



Nonlinear Effects of Agricultural Land and Value Added on Freshwater Withdrawals in Azerbaijan: An XGBoost–SHAP Analysis

Anar Eminov¹, Ramil I. Hasanov^{1,2*}, Jeyhun Mahmudov¹, Rodolfo M. Nayga Jr.³,
Abdullhuseyn Zamanov⁴

¹ Department of Business and Management, Mingachevir State University, 4500 Mingachevir, Azerbaijan

² UNEC Research Center on Global Environmental Issues, Azerbaijan State University of Economics (UNEC), AZ 1001 Baku, Azerbaijan

³ Department of Agricultural Economics, Texas A&M University, College Station, TX 77843 Texas, United States

⁴ Department of Economics, Nakhchivan State University, AZ7012 Nakhchivan, Azerbaijan

* Correspondence: Ramil I. Hasanov (ramil.hasanov@mdu.edu.az)

Received: 12-30-2025

Revised: 04-13-2026

Accepted: 04-25-2026

Citation: Eminov, A., Hasanov, R. I., Mahmudov, J., Nayga, R. M., Jr. & Zamanov, A. (2026). Nonlinear effects of agricultural land and value added on freshwater withdrawals in Azerbaijan: An XGBoost–SHAP analysis. *Chall. Sustain.*, 14(2), 424–437. <https://doi.org/10.56578/cis140213>.



© 2026 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: Effective management of freshwater resources in agriculture is essential for ensuring sustainable economic development and environmental resilience, particularly in transitional economies such as Azerbaijan. Over the period 2000–2021, agricultural land area in Azerbaijan exhibited a steady increase, while the sector’s contribution to GDP declined, indicating structural transformation and potential inefficiencies in resource utilization. This study investigates the nonlinear effects of agricultural land use and agricultural value added on freshwater withdrawals using an interpretable machine learning framework. Specifically, Extreme Gradient Boosting (XGBoost) is employed to model complex relationships, while Shapley Additive Explanations (SHAP) quantify feature importance and elucidate threshold and asymmetric effects. The analysis draws on annual country-level data integrating national and international statistics to ensure temporal consistency and comparability. Results indicate that agricultural land area constitutes the dominant driver of freshwater withdrawals, contributing 57% of the model’s predictive gain, whereas agricultural value added accounts for 43%. SHAP dependence plots reveal pronounced nonlinearities: moderate land expansion exacerbates freshwater stress, whereas allocations beyond a critical threshold mitigate pressure, reflecting potential efficiency gains at scale. Agricultural value added exhibits a U-shaped relationship, wherein both low and high productivity levels are associated with increased freshwater use, while intermediate productivity generates the greatest negative impact. The XGBoost model achieves substantial predictive performance (Coefficient of Determination (R^2) = 0.78, Root Mean Squared Error (RMSE) = 0.806, Mean Absolute Percentage Error (MAPE) = 0.86%), demonstrating its capacity to capture heterogeneous, nonlinear dynamics that linear models fail to detect. The robustness of the model was further assessed using Leave-One-Out Cross-Validation (LOOCV) to evaluate its out-of-sample predictive performance and mitigate potential overfitting arising from the limited sample size. These findings underscore the necessity of adaptive water management strategies that incorporate scale-dependent effects and productivity heterogeneity. Policies optimizing land allocation and promoting efficient agricultural practices can enhance water-use efficiency while sustaining sectoral output. The study highlights the value of interpretable machine learning in advancing empirical understanding of the water–agriculture nexus under conditions of structural economic change.

Keywords: Sustainability; Agriculture; Freshwater; Land; Agricultural value added; Machine learning; XGBoost; SHAP analysis

1. Introduction

Freshwater resources are vital for sustaining agricultural productivity, ensuring food security, supporting industrial and economic development, and maintaining ecological balance by preserving biodiversity and the

integrity of natural ecosystems (Khondoker et al., 2023). Although agricultural land area has grown over the past two decades, the sector's declining contribution to GDP highlights inefficiencies and underscores the need to understand the complex interactions between land allocation, productivity, and water demand (Niftiyev, 2020).

Water resources management has emerged as a critical global challenge, as climate change-induced pressures on freshwater availability increasingly threaten environmental sustainability and socioeconomic stability. Recognized as a national strategic priority in many countries, water governance is formally embedded in Sustainable Development Goal 6 of the United Nations Sustainable Development Goals, which emphasizes the efficient, sustainable, and inclusive management of water resources (United Nations, 2015). In this context, Azerbaijan faces acute water stress, with a long-term average Water Exploitation Index exceeding 52.7% and peaking at 78% in 2019, while agriculture and irrigation dominate water use, accounting for over 90% of total withdrawals. Moreover, between 2000 and 2019, total water abstraction increased by 15% despite a 30.4% reduction in renewable water resources, intensifying reliance on surface water and resulting in a fourfold expansion in groundwater utilization (METEO, 2022). Agriculture represents the largest global demand on freshwater resources, utilizing nearly 70% of the world's total freshwater withdrawals, which underscores its central role in water resource management and the critical need for sustainable irrigation practices (Azertag, 2023).

Efficient utilization of agricultural land is essential for sustainable food security and rural development, yet in Azerbaijan, substantial gaps between potential and actual land productivity persist due to ecological degradation, institutional fragmentation, and outdated management systems. Although 59.2% of Azerbaijan's territory is classified as agricultural land, only 30% to 65% of its productive potential is realized, with the greatest spatial disparities occurring in the Kura–Araz lowland and foothill plains, where salinization, erosion, and inadequate irrigation limit crop yields (CEIC, 2021; Valiyev et al., 2025). A core principle of sustainable development is adherence to environmental regulations alongside the efficient utilization of natural resources. In this context, the strategic development of the agricultural sector is particularly significant, as its performance is directly contingent on the availability of freshwater resources. While the study of water use in agriculture is globally important, empirical research in Azerbaijan remains limited.

The analysis is based on a limited set of variables due to data availability constraints, so it does not fully capture all key drivers of freshwater withdrawals such as climate, population, technology, and governance factors. Therefore, the results should be understood as conditional relationships rather than complete causal effects, and future studies could improve robustness by including these additional determinants.

Traditional linear econometric techniques, including OLS (Ordinary Least Squares) and ARDL (Autoregressive Distributed Lag), are constrained in their ability to capture nonlinear interactions and heterogeneous effects, whereas interpretable machine learning approaches provide a powerful alternative for modeling complex relationships. This study utilizes an AI-driven machine learning framework to examine the nonlinear influences of agricultural land use and value added on freshwater withdrawals, uncovering threshold behaviors and quantifying the relative contribution of key factors, thereby offering actionable insights for evidence-based water management amid structural economic transformation.

2. Literature Review

Previous research has underscored the critical interconnection between agricultural management, water resource utilization, and food security. Sarwar (2025) highlighted the significance of agricultural value-added and the strategic management of water reservoirs in strengthening food security in the United States, demonstrating that economic growth, land-use practices, and technological innovations collectively supported a stable food supply. In a complementary study, Eslamifar et al. (2024) applied system dynamics modeling to assess a fallowing strategy in the Mesilla–Rincon Valley, revealing that temporarily leaving cropland fallow markedly reduced surface and groundwater withdrawals for cotton, alfalfa/hay, and Chile, with only marginal reductions in agricultural income. Similarly, Li et al. (2020) proposed a comprehensive optimization framework for allocating agricultural water and land under uncertainty, effectively reconciling economic, environmental, and social objectives while accounting for non-linearities and fuzzy uncertainties to facilitate sustainable resource management. Collectively, these studies demonstrated that the integration of economic, environmental, and technological strategies with deliberate water and land-use planning is essential for advancing sustainability and conserving resources without substantially undermining agricultural productivity.

