

Data-Driven Modeling of Reservoir Evaporation Dynamics at Amir Kabir Dam

Maziar Asmani¹, Seyed Ahmad Kebriyae², Seyedeh Mohadeseh Miri³, Kazem Bashirnezhad⁴

¹ Department of Computer Engineering, Qeshm Branch, Islamic Azad University, Qeshm, Iran. maziarasmani@iau.ac.ir

² Department of Mechanical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.
seyedahmad.kebriyae@alumni.um.ac.ir

³ Department of Mechanical Engineering, University of Zabol, Zabol, Iran. mhdsh.miri@uoz.ac.ir

⁴ Department of Mechanical Engineering, Ma.C., Islamic Azad University, Mashhad, Iran. bashirnezhad@iau.ac.ir

Received: 03 September 2025

Accepted: 5 October 2025

Available online: 2 December 2025

Abstract: Evaporation from dam reservoirs is a major source of water loss in arid and semi-arid regions, which makes accurate prediction essential for sustainable water resource management. In this study, evaporation from the Amir Kabir Dam reservoir was predicted for 2020–2024 using statistical and machine learning approaches. Monthly evaporation was selected as the target variable, while month, year, air temperature, solar radiation, relative humidity, and wind speed were used as input parameters. Three models, Multivariate Linear Regression (MLR), Random Forest (RF), and XGBoost, were evaluated using mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination (R^2). Correlation analysis and feature importance assessment were also performed to identify the dominant factors affecting evaporation. The results showed that the machine learning models outperformed the linear regression model in predicting evaporation dynamics. Among the tested models, RF achieved the highest accuracy (MAE = 115.35, RMSE = 156.02, R^2 = 0.8755), followed by XGBoost (R^2 = 0.8651), while MLR showed the lowest performance (R^2 = 0.5157). The findings confirmed the nonlinear and multifactorial nature of reservoir evaporation. Correlation analysis indicated positive relationships between evaporation and solar radiation, wind speed, and air temperature, whereas relative humidity showed a negative correlation. Feature importance analysis identified solar radiation as the most influential variable in both the RF and XGBoost models, while the month variable highlighted the role of seasonal variation in evaporation behavior. Overall, the results demonstrate that machine learning approaches, and particularly RF, are effective tools for reservoir evaporation prediction and can support water resource management and the monitoring of high-risk evaporative conditions.

Keywords: Reservoir evaporation; Machine learning; Random Forest; XGBoost; Water resources management.

1. Introduction

Water is a critical resource for agricultural, industrial, and domestic activities, particularly in arid and semi-arid regions. However, excessive consumption and mismanagement of water resources, together with climate change, population growth, and uneven precipitation distribution, have significantly reduced water availability worldwide [1–5]. Climate change has intensified this challenge by increasing evaporation rates and reducing accessible water resources, especially in dry-climate countries [6–8]. In some regions, such as Australia, annual evaporation can reach 2 m per year while annual precipitation may be as low as 500 mm per year, so that nearly 95 % of the precipitation is lost through evaporation. Furthermore, more than 50 % of the water stored in dams and small reservoirs may evaporate [2, 10].

Iran lies within the world's arid belt and receives an average annual precipitation of only 222 mm, which is less than one third of the global average [11]. Although Iran occupies approximately 1.1 % of the Earth's surface, it holds only 0.3 % of the world's renewable freshwater resources [12]. Evaporation rates across the country range from 2400 to 4500 mm per year and may exceed 5000 mm per year in certain regions [13]. Approximately 126 million m³ of water are lost annually

from the Karun-3 Dam alone, which corresponds to 4.2 % of the reservoir volume [12]. The average annual evaporation rate in Iran is approximately 2100 mm, nearly three times the global average [14], and it is estimated that around 3.5 billion m³ of water evaporates annually from Iranian dam reservoirs [12]. Effective evaporation reduction strategies are therefore essential [15].

