qurulan model digər üç rayon üçün tətbiq edilmiş və real-proqnoz göstəriciləri arasındakı uyğunluq R^2 və RMSE metrikləri ilə qiymətləndirilmişdir. Nəticələr göstərir ki, modelin proqnoz dəqiqliyi 90% intervalında dəyişir ki, bu da regional səviyyədə pambıq məhsuldarlığı proqnozlaşdırılması üçün qənaətbəxş dəqiqlik hesab olunur. Tədqiqatın nəticələri peyk məlumatlarına əsaslanan operativ monitorinqin və aqrotexniki qərarların planlaşdırılmasının təkmilləşdirilməsi baxımından əhəmiyyətlidir.

Açar sözlər: pambıq, məhsuldarlıq, peyk təsviri, iqlim məlumatları, reqressiya modeli.

Прогнозирование урожайности хлопка в Миль-Муганском экономическом районе с использованием индексов дистанционного зондирования и климатических переменных

Б.Х. Бабашлы
Азербайджанская Национальная академия авиации (Баку, Азербайджан)

Аннотация

Для оценки и прогнозирования продуктивности сельскохозяйственных культур спутниковые снимки считаются одним из наиболее надёжных и эффективных источников данных. Целью данного исследования является прогнозирование урожайности хлопковых полей в Миль-Муганском экономическом районе с использованием вегетационных индексов и климатических данных. Для достижения этой цели были использованы показатели температуры (T) из данных ERA5-Land и осадки из базы данных CHIRPS. Индекс NDVI был рассчитан на основе снимков спутника Sentinel-2. После предварительной обработки данных инструменты ArcGIS Pro, Python и Google Earth Engine были применены для формирования единого пространственно-временного набора данных. Для анализа взаимосвязей между температурой, осадками, продуктивностью и NDVI была использована многомерная модель. Согласно результатам моделирования, NDVI является основным фактором, определяющим продуктивность хлопка. Разработанная модель, созданная для одного региона, была затем применена к трём соседним регионам. Для оценки соответствия между фактическими и прогнозируемыми значениями использовались показатели R^2 и RMSE. Результаты показывают, что прогнозирование урожайности хлопка на региональном уровне является удовлетворительной точностью, Поскольку точность прогноза модели изменяется в пределах 90%. Полученные выводы имеют важное значение для планирования и оперативного мониторинга агротехнических решений на основе спутниковых данных.

Ключевые слова: хлопок, урожайность, спутниковые снимки, климатические данные, регрессионная модель.

Introduction

Cotton is one of the world's most important agricultural and industrial raw resources, utilized in textiles, oil, seed processing, and livestock feed. The cotton sector holds substantial importance globally due to its impact on economic revenue, employment, and export income [1]. Cotton farming, alongside its historical traditions, significantly contributes to the socio-economic development of the Aran region, particularly the Mil-Mugan economic area [2]. Cotton crops are strategically positioned within the region's agricultural framework, owing to farm profitability and the sustainability of agro-industrial supply chains.

Modern agricultural science analyzes agrotechnical, ecological, meteorological, and biophysical factors to understand production variability. The advancement of remote sensing technologies has elevated the monitoring of agricultural crop conditions, particularly through the widespread use of indicators such as NDVI (Normalized Difference Vegetation Index) to evaluate vegetation cover, intensity, and photosynthetic activity [3, 4]. The dynamics of NDVI values precisely reflect plant biomass growth, alterations in leaf indicators, and important phases of the vegetation cycle, demonstrating a strong link in yield modelling.

However, recent research has shown that yield should be based on a combination of vegetative indices and climatological variables as temperature and precipitation [5, 6]. Temperature serves as a determinant of the phenological development rate, growing season, and physiological responses of cotton plants. Precipitation or water supply, conversely, influences soil moisture balance, water stress, and vegetation dynamics [7]. Extensive modeling studies indicate that the precision of yield projections markedly improves when NDVI, temperature, and precipitation are combined [8]. This method has become more advantageous for crop planning and optimizing resource consumption, particularly in irrigated agricultural areas.

The Mil-Mugan economic region is characterized by optimal climate conditions for cotton farming, an extensive irrigation infrastructure, substantial arable land, and significant heat accumulation throughout the growing season. Productivity disparities among regions arise from varying water availability, climate variability, soil characteristics, and diverse management strategies. Utilizing statistical and satellite-based methodologies to ascertain these differences is crucial for enhancing agricultural management, precisely forecasting yield, and empirically validating the decision-making process [9].