In recent years, AI-driven machine learning approaches have become increasingly central to environmental and resource management research, offering advanced capabilities to capture complex, nonlinear, and threshold interactions among ecological, climatic, and anthropogenic factors. Studies in China have demonstrated the effectiveness of these approaches in diverse contexts: Zhou et al. (2024) revealed that land use changes, particularly the conversion of cropland to construction land, along with variations in rainfall and sunshine duration, significantly influenced habitat quality, soil retention, water yield, and carbon fixation in the Karst region, with Extreme Gradient Boosting–Shapley Additive Explanations (XGBoost–SHAP) models identifying critical nonlinear and threshold effects to inform ecological conservation strategies. Similarly, Li et al. (2025) applied the

same interpretable machine learning framework to assess seasonal water quality in Tai Lake Basin, highlighting dissolved oxygen, total phosphorus, permanganate index (COD_{Mn}), and ammonia nitrogen as key determinants with distinct seasonal impacts, offering actionable guidance for freshwater management. Complementing these ecosystem-focused studies, Wu et al. (2025) demonstrated that environmental regulation significantly enhanced urban ecological resilience in the Yangtze River Delta once regulatory intensity exceeded 0.8%, with additional synergistic effects from industrial structure, urbanization, and public investment, and XGBoost–SHAP effectively elucidated the nonlinear and interactive mechanisms underlying these outcomes. Collectively, these findings underscore the growing value of interpretable machine learning in capturing the complex dynamics of ecosystem services, water quality, and urban resilience, providing robust empirical evidence to guide sustainable resource management and policy development across diverse environmental contexts in China.

Recent studies highlighted the use of machine learning and explainable AI to analyze complex interactions between human activities, climate variability, and water or ecosystem resources, providing actionable insights for sustainable management. Applications included assessing groundwater drought responses, long-term river discharge, and dietary water footprints, with key drivers such as land use, climate teleconnections, socioeconomic factors, and demographic attributes identified to guide targeted interventions (Banda et al., 2025; Huang et al., 2025; Poursaeid, 2025). Nevertheless, the application of machine learning–based modeling to the analysis of water management and sustainable agricultural systems in Azerbaijan has remained largely underexplored, indicating a notable gap in the existing literature. Accordingly, this article sought to bridge this gap by applying advanced machine learning techniques to examine the complex interrelationships among water resources, agricultural activities, and environmental factors within the Azerbaijani context.

3. Data and Methodology

3.1 Data Formation

The study employs annual country-level data for Azerbaijan spanning 2000–2021, integrating both national statistics and international indicators to ensure consistency, coverage, and comparability over time. The dataset links freshwater withdrawals in agriculture with measures of agricultural land area and sectoral value added, facilitating an examination of the interplay between physical land expansion, economic performance, and water use dynamics throughout the study period.

The variables and their measurement are as follows:

- Dependent variable: Freshwater withdrawals (FR_WATER) in agriculture. This represents the total annual volume of freshwater extracted for agricultural purposes and serves as the primary outcome for prediction and interpretation of sector-level water use trends.
- Independent variable: Agricultural land use (AGR_LAND). Defined as the proportion of total land area designated for agriculture, this series exhibits low variability and approximates a normal distribution, indicating relative structural stability in land utilization over time.
- Independent variable: Agricultural value added (AGR_VA). Capturing the sectoral contribution of agriculture, forestry, and fishing, AGR_VA is measured either in national currency or constant-price 2015 USD, depending on data harmonization. The series is right-skewed with higher dispersion than AGR_LAND, reflecting greater volatility in sectoral performance. AGR_VA is measured in constant prices to ensure temporal comparability and to eliminate distortions arising from inflationary pressures and price-level fluctuations over the study period (2000–2021). This adjustment allows for a more consistent assessment of real sectoral performance and its relationship with freshwater withdrawals.

The dataset primarily draws on the World Bank (World Bank Group, 2025) and the State Statistical Committee of the Republic of Azerbaijan (SSCRA, 2025), combining these sources to ensure comprehensive coverage and cross-temporal comparability of agricultural and water-use indicators.

The analysis is constrained by a small sample size of 22 annual observations, which limits the suitability of machine learning methods for strong predictive performance. Accordingly, XGBoost is used mainly to identify nonlinear patterns within the observed data, and the results are interpreted cautiously without extending them to broad out-of-sample generalizations.

Figure 1 illustrates the temporal trends in agricultural land use, freshwater withdrawals, and sectoral GDP contribution. The analysis indicates a steady increase in agricultural land area alongside a fluctuating but overall rising dependence on freshwater, reflecting intensifying resource utilization over the twenty-year period. In contrast, the declining share of agriculture, forestry, and fishing in total GDP highlights a structural transformation, where sectoral value-added has contracted despite continued physical expansion in agricultural activities.

Although the STROBE reporting checklist is not formally applied in this study, the reporting of data sources, variable construction, and data preprocessing procedures is aligned with its core principles of transparency and reproducibility (Von Elm et al., 2007). This approach ensures that the process of dataset construction is clearly documented and can be readily traced. As a result, methodological transparency is strengthened, enhancing the

overall clarity and credibility of the observational econometric and machine learning framework employed.