Globally, evaporation losses account for nearly 25 % of freshwater consumption in the agricultural, industrial, and domestic sectors [16]. In highly arid regions, reservoir evaporation losses can reach up to 50 % of storage capacity [17, 18]. According to the Forest, Rangeland and Watershed Management Organization of Iran, the country ranks 11th among 112 countries facing severe water stress, and available water resources would need to increase by approximately 112 % to maintain current conditions by 2028 [19]. Substantial attention has consequently been directed toward the development of evaporation mitigation techniques.

Various physical and chemical approaches have been proposed to reduce evaporation losses. Physical methods include floating balls, metal sheets, polypropylene covers, and natural materials that reduce solar radiation exposure, limit airflow, and suppress evaporation. Chemical techniques generally involve monolayer

compounds that form a thin protective film on the water surface. Previous studies have investigated the effectiveness of physical covers, including polypropylene sheets and floating balls, in reducing evaporation under different meteorological conditions [20–22]. Floating covers may be classified as continuous, suspended, or modular systems, and modular covers contain open spaces that permit gas and light exchange [23, 24]. Wooden and steel covers have been reported to reduce evaporation by up to 90 % without adversely affecting water quality [25, 26]. In the Karun River basin alone, annual evaporation has been estimated at approximately 1521 mm, which corresponds to billions of cubic metres of water lost each year [27].

Water availability is influenced by climatic factors, such as temperature and precipitation, and by non-climatic factors, including population growth, technological development, and water resource management practices [28]. Reservoirs play a vital role in storing water during wet seasons for use during dry periods, and they are also essential for hydropower generation, flood control, transportation, and water supply [29]. Evaporation nevertheless remains one of the largest sources of water loss from storage reservoirs. In some countries, such as Australia and Spain, evaporation losses may account for up to 50 % and 10 % of stored water resources, respectively [9, 30, 31].

Research on evaporation reduction technologies has expanded since the introduction of chemical surface films in the 1950s [32]. Numerous mitigation methods have been developed since then, including monomolecular films, suspended shading systems, and floating modular covers [25, 33, 34]. Many existing studies remain largely empirical, however, and do not provide sufficient precision for engineering or economic optimization [9]. Previous investigations have shown that evaporation reduction efficiency is generally proportional to the coverage area [35], but comprehensive evaluations comparing the performance of different cover systems under varying environmental conditions remain limited. Biological approaches, such as the use of aquatic plants including duckweed and mosses, have also been explored but have received limited attention because of their relatively low efficiency [35]. Other natural and artificial materials, including cordia leaves and granular polystyrene, have demonstrated acceptable evaporation reduction performance. Given the increasing pressures of population growth, climate change, and water scarcity, the development and implementation of effective strategies to minimize evaporation losses from reservoirs has become increasingly important.

Despite the growing number of studies on reservoir evaporation and evaporation mitigation strategies, several important research gaps remain. Most previous investigations have focused on empirical analyses,

physical evaporation reduction techniques, or conventional statistical approaches, while comprehensive data-driven modeling studies capable of capturing the nonlinear behavior of reservoir evaporation under varying hydroclimatic conditions remain limited. Comparative evaluations of traditional statistical models and advanced machine learning algorithms for reservoir evaporation prediction in arid and semi-arid regions, and particularly in Iran, are still insufficient. The combined analysis of temporal reservoir variations, hydroclimatic interactions, correlation structures, and feature importance within machine learning frameworks has also received relatively limited attention. The present study therefore aims to address these gaps by developing and comparing Multivariate Linear Regression (MLR), Random Forest (RF), and XGBoost models for predicting evaporation from the Amir Kabir Dam reservoir using hydroclimatic and temporal variables. In addition to predictive modeling, the study integrates correlation analysis and feature importance evaluation to improve the physical interpretability of the machine learning predictions and to identify the dominant factors controlling reservoir evaporation dynamics. The findings are expected to contribute to more reliable evaporation forecasting and to support sustainable reservoir and water resource management in arid and semi-arid environments.