The primary objective of this research is to use NDVI, temperature, and precipitation variables to evaluate cotton crop productivity in four administrative regions of the Mil-Mugan economic region. This study aimed to determine the correlations among the variables and forecast productivity based on the established statistical relationships. The research demonstrates that the combination of satellite-derived vegetation indices and meteorological data yields great precision in the classification and forecasting of cotton productivity, hence enhancing both scientific foundations and practical decision-making.

Research methods

The study covered four administrative districts within the Mil-Mugan economic region: Saatli, Sabirabad, Imishli, and Beylagan (Fig. 1). These districts are situated in the

lowland climatic zone of Azerbaijan and have been identified as the principal regions for the advancement of cotton cultivation. The mean annual temperature in the region is from 14 to 16°C, while the yearly precipitation fluctuates between 250 and 350 mm. The soil cover is predominantly characterized by gray-brown and alluvial-peat varieties. Cotton farming primarily grows on irrigated fields, with the irrigation regime being a primary component that directly influences productivity.

The data used in the study were collected from the following sources (Fig. 2):

Statistical indicators: Cotton yield (cents/ha), cotton production (tons), and cultivated area (hectares) per region as reported by the State Statistical Committee of the Republic of Azerbaijan [2].

Remote sensing: NDVI (Normalized Difference vegetative Index) indices collected from the Sentinel-2 satellite covered the main phase of the vegetative period (August) [10]. Temperature and precipitation data have been collected from the ERA-5 and CHİRPS satellites:

$$\text{NDVI} = \frac{NİR - RED}{NİR + RED} \quad (1)$$

The research employed a multi-stage methodology framework to identify and model cotton yield (t/ha):

1. Assessment of production. The average productivity for each region has been calculated:

$$Y = P \setminus A \quad (2)$$

where Y – yield (t/ha), P – total cotton production (tons), A – area under cultivation (ha).

2. Correlation analysis. The relationship between NDVI indicators [11] and yield was assessed using the Pearson correlation coefficient (r) [12].

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2 \ \sum(Y_i - \bar{Y})^2} \quad (3)$$

3. Predictive modeling. A linear regression model developed using actual data from one region was utilized to forecast the productivity of other regions [13]:

$$Y = \beta_0 + \beta_1 x\ NDVİ + \beta_2 x\ T + \beta_3\ x\ R \quad (4)$$

The model fit was assessed using R^2, p-value and RMSE. The model was calibrated for one region and afterwards applied to three more regions to assess the predictions. The comparison utilized actual-predicted values, relative error percentages, and confidence intervals.

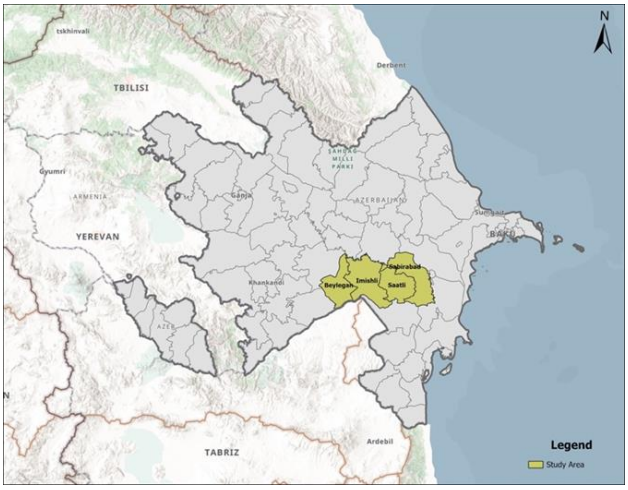

Figure 1 – Study Area

Data acquisition and initial analysis were conducted on the Google Earth Engine (GEE) platform. The GEE environment was used to extract and filter NDVI indicators from Sentinel-2 satellite pictures, as well as temperature and precipitation records from ERA5-Land and CHIRPS data. The acquired raster and tabular data were incorporated into the ArcGIS Pro software, where spatial analysis, zonal statistics, and the generation of spatial data necessary for the model were conducted. The creation of models, computation of productivity metrics, regression

analysis, and graphical representation of results were conducted exclusively within the Python environment. The spatial distribution of NDVI and climatic indices by region was presented in ArcGIS Pro.