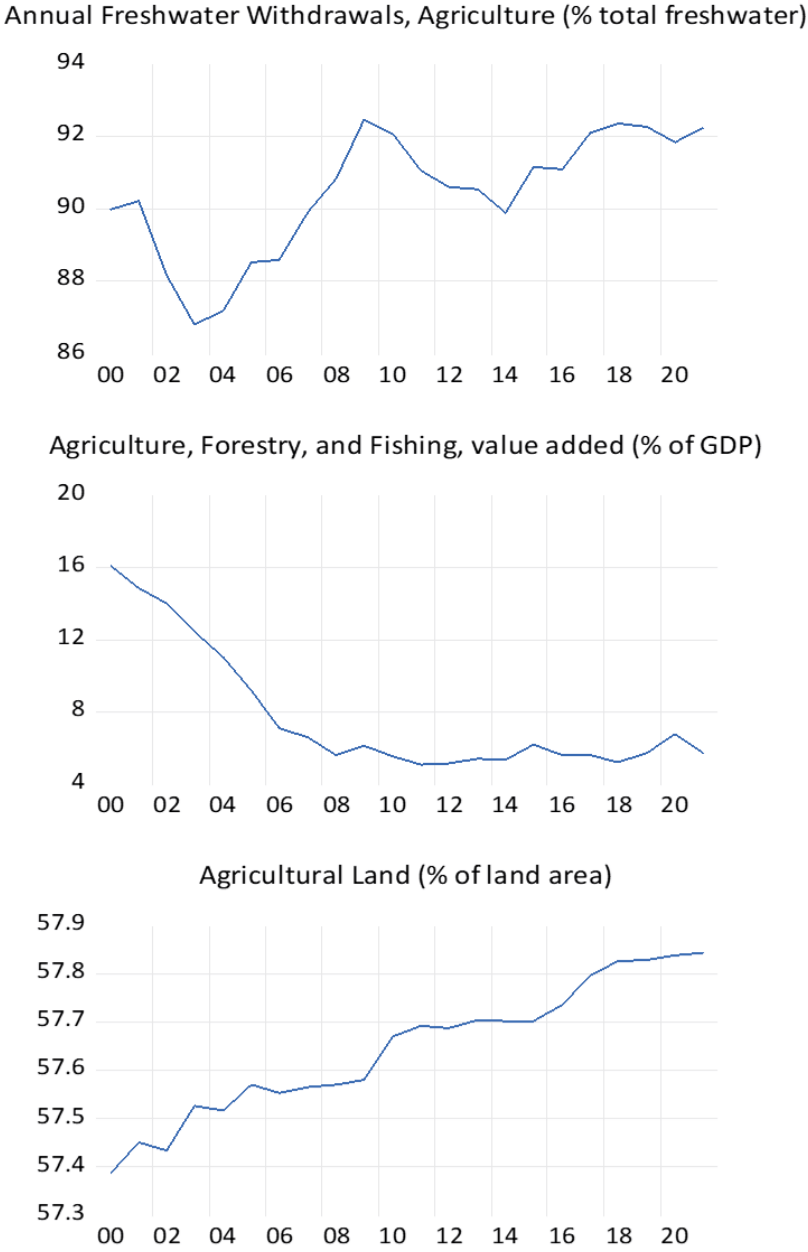


Figure 1. Trends in agricultural land use, freshwater withdrawals, and sectoral contribution to GDP (2000–2021)

3.2. Methods

This study adopts a machine learning framework to investigate the nonlinear impacts of AGR_LAND and AGR_VA on freshwater withdrawals within Azerbaijan’s agricultural sector. The modeling strategy combines XGBoost with SHAP. XGBoost, introduced by Chen & Guestrin (2016), is a scalable, regularized gradient boosting algorithm known for its high predictive performance, while SHAP, developed by Lundberg & Lee (2017), provides consistent and locally accurate feature attributions based on cooperative game theory principles.

The analysis follows a structured workflow:

- Data preprocessing: The dataset undergoes cleaning, normalization, and imputation to address missing values and ensure data quality.
- Model training: AGR_LAND and AGR_VA are used as input features, while FR_WATER is defined as the target variable for model training.
- Performance evaluation: Model performance is assessed using standard metrics, including Root Mean

Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R²), and Mean Absolute Percentage Error (MAPE).

- Model interpretation: SHAP values are computed to quantify the relative importance of predictors and to visualize their nonlinear effects through summary and dependence plots.

This approach enables a rigorous, data-driven assessment of how land use patterns and sectoral performance influence agricultural water demand, capturing both linear and nonlinear interactions. The XGBoost algorithm seeks to minimize a regularized objective function that integrates both the model's training loss and a complexity penalty, thereby controlling overfitting. Formally, this objective can be expressed as:

$$Obj(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

$$\Omega(f_k) = \gamma T + \left(\frac{1}{2}\right) \lambda ||w||^2 \quad (2)$$

In this context, $l(y_i, \hat{y}_i)$ represents the loss function for the i observation, f_k denotes the k decision tree in the ensemble, T is the total number of leaves within a tree, and w is the vector of weights assigned to each leaf. The parameters γ and λ serve as regularization terms that penalize model complexity to reduce overfitting. SHAP values are based on Shapley's cooperative game theory (Shapley, 1953) and adapted for machine learning to quantify the contribution of each feature to individual predictions. For a given feature i , the SHAP value is calculated as:

$$\phi_{i(f,x)} = \sum_{S \subseteq M \setminus \{i\}} \frac{|S|! (M - |S| - 1)!}{M!} (f_x(S \cup \{i\}) - f_x(S)) \quad (3)$$

where, M denotes the total number of features in the dataset, S represents a subset of features that excludes the i feature, and $f(S)$ corresponds to the model's prediction when only the features in subset S are used.

SHAP values are computed for interpretability purposes using the full dataset. They are not used for predictive validation, which is instead conducted using LOOCV (Leave-One-Out Cross-Validation). LOOCV is applied to evaluate the out-of-sample predictive performance of the XGBoost model given the limited sample size. This approach ensures that each observation was sequentially used as a validation point, thereby providing a robust assessment of model stability and reducing the risk of overfitting in a small-sample setting.

4. Results

The correlation results in Table 1 reveal a strong negative relationship between agricultural land use and agricultural value added, indicating that increases in cultivated land are associated with lower productivity intensity during the study period. This marked interdependence points to underlying structural trade-offs and underscores the suitability of nonlinear modeling frameworks, as conventional linear specifications may be inadequate for capturing such complex dynamics.

Table 1. Covariance and correlation matrix (2000–2021, $n = 22$)

Variables	AGR_LAND	AGR_VA
Panel A: Covariance		
AGR_LAND	0.018172	-0.368897
AGR_VA	-0.368897	11.80199
Panel B: Correlation		
AGR_LAND	1.000000	-0.796583
AGR_VA	-0.796583	1.000000

Note: agricultural land use (AGR_LAND); agricultural value added (AGR_VA)

Table 2 indicates that the XGBoost model exhibits robust predictive capability, accounting for nearly 78% of the variability in freshwater resources. The model's low RMSE and MAE values reflect minimal prediction error, while a MAPE below 1% underscores both high accuracy and stability. Notably, achieving an R² of approximately 0.78 with only two predictors and 22 observations highlights the model's remarkable explanatory power within this dataset.

Table 3 and Figure 2 present the feature importance analysis, quantifying the relative contribution of each agricultural variable to the model's predictive performance. The results demonstrate that agricultural land area is the most influential predictor of the target outcome, surpassing agricultural value added in terms of information gain during the decision-tree splitting process.

Table 2. Model performance

Metric	Value	Interpretation
RMSE	0.806	Low prediction error relative to the scale of freshwater
MAE	0.780	Stable average absolute deviation
R ²	0.783	Strong explanatory power for a nonlinear machine learning model
MAPE	0.86%	Excellent predictive accuracy

Note: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R²), Mean Absolute Percentage Error (MAPE)

Table 3. Feature importance

Variable	Gain	Frequency
AGR_LAND	0.569	0.490
AGR_VA	0.431	0.510

Note: agricultural land use (AGR_LAND); agricultural value added (AGR_VA)

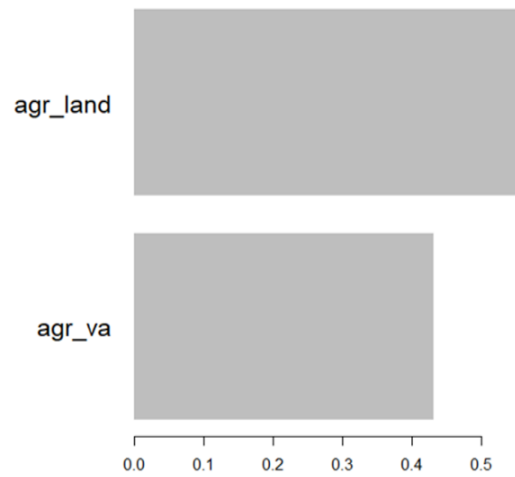


Figure 2. Feature importance of predictors in the XGBoost test dataset

Note: agricultural land use (AGR_LAND); agricultural value added (AGR_VA); Extreme Gradient Boosting (XGBoost)