2. Methodology and Data Preparation

The city of Karaj in Alborz Province, located near the Amir Kabir Dam, was selected as the study area. Because of its critical role in regional water supply and its sensitivity to climatic and hydrological variation, the reservoir provides a suitable case for investigating and modeling evaporation from dam surfaces. The location of the Amir Kabir Dam is presented in Figure 1. Meteorological and climatic data for the period 2020–2024 were obtained from widely validated reanalysis and satellite-based datasets, including the ERA5 data provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) and the NASA POWER datasets. Data processing and spatial extraction were carried out on the Google Earth Engine platform, and monthly averages of the main climatic variables were calculated within an approximately 3 km radius around the dam coordinates.

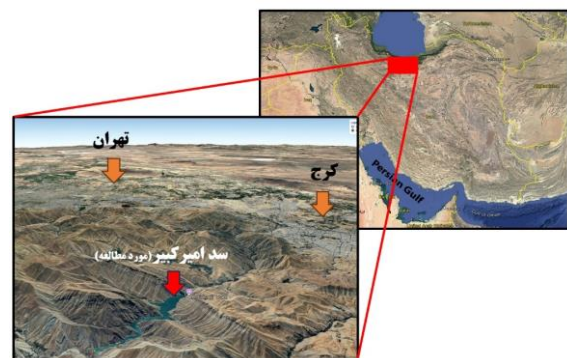


Figure 1. Location of the Amir Kabir Dam (case study area).

The selected input parameters were air temperature, surface solar radiation, relative humidity, wind speed, and the temporal variables month and year. Monthly evaporation was taken as the target variable. These parameters were selected because reservoir evaporation is a multifactorial process influenced by thermal, radiative, atmospheric, and humidity-related conditions. Increases in air temperature and solar radiation provide the energy required for evaporation and generally lead to higher evaporation rates. Wind speed enhances the transport of water vapour from the water surface to the atmosphere and thereby accelerates evaporation. Higher relative humidity, in contrast, reduces the vapour absorption capacity of the air and consequently decreases evaporation rates. After extraction, the data were organized into a monthly integrated database, and quality control procedures were performed to evaluate data completeness, temporal consistency, and potential anomalies. The prepared dataset was then divided into training and testing subsets to assess the predictive capability and generalization performance of the developed models. Three predictive approaches, Multivariate Linear Regression (MLR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), were implemented and compared. Statistical error indices and goodness-of-fit criteria were used to compare their performance in predicting reservoir evaporation.

3. Models Employed and Evaluation Criteria

Three predictive models were implemented to predict evaporation from the Amir Kabir Dam reservoir: Multivariate Linear Regression (MLR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). These models were selected to compare a conventional linear statistical approach with two tree-based machine learning techniques in representing the relationships between meteorological variables and reservoir evaporation. The independent variables were month, year, air temperature (TEMP), solar radiation (SOLAR), relative humidity (RH), and wind speed (WS), while evaporation (EVAP) was the dependent variable. The selection of these variables was based on the physical mechanisms governing evaporation and on the results of the correlation analysis between the meteorological parameters, since reservoir evaporation is a multifactorial phenomenon influenced by thermal, radiative, humidity-related, and atmospheric dynamic conditions.

Multivariate Linear Regression was used as the baseline model. This approach establishes a linear relationship between the dependent variable and multiple independent variables. Because of its simplicity, interpretability, and widespread application in hydrological studies, MLR is commonly used as a benchmark for evaluating more advanced predictive methods. Since evaporation is governed by complex and often nonlinear interactions among climatic variables, however, the predictive capability of linear regression

may be limited under such conditions. The Random Forest model is an ensemble machine learning approach based on multiple decision trees. By aggregating the outputs of several trees, the model improves prediction