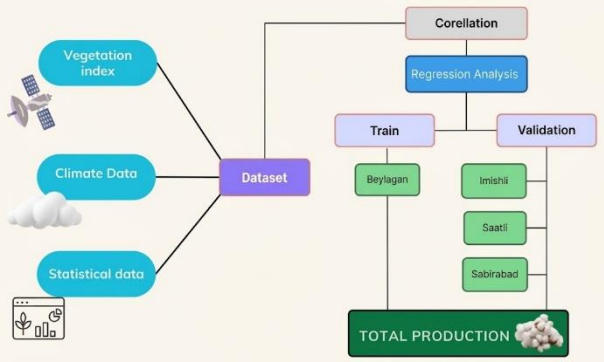

Figure 2 – Flowchart of predicting cotton production

Reuslt and Discussion

The analysis of four administrative units within the Mil-Mugan economic region revealed a clear positive association between NDVI values and cotton yield. The statistical data derived from the cultivated regions indicate that the rise in NDVI values during the peak vegetative growth period directly influenced yield metrics. The correlation coefficient between NDVI and yield was first computed for the Beylagan region. The result falling within the 0.96 interval corroborated that the vegetation index is a substantial predictor in yield models. Subsequently, the production of other locations was assessed using this correlation.

This outcome matches the international research and demonstrates that NDVI is a dependable predictor of cotton output. ERA5-Land (T) and CHIRPS data (R), which directly influence the vegetative development dynamics of plants, were incorporated as primary explanatory variables in the multivariate regression model alongside NDVI. The integration of NDVI, temperature, and precipitation data inside a single model markedly increased the precision of yield forecasts. The study employed a multivariate regression model. The model assessment provided the following results:

- Model explanatory power (R^2): 0.96
- RMSE: average 0.4-4.3 sent/ha
- It was determined that the coefficients β2 (Temperature) and β3 (Precipitation) were 0.0001, which was the most appropriate.

The model developed for Beylagan region was also applied to three other regions, and the comparison of the forecasts with the actual values showed a high level of agreement:

- Accuracy of cultivated area: 95–99%
- Accuracy of yield forecasts: 87–98%
- Accuracy of production: 85–99%

Table 1 – Estimated cotton yield (t/ha)

		Area (ha)		Yield (S/ha)		Production (ton)		RMSE
Rayon	Dataset	Actual	Predict	Actual	Predict	Actual	Predict	
Beylagan	TRAIN	8582.9	8376	35.09	33.06	28427	27253	2.8
Imishli	TEST	7033.7	6,903.28	31.23	29.46	21968	20237	1.8
Saatli	TEST	12969	12,925.48	33.47	29.17	43406	36716	4.3
Sabirabad	TEST	13357.2	14,044.29	25.45	25.05	33994	33773	0.4

The modeling findings indicated that the multivariate regression model that included NDVI, temperature, and precipitation variables performed highly in forecasting cotton

production across areas. The coefficient of determination (R^2) for the model ranged from 0.87 to 0.98 across regions, indicating that the model accounted for 87% to 98% of the variability in actual yield. The elevated R^2 validates the robust prediction capacity of the model and the presence of consistent statistical correlations between NDVI and meteorological variables. International cotton yield studies indicate that R^2 above 0.70 in models using satellite-based vegetation indices and climate indicators represents high, while a variation of 0.85–0.90 is acceptable for practical application.

RMSE is generally evaluated using the relative error (RMSE/mean yield) in agricultural yield modeling studies. Current international research suggests that the relative error (RRMSE = RMSE / mean yield) in yield models utilizing remote sensing and climate indicators is a primary metric for assessment, with an RRMSE < 10% signifying strong model accuracy [14]. Numerous studies deem a 10–20% accuracy range as acceptable, a standard frequently referenced in the literature about remote sensing-based yield forecasts [15]. A relative error beyond 20% signifies that the forecasting model is ineffective and exhibits a lack of concordance with actual observations. This method is reasonable because relative error is considered a more objective measure than absolute RMSE, given the variations in productivity levels among different crops and locales. This methodological approach is common because various crops exhibit distinct biomass potentials and productivity levels. For instance, an RMSE of 0.5 t/ha in high productivity conditions may correspond to a relative error of merely 10% due to the elevated average productivity, thereby allowing the model to remain classified within the high accuracy category.

The model's forecast error, assessed by the RMSE indicator, ranged from 2.72 sent/ha across several locations. The typical cotton yield values (2.4–2.8 centners/ha) indicate a relative error of roughly 7%, considered acceptable for agricultural forecasting research. The low RMSE signifies that the model effectively captures both overarching trends and regional variability within distinct regions.