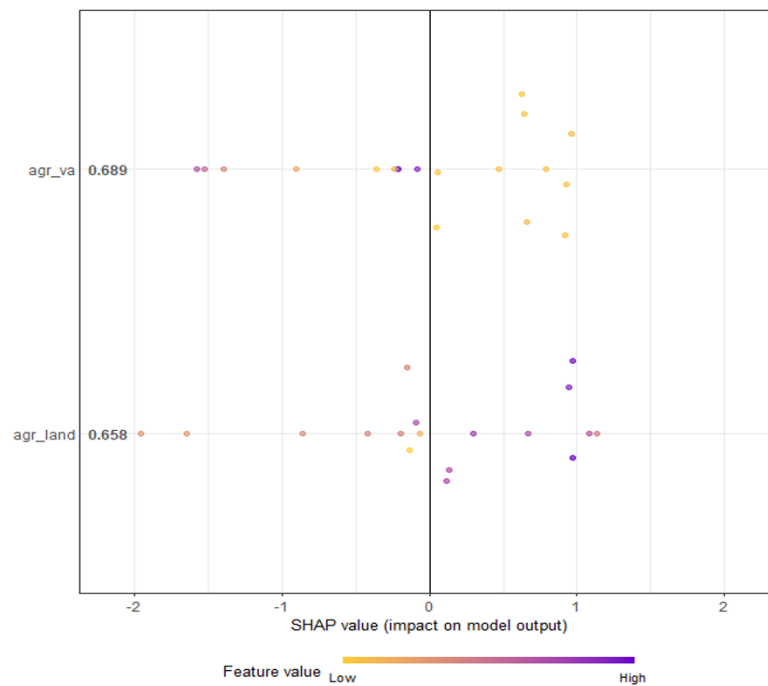


Figure 3. SHAP summary plot of agricultural predictors on freshwater resources

Note: agricultural land use (AGR_LAND); agricultural value added (AGR_VA); Shapley Additive Explanations (SHAP)

Figure 3 presents the SHAP summary plot, showing that both AGR_VA and AGR_LAND have substantial and roughly comparable impacts on the predicted freshwater resources. AGR_VA exhibits a slightly higher mean absolute SHAP value, indicating marginally greater overall importance. The color gradient illustrates that higher values of both variables are generally associated with positive SHAP contributions, suggesting that increases in agricultural value added and land area tend to elevate the model’s predicted freshwater withdrawals. The spread of SHAP values around zero highlights heterogeneous and nonlinear effects, indicating that the influence of agricultural activity on freshwater resources varies across observations rather than being uniform. The SHAP summary plot represents global feature attribution across the full dataset rather than a train–test split.

Figure 4 presents the SHAP dependence plots for AGR_LAND and AGR_VA, revealing pronounced nonlinear relationships with freshwater resources. For AGR_LAND, moderate land area is initially associated with a decrease in predicted freshwater withdrawals, whereas values beyond a threshold correspond to positive effects, indicating potential land-use tipping points. This pattern suggests that limited agricultural expansion may intensify pressure on freshwater systems, while larger land allocations may enhance water outcomes, possibly through more efficient management. The dependence plot for AGR_VA exhibits a U-shaped relationship, with both low and high levels contributing positively to freshwater resources, whereas intermediate productivity levels produce the most negative impact. These nonlinear trajectories emphasize the complex adjustment mechanisms in the water–agriculture nexus, highlighting the limitations of linear models in capturing threshold and rebound effects.

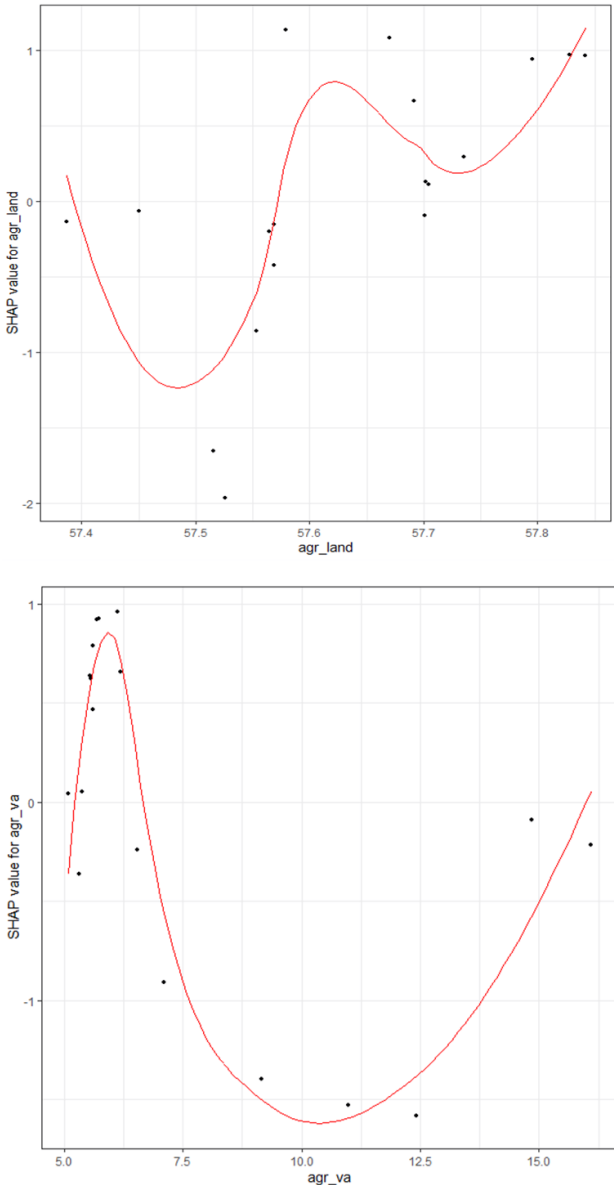


Figure 4. Nonlinear effects of agricultural land use and agricultural value added on model outcomes: A Shapley Additive Explanations (SHAP)-based dependence analysis
 Note: agricultural land use (AGR_LAND); agricultural value added (AGR_VA)

Feature importance analysis demonstrates that agricultural land use represents the primary driver of freshwater resource predictions, with agricultural value added remaining influential but secondary. The high frequency of both variables across decision trees indicates their consistent contribution to model performance. SHAP summary results further reveal that variations in AGR_LAND exert the strongest global effect on freshwater withdrawals, while AGR_VA provides a secondary yet notable impact, with the dispersion of SHAP values highlighting heterogeneous and nonlinear effects across observations. Dependence plots show that AGR_LAND follows a smooth nonlinear trajectory, suggesting diminishing marginal effects beyond specific land-use thresholds, whereas AGR_VA exhibits a weaker, asymmetric relationship, indicating that increases in agricultural productivity do not uniformly translate into enhanced freshwater availability. These nonlinear and observation-specific effects underscore the limitations of traditional linear approaches, such as OLS or ARDL, which cannot capture threshold behavior or variable-specific heterogeneity. Overall, the XGBoost–SHAP framework provides a nuanced understanding of the water–agriculture nexus, revealing complex dynamics that linear models are unable to detect.

The LOOCV results in Table 4 show that the XGBoost model achieves a moderate-to-strong out-of-sample performance, with an R^2 of 0.663, indicating that about 66% of the variation in freshwater withdrawal is explained by agricultural land and agricultural value added. The relatively low RMSE (1.003) and MAE (0.853) suggest that prediction errors are small and consistent across all leave-one-out iterations. Overall, these findings indicate that the model is stable and not driven by individual yearly observations, supporting the robustness of the estimated nonlinear relationships.

In Figure 5, The scatter plot exhibits a pronounced positive linear relationship between the observed and predicted values, indicating that the model adequately explains a substantial proportion of the variability in freshwater withdrawal ($R^2 = 0.663$). The clustering of observations around the implicit 45-degree identity line further suggests a satisfactory level of model consistency. This pattern implies that the explanatory variables retain meaningful predictive capacity beyond the estimation sample.

Table 4. Leave-One-Out Cross-Validation (LOOCV) performance of Extreme Gradient Boosting (XGBoost) model

Metric	Value	Interpretation
RMSE	1.003	Average prediction error in original units of freshwater withdrawals
MAE	0.853	Mean absolute deviation between observed and predicted values
R^2	0.663	Proportion of variance explained in out-of-sample prediction

Note: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2)

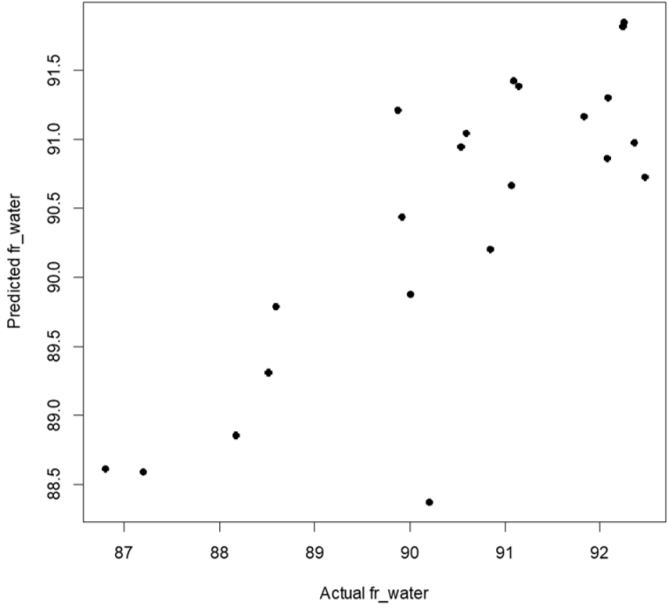


Figure 5. Actual vs. predicted values
Note: freshwater withdrawals (FR_WATER)

In Figure 6, the line chart presents the temporal pattern of absolute errors, indicating that the residuals are randomly distributed over time rather than exhibiting systematic bias or a discernible temporal trend. The noticeable decline in error variability in the period 2011–2021 further suggests a greater stabilization of the underlying relationship between agricultural indicators and water consumption during the later part of the sample period.

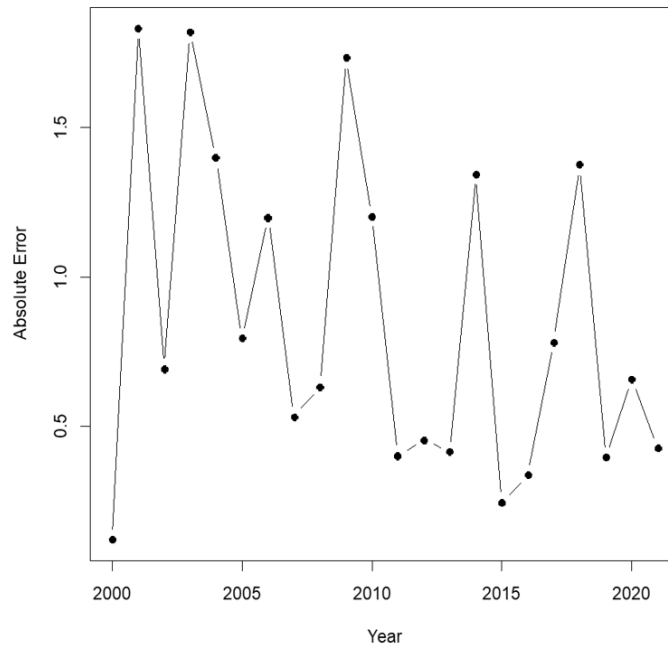


Figure 6. Leave-One-Out Cross-Validation (LOOCV) error by year

To complement the machine learning analysis, a traditional econometric framework is also applied to ensure robustness and provide broader empirical validation. Table 5 indicates that FR_WATER and AGR_LAND are non-stationary in levels but become stationary after first differencing, implying that both variables are integrated of order I(1), while AGR_VA is stationary at level (I(0)). Table 6 further confirms the existence of a stable long-run relationship among the variables, as the ARDL bounds test F-statistic (6.7066) exceeds the upper critical values at the 1% significance level, providing strong evidence of cointegration.

Table 5. Augmented Dickey-Fuller unit root test results

Variable	Level (t-stat)	Level (p-value)	First Difference (t-stat)	First Difference (p-value)
FR_WATER	-1.035	0.720	-3.260	0.031
AGR_VA	-4.031	0.005	-	-
AGR_LAND	-1.291	0.613	-4.830	0.001

Note: agricultural land use (AGR_LAND); agricultural value added (AGR_VA); freshwater withdrawals (FR_WATER)

Table 6. ARDL bounds cointegration test and long-run estimation results

Panel	Variable / Statistic	Coefficient	Std. Error	t-Statistic	Prob.	Interpretation
A. Bounds Test for Cointegration	F-Statistic	6.7066	-	-	-	Cointegration Confirmed
	Critical Value 10% I(0) / I(1)	2.915 / 3.695	-	-	-	
	Critical Value 5% I(0) / I(1)	3.538 / 4.428	-	-	-	
	Critical Value 1% I(0) / I(1)	5.155 / 6.265	-	-	-	
B. Long-Run Coefficients	AGR_LAND	4.6699	2.7739	1.6835	0.1105	Positive but Insignificant
	Constant	-176.7952	160.7049	-1.1001	0.2866	-
C. Adjustment Parameter	AGR_VA(-1)	-0.2455	0.1521	-1.6143	0.1249	Negative Adjustment Speed

Note: Bounds test critical values are based on Narayan (2005). The F-statistic exceeds the upper bound critical value at the 1% significance level, confirming long-run cointegration. Long-run coefficients are derived from the conditional error correction representation of the ARDL (Autoregressive Distributed Lag) model. Agricultural land use (AGR_LAND); agricultural value added (AGR_VA)

Figure 7 presents the Cumulative Sum of Recursive Residuals (CUSUM) and CUSUM of squares stability tests for the estimated model, with both test statistics remaining within the 5% critical boundaries over the entire sample period. These findings suggest that the model coefficients are stable across time and that no meaningful structural breaks or parameter instability are detected during the study period.