The results were spatially integrated to provide a productivity map covering the entire research area (Fig. 3). The combination of forecast indicators at a regional scale facilitated the visualization of spatial variability, the detection of disparities in productivity levels, and the evaluation of production dynamics across several zones of the research area. The predicted cotton yield data in the research area were normalized for statistical analysis and categorized into five classes (Fig. 4). This classification is meant to make it easier to determine how yield changes within area, find different output zones. Subsequent to classification, the yield distribution across regions was examined, and the geographical dispersion and intensity level of each category were evaluated. This method made it possible to find both places with high yields and areas with low yields. The comparison of yield classes across regions established a dependable analytical foundation for assessing the structural aspects of cotton production, the extent of variability, and potential intervention strategies. According to the data, the area of class 5 (10000 ha), which is high-yielding, was bigger than the others.

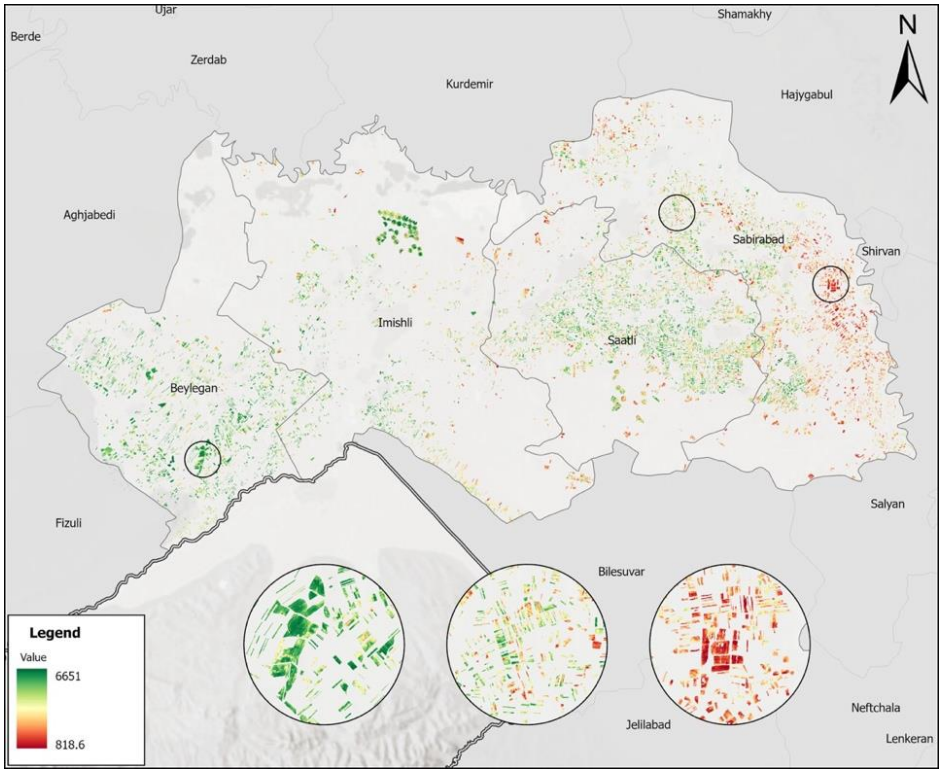

Figure 3 – Cotton Production

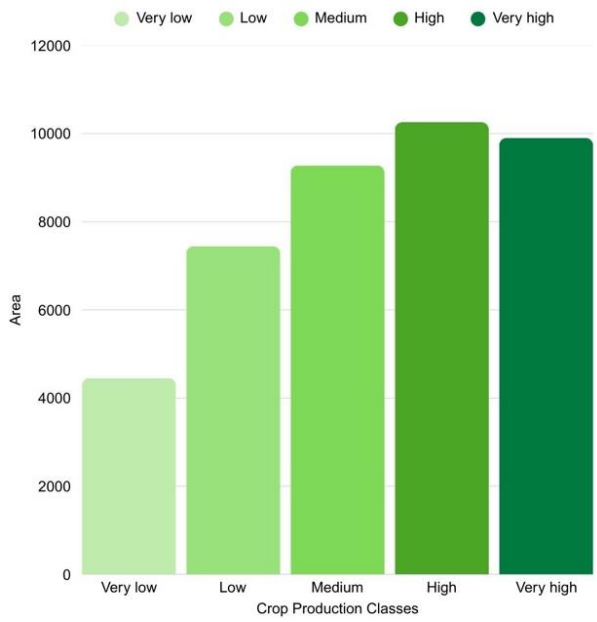

Figure 4 – Cotton Production Classes

Class 4 was notably higher in Saatli. The minimum area of class 1 was measured (4000 ha), mainly in region Sabirabad.

The large area of cotton fields in the Saatli district contributes to the widespread distribution of high-yield areas. In terms of area distribution, 32 percent of the Beylagan district is classified as high-yield, while only 1 percent is categorized as low-yield. Simultaneously, conversely, in the Sabirabad district, the high yield is recorded at 16 percent, while the lowest yield stands at 22 percent. The data indicates that Beylagan district exhibits the highest level of productivity compared to the other districts, while Sabirabad district demonstrates the lowest level of productivity.

Model performance was determined to be reliable when comparing the training and test regions. The R^2 metric in the training area was 0.96, whereas in the testing areas it ranged from 0.85 to 0.99. The peak measurement was observed in the Sabirabad region (99%). This indicates that the model achieved good accuracy not only in its training area but also

when applied in other places. This outcome demonstrates that the model holds strong spatial generalization capabilities and has the potential to clarify inter-regional production variations. A comparable consistency is noted in the RMSE values. The Sabirabad region exhibits higher predictive accuracy relative to other regions.

The results indicate that NDVI's representation of vegetation dynamics, the influence of temperature on biomass development and photosynthesis, and the correlation between precipitation and water availability are significant explanatory variables for the model. The model's elevated R^2 and decreased RMSE show that the combination of meteorological factors with satellite data enhances the precision of yield predictions. This aligns with the global utilization of NDVI-based yield models and enhances the practical applicability of the study. Satellite data and climate indicators are reliable analytical tools for cotton yield evaluation and risk management.

The p-value signifies how reliable the model is and whether the statistically-observed difference is random or not. In scientific research concerning cotton yield prediction, a p-value below 0.05 is typically regarded as statistically significant, indicating that the model demonstrates substantial concordance with actual data and that the likelihood of prediction error being random is minimal. Numerous studies in agriculture, particularly in modeling vegetation-climate relationships, indicate that p-value values may not be within the <0.05 range due to limited observations, heterogeneous soil-climate conditions, and variable meteorological variables. In agro-meteorological models, a p-value between 0.05 and 0.10 represents "borderline significance" suggesting that the model aligns with the actual indicators, if not entirely, to a considerable degree.

In the current study, the p-value that was derived from the model that was used to forecast cotton production was at the level of 0.08. This value means that the model has a weak but acceptable statistical prediction. The yield in cotton fields is influenced by many climatic, irrigation, soil fertility, and genetic factors, rendering the models not entirely deterministic, which may result in elevated p-values.

If the important factors (NDVI and temperature) in the model are statistically significant and the model correctly explains the overall trend, then a p-value of 0.08 is fine for agricultural forecasting studies, and the model can be used in a lot of different ways. Consequently, $p = 0.08$ is regarded as a weak yet acceptable statistical prediction in the context of cotton yield forecasting. This number signifies that the disparity between the actual and projected indicators is not entirely consistent, however the model accurately reflects the overarching trend. Scientific studies indicate that in regions characterized by significant natural variability, such as agriculture, elevated p-values are common and have no impact on the acceptance of model performance.

Conclusion

This study validated the viability and efficiency of forecasting cotton yields in the Mil-Mugan economic zone through the combination of vegetation indicators and climatic variables derived from satellite imagery. The NDVI data from the Sentinel-2 satellite were used as a primary factor of yield. It was found through correlation and

multivariate regression analysis that there is a stable statistical connection between NDVI and yield. This relationship makes forecast models more reliable. The implementation of the model developed for one region in three more locations yielded a strong correlation between actual and predicted outcomes, and the assessment metrics (R^2 and RMSE) validated the model's complete appropriateness for practical use. The fluctuation in RMSE within the range of 0.4–4.3 cents/ha and the relative error of 7% suggest that the modeling is quite precise at the regional level.

The productivity parameters collected from the study area were categorized into five classes, and a geographical assessment was conducted for the regions according to these classifications. This method facilitated a more accurate assessment of production variations, the identification of high and low productivity areas, and the enhancement of agrotechnical decision-making at the regional scale.

The findings highlight that the integration of satellite data with climatic factors functions as a strong analytical method for the operational evaluation of cotton output, identification of field-level variability, and risk management.

This method has extensive use for maximizing resources in agricultural management, planning irrigation methods, and enhancing the precision of yield-related decisions. Future studies will enhance forecast accuracy by incorporating soil characteristics, irrigation intensity, and multi-year NDVI series into the model.

Conflict of Interest

The author declares there is no conflict of interests related to the publication of this article.