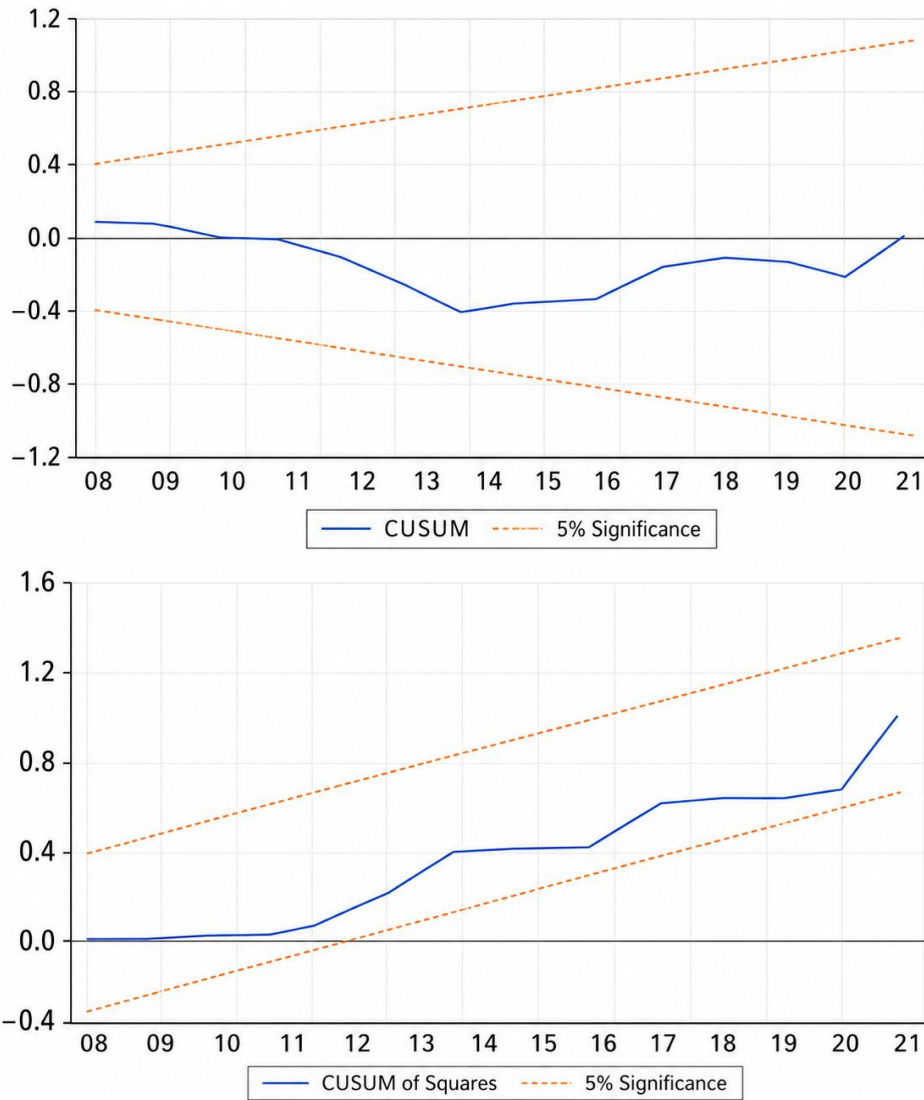


Figure 7. The Cumulative Sum of Recursive Residuals (CUSUM) and CUSUM of squares diagnostic plots

The comparative results in Table 7 indicate that the ARDL model achieves lower in-sample error (RMSE = 0.536; MAE = 0.410), reflecting its efficiency in capturing linear relationships under small-sample econometric assumptions. In contrast, XGBoost produces higher error values but is designed to capture nonlinear and threshold effects, making both approaches complementary rather than directly comparable in terms of predictive superiority.

Hyperparameter tuning reported in Table 8 was carried out through a grid-search procedure integrated with 5-fold cross-validation within the XGBoost framework. Model selection was based on minimizing the RMSE across validation folds. The best-performing specification was obtained with a learning rate of 0.05, a maximum tree depth of 2, and 97 boosting rounds. Both subsampling and column sampling ratios were fixed at 0.8, while L2 regularization and the split penalty parameter were retained at their default levels to ensure model robustness given the limited sample size.

Table 7. Comparative model performance: ARDL vs XGBoost

Metric	ARDL	XGBoost
RMSE	0.536	0.806
MAE	0.410	0.780
Model Type	Linear econometric (ARDL)	Nonlinear machine learning (XGBoost)
Key Strength	Efficient linear inference	Captures nonlinear and threshold effects
Limitation	Misses nonlinear structure	Higher in-sample error in small samples

Note: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Autoregressive Distributed Lag (ARDL), Extreme Gradient Boosting (XGBoost)

Table 8. Optimized hyperparameters of Extreme Gradient Boosting (XGBoost) model

Hyperparameter	Value
n_estimators (nrounds)	97
learning_rate (eta)	0.05
max_depth	2
subsample	0.8
colsample_bytree	0.8
min_child_weight	1
gamma	0
objective	reg:squarederror

5. Discussion

The results reveal a complex, nonlinear interplay between agricultural expansion, sectoral productivity, and freshwater withdrawals in Azerbaijan. While agricultural land use has remained relatively stable, the declining share of agriculture in GDP reflects ongoing structural transformation, suggesting that physical expansion does not necessarily yield proportional economic gains. The pronounced negative correlation between land area and value added highlights trade-offs in productivity intensity, consistent with prior evidence on resource-use efficiency in transitional economies. The XGBoost–SHAP framework proves effective in capturing these nonlinear dynamics, achieving robust predictive performance even with limited data. Feature importance analysis identifies agricultural land as the principal driver of freshwater withdrawals, whereas SHAP dependence plots reveal threshold and rebound effects. Specifically, moderate land expansion exacerbates water stress, while larger allocations appear to alleviate pressure, potentially reflecting efficiency gains at scale. Agricultural value added exhibits a U-shaped relationship, in which both low and high productivity levels are associated with increased water use, whereas intermediate productivity intensifies water stress. These heterogeneous, nonlinear effects would likely remain obscured under conventional linear models, underscoring the methodological advantage of machine learning for resource–economic analyses.

Hybrid and optimization-enhanced machine learning frameworks, particularly those combining boosting algorithms with metaheuristic or physics-inspired optimization methods, improve robustness, reduce hyperparameter sensitivity, and enhance generalization in small-sample settings, as evidenced in recent engineering applications such as strain and deformation prediction models. In parallel, physics-informed and domain-guided learning approaches strengthen interpretability by embedding system-specific knowledge from fields such as irrigation mechanics, land productivity, and hydrological processes, thereby improving both predictive reliability and structural consistency. Within this context, the present study aligns with emerging hybrid ML paradigms and suggests that integrating XGBoost–SHAP with optimization and domain-constrained learning could further enhance robustness, reduce uncertainty, and improve policy relevance in water–agriculture sustainability assessments.

The XGBoost–SHAP framework suggests that future freshwater withdrawals in Azerbaijan’s agricultural sector are shaped by complex, nonlinear, and threshold-dependent relationships between land expansion and agricultural value added. While moderate increases in cultivated land tend to intensify water stress, expansion beyond a certain threshold can reduce pressure, indicating potential efficiency improvements through economies of scale, enhanced irrigation infrastructure, and integrated water management (Grafton et al., 2018). Likewise, agricultural value added demonstrates a U-shaped relationship with water use: both low and high productivity levels are associated with increased freshwater consumption, whereas intermediate productivity exacerbates water stress, underscoring the asymmetric and heterogeneous dynamics of resource–productivity interactions. The nonlinear dynamics observed indicate that traditional linear econometric approaches are inadequate for capturing tipping points and rebound effects within water–agriculture systems, particularly in transitional economies experiencing structural change. The significant role of agricultural land in driving freshwater withdrawals highlights the necessity of policy measures that enhance land-use efficiency, such as land consolidation initiatives and reforms to irrigation pricing, which can internalize scarcity costs and promote optimal water utilization without limiting agricultural output (Djanibekov et al., 2012). Furthermore, given the U-shaped relationship between productivity and water use, targeted support for water-saving technologies, such as drip irrigation and precision farming, is essential, as mere productivity growth does not automatically translate into reduced water consumption (Ward & Pulido-Velazquez, 2008). The identified U-shaped relationship between agricultural value added and freshwater withdrawals can be understood through the framework of the Jevons Paradox. This concept posits that gains in resource-use efficiency may unintentionally lead to higher overall consumption as a result of expansion and rebound effects (Alcott, 2005). In the case of Azerbaijan, improvements in agricultural productivity at more advanced stages of development may encourage further intensification of production and expansion of cultivated land, thereby increasing aggregate water demand despite efficiency gains. This perspective offers a theoretical explanation for the nonlinear and asymmetric patterns observed in the SHAP-based analysis.

Overall, the study provides empirical support for Azerbaijan's ongoing national water and agricultural reforms and aligns with international commitments. By integrating interpretable machine learning with policy-relevant scenario analysis, the research demonstrates how AI-driven decision-support tools can translate complex, nonlinear relationships into actionable policy recommendations, enhancing both the targeting and effectiveness of sustainable water management strategies under conditions of climatic and structural uncertainty. The findings suggest that agricultural water management cannot be adequately captured by linear trends alone. Nonlinear thresholds and asymmetric responses underscore the need for adaptive policies that account for scale effects and productivity heterogeneity. By integrating AI-driven predictive modeling with interpretable analytical frameworks, this study advances the empirical understanding of the water–agriculture nexus under conditions of structural economic change. Agricultural land emerges as the dominant determinant of freshwater withdrawals, while agricultural value added exerts nonlinear, U-shaped effects.

SHAP values should be interpreted as model-derived, associative contributions rather than causal impacts. As a result, the detected nonlinearities and threshold effects reflect patterns learned within the dataset and describe how the model behaves under given conditions.

The LOOCV results, along with the diagnostic plots, indicated that the XGBoost model demonstrated satisfactory out-of-sample predictive performance, with relatively low error metrics and a stable fit between observed and predicted values. The absence of systematic patterns in the residuals further suggested that the estimated nonlinear relationships were not primarily driven by individual annual observations or structural bias.

These findings align with comparable studies employing XGBoost–SHAP to investigate land–water interactions. Du et al. (2025) examined ecosystem service trade-offs in ecologically sensitive regions, demonstrating that land-use expansion produces nonlinear effects with thresholds beyond which resource efficiency improves. This pattern is mirrored in Azerbaijan, where moderate land expansion increases water stress, while larger allocations can mitigate impacts through efficiency gains. Additionally, the nonlinear relationship between freshwater withdrawals and agricultural land use can be partly attributed to soil salinization, a prevalent constraint in irrigated areas such as the Kura–Araz lowland. Increasing salinity levels reduce soil productivity, which in turn requires higher irrigation inputs to sustain agricultural yields. As a result, water demand tends to rise even under conditions of efficiency improvements or stable land expansion. This environmental constraint supports the view that land expansion, in isolation, does not necessarily lead to efficiency gains and may instead contribute to greater water stress in the presence of soil degradation. Similarly, Zhao et al. (2025) analyzed cropland water-use efficiency in Xinjiang, identifying heterogeneous nonlinear relationships between productivity and water efficiency, where both low and high productivity levels were associated with improved outcomes, whereas intermediate productivity heightened stress. Collectively, these parallels suggest that nonlinear thresholds and rebound effects are a pervasive phenomenon in agricultural water management, reinforcing the generalizability of the present findings and highlighting the advantages of interpretable machine learning over traditional linear econometric approaches.

6. Conclusion

The findings indicate that the relationship between agricultural activity and freshwater resources is characterized by nonlinear and heterogeneous dynamics. The correlation analysis reveals a strong negative association between agricultural land use and agricultural value added, pointing to structural trade-offs in the agricultural system during the study period. This interdependence suggests that changes in land expansion and productivity intensity interact in complex ways, motivating the application of modern, nonlinear modeling techniques. In this context, the XGBoost framework demonstrates robust predictive performance, explaining nearly 78% of the variation in freshwater resources with a limited number of predictors and observations, which highlights its effectiveness in capturing complex adjustment processes. The feature importance results show that agricultural land use represents the dominant contributor to freshwater resource predictions, while agricultural value added remains a consistently influential secondary factor. SHAP-based interpretation further refines these insights by revealing non-monotonic and threshold-dependent effects. The dependence patterns indicate that moderate levels of land expansion and intermediate productivity intensities are associated with negative freshwater outcomes, whereas higher land allocations and extreme productivity levels tend to exert more favorable effects. The empirical evidence supported the robustness of the model in capturing the relationship between agricultural indicators and freshwater withdrawal, with acceptable predictive accuracy under LOOCV validation. However, these findings were still interpreted with caution given the limited sample size, while recognizing the model's potential for providing reliable initial inference in a data-constrained context. These results reflect the presence of tipping points and asymmetric responses within the water–agriculture nexus, emphasizing that the impacts of agricultural activity vary across observations rather than remaining constant. From a policy perspective, the evidence suggests that land-use and agricultural productivity policies should be designed with greater sensitivity to nonlinear effects. Agricultural expansion strategies should prioritize efficiency and water-conscious land management, particularly at stages where freshwater pressure intensifies. Likewise, productivity-enhancing measures should be aligned with investments in water-saving technologies and sustainable practices. Looking ahead, future research should enlarge

the scope of analysis by extending the time horizon, increasing the number of explanatory variables, and applying this modern modeling framework across regions. Such extensions would strengthen policy relevance and support more effective strategies for sustainable water and agricultural management. Nevertheless, it is acknowledged that the exclusion of climatic and institutional variables and the limited sample size impose constraints on causal interpretation and external generalization, suggesting that future studies should adopt multi-factor and higher-frequency datasets to further validate and extend these findings.

Author Contributions

Conceptualization, A.E. and R.I.H.; methodology, R.I.H.; software, R.I.H.; validation, J.M., R.M.N., and A.Z.; formal analysis, A.E.; investigation, R.I.H.; resources, J.M.; data curation, R.I.H.; writing—original draft preparation, R.I.H.; writing—review and editing, A.E., J.M., R.M.N., and A.Z.; visualization, R.I.H.; supervision, A.E. and R.M.N.; project administration, R.I.H.; funding acquisition, J.M. All authors have read and agreed to the published version of the manuscript.