REFERENCES

1. **S. Akin.** The Strategic Importance of Cotton Production for the World and Türkiye, 2024. doi: 10.5772/intechopen.114084.
2. **State Statistical Committee of the Republic of Azerbaijan.** 'THE AGRICULTURE OF AZERBAIJAN', Baku, Aug. 2025.
3. **S. Leo, M. Migliorati, and P. Grace.** Predicting within-field cotton yields using publicly available datasets and machine learning, Agron J, vol. 113, Mar. 2021, doi: 10.1002/agj2.20543.
4. **S. Khodjaev, I. Bobojonov, L. Kuhn, and T. Glauben.** An accuracy assessment of satellite-based cotton yield estimation using panel data regression: a case study of Uzbekistan, Environ Dev Sustain, 2024, doi: 10.1007/s10668-024-05220-1.
5. **K. A. Asgerov and E. F. Fatiyeva,** "Assessment of the Parameters of the State of Natural Objects in the Territory of Agdam Region Based on Multi-Temporal Space Images," *Herald of Azerbaijan Engineering Academy*, vol. 14, no. 1, pp. 125–130, Mar. 2022, doi: 10.52171/2076-0515_2022_14_01_125_130.
6. **P. Lang *et al.*,** "Integrating environmental and satellite data to estimate county-level cotton yield in Xinjiang Province," *Front. Plant Sci.*, vol. 13, Jan. 2023, doi: 10.3389/fpls.2022.1048479.

7. **X. Zhu, R. Guo, T. Liu, and K. Xu,** "Crop Yield Prediction Based on Agrometeorological Indexes and Remote Sensing Data," *Remote Sens. (Basel).*, vol. 13, p. 2016, May 2021, doi: 10.3390/rs13102016.
8. **B.M. Azizov, J. S. Mekhtiyev, H. N. Mammadov, and S. H. Guliyeva,** "A Mathematical Model for Accuracy Improvement of Determining the Radiative Characteristics of Objects on the Earth's Surface," *Herald of Azerbaijan Engineering Academy*, vol. 14, no. 2, pp. 117–123, Jun. 2022, doi: 10.52171/2076-0515_2022_14_02_117_123.
9. **A.M. Ali *et al.*,** "Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article)," *The Egyptian Journal of Remote Sensing and Space Sciences*, vol. 25, no. 3, pp. 711–716, 2022, doi: https://doi.org/10.1016/j.ejrs.2022.04.006.
10. **H. Burhan,** "Crop Yield Prediction by Integrating Meteorological and Pesticides Use Data with Machine Learning Methods: An Application for Major Crops in Turkey," *Ekonomi Politika & Finans Araştırmaları Dergisi*, vol. 7, pp. 1–18, Oct. 2022, doi: 10.30784/epfad.1148948.
11. **H. Wang, Y. Dai, Q. Yao, L. Ma, Z. Zhang, and X. Lv,** "Multi-task learning model driven by climate and remote sensing data collaboration for mid-season cotton yield prediction," *Field Crops Res.*, vol. 333, p. 110070, 2025, doi: https://doi.org/10.1016/j.fcr.2025.110070.
12. **M. Drusch *et al.*,** "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, Feb. 2012, doi: 10.1016/j.rse.2011.11.026.
13. **A. Badalova and B. Babashli,** "Peyk təsviri əsasında vegetasiya indekslərindən istifadə edərək pambıq sahələrinin monitorinqi," *Scientific Journal (National Aviation Academy)*, vol. 25, no. 4, pp. 23–30, Dec. 2024, doi: 10.30546/emnaa.2023.25.4.3.
14. **G. James, D. Witten, T. Hastie, and R. Tibshirani,** *An Introduction to Statistical Learning*. 2013.
15. **Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining,** *Introduction to Linear Regression Analysis, Fifth Edition.*, 5th ed. 2012: A JOHN WILEY & SONS, INC, 2012. doi: 10.1111/insr.12020_10.
16. **T. Mn *et al.*,** "Comparative analysis of machine learning and statistical models for cotton yield prediction in major growing districts of Karnataka, India," *Journal of Cotton Research*, vol. 8, Feb. 2025, doi: 10.1186/s42397-024-00208-8.
17. **A. Nazir *et al.*,** "Estimation and Forecasting of Rice Yield Using Phenology-Based Algorithm and Linear Regression Model on Sentinel-II Satellite Data," *Agriculture*, vol. 11, Oct. 2021, doi: 10.3390/agriculture11101026.