Data Availability

The datasets analyzed during the current study are available from publicly accessible databases and from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- Alcott, B. (2005). Jevons' paradox. *Ecol. Econ.*, *54*(1), 9–21. <https://doi.org/10.1016/j.ecolecon.2005.03.020>.
- Azertag. (2023). *Approximately 70 percent of freshwater resources are used in agriculture (Azerbaijani)*. https://azertag.az/xeber/sirin_su_ehtiyatlarinin_texminen_70_faizi_kend_teserrufatinda_istifade_olunur-2789074
- Banda, K., Shilengwe, C., & Nyambe, I. (2025). Integrating environmental and LULC drivers of groundwater droughts in groundwater-dependent ecosystems: A machine learning (XGBoost)-SEM analysis with ecosystem implications. *Ecol. Process.*, *14*, 64. <https://doi.org/10.1186/s13717-025-00633-w>.
- CEIC. (2021). *Azerbaijan Agricultural Irrigated Land: % of Total Agricultural Land*. <https://www.ceicdata.com/en/azerbaijan/environmental-land-use-protected-areas-and-national-wealth/az-agricultural-irrigated-land--of-total-agricultural-land>
- Chen, T. & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>.
- Djanibekov, N., Van Assche, K., Bobojonov, I., & Lamers, J. P. A. (2012). Farm restructuring and land consolidation in Uzbekistan: New farms with old barriers. *Eur.-Asia Stud.*, *64*(6), 1101–1126. <https://doi.org/10.1080/09668136.2012.691720>.
- Du, P., Huai, H., Wu, X., Wang, H., Liu, W., & Tang, X. (2025). Using XGBoost–SHAP for understanding the ecosystem services trade-off effects and driving mechanisms in ecologically fragile areas. *Front. Plant Sci.*, *16*, 1552818. <https://doi.org/10.3389/fpls.2025.1552818>.
- Eslamifar, G., Balali, H., & Fernald, A. (2024). Fallowing strategy and its impact on surface water and groundwater withdrawal, and agricultural economics: A system dynamics approach in Southern New Mexico. *Water*, *16*(1), 181. <https://doi.org/10.3390/w16010181>.
- Grafton, R. Q., Williams, J., Perry, C. J., Molle, F., Ringler, C., Steduto, P., Udall, B., Wheeler, S. A., Wang, Y., Garrick, D., et al. (2018). The paradox of irrigation efficiency. *Science*, *361*(6404), 748–750. <https://doi.org/10.1126/science.aat9314>
- Huang, K., Wang, D., & Kapelan, Z. (2025). A novel machine learning framework for advanced driving force analysis of individuals' dietary water footprint. *Earth's Future*, *13*(12), e2024EF005061. <https://doi.org/10.1029/2024EF005061>.
- Khondoker, M., Mandal, S., Gurav, R., & Hwang, S. (2023). Freshwater shortage, salinity increase, and global food production: A need for sustainable irrigation water desalination—A scoping review. *Earth*, *4*(2), 223–240. <https://doi.org/10.3390/earth4020012>.
- Li, M., Fu, Q., Singh, V. P., Liu, D., Li, T., & Zhou, Y. (2020). Managing agricultural water and land resources with tradeoff between economic, environmental, and social considerations: A multi-objective non-linear optimization model under uncertainty. *Agric. Syst.*, *178*, 102685. <https://doi.org/10.1016/j.agsy.2019.102685>.

- Li, W., Deng, M., Liu, C., & Cao, Q. (2025). Analysis of key influencing factors of water quality in Tai Lake Basin based on XGBoost-SHAP. *Water*, 17(11), 1619. <https://doi.org/10.3390/w17111619>.
- Lundberg, S. M. & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA*.
- METEO. (2022). *Freshwater withdrawals*. National Hydrometeorological Service (Azerbaijani). <https://meteo.az/index.php?ln=az&pg=122>
- Narayan, P. K. (2005). The saving and investment nexus for China: Evidence from cointegration tests. *Appl. Econ.*, 37(17), 1979–1990. <https://doi.org/10.1080/00036840500278103>.
- Niftiyev, I. (2020). The role of agriculture and agrarian sectors in the Azerbaijan economy: Main trends and dynamics since 1991. *SSRN*. <https://doi.org/10.2139/ssrn.3681896>.
- Poursaeid, M. (2025). Long-term study of water resource by machine learning: A new application of explainable artificial intelligence model towards sustainable development goals. *Water Resour. Manage.*, 39(15), 7995–8016. <https://doi.org/10.1007/s11269-025-04326-z>.
- Sarwar, S. (2025). Consequences of land utilization, agriculture and water to handle the food security issues. *Land Degrad. Dev.*, 36(6), 1962–1976. <https://doi.org/10.1002/ldr.5475>.
- Shapley, L. (1953) A value for n-Person games. In *Contributions to the Theory of Games II* (pp. 307–317). Princeton University Press.
- SSCRA. (2025). *Agriculture, forestry and fishery*. The State Statistical Committee of the Republic of Azerbaijan. <https://www.stat.gov.az/source/agriculture/?lang=en>
- United Nations. (2015). *Transforming Our World: The 2030 Agenda for Sustainable Development (A/RES/70/1)*. General Assembly. <https://docs.un.org/en/A/res/70/1>
- Valiyev, A., Dadashov, A., Jafarova, G., Babayeva, V., & Javadzade, K. (2025). Agricultural lands in Azerbaijan: Current usage situation, problems and solutions. *Agric. Rural Stud.*, 3(4), 17. <https://doi.org/10.59978/ar03040021>.
- Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *PLoS Med.*, 4(10), e296. <https://doi.org/10.1371/journal.pmed.0040296>.
- Ward, F. A. & Pulido-Velazquez, M. (2008). Water conservation in irrigation can increase water use. In *Proceedings of the National Academy of Sciences (Vol. 105, No. 47)* (pp. 18215–18220). <https://doi.org/10.1073/pnas.0805554105>.
- World Bank Group. (2025). *World Development Indicators: Agriculture and freshwater withdrawals*. <https://databank.worldbank.org/source/world-development-indicators>
- Wu, N., Zhou, Y., Yin, S., Gong, H., & Zhang, C. (2025). Revealing the nonlinear impact of environmental regulation on ecological resilience using the XGBoost-SHAP model: Evidence from the Yangtze River Delta region, China. *J. Clean. Prod.*, 541, 145700. <https://doi.org/10.1016/j.jclepro.2025.145700>.
- Zhao, Q., Gao, F., He, B., Li, Y., Li, H., Xiao, Y., & Lin, R. (2025). Analysis of temporal and spatial variations in cropland water-use efficiency and influencing factors in Xinjiang based on the XGBoost-SHAP model. *Agronomy*, 15(8), 1902. <https://doi.org/10.3390/agronomy15081902>.
- Zhou, B., Chen, G., Yu, H., Zhao, J., & Yin, Y. (2024). Revealing the nonlinear impact of human activities and climate change on ecosystem services in the Karst Region of Southeastern Yunnan using the XGBoost-SHAP model. *Forests*, 15(8), 1420. <https://doi.org/10.3390/f15081420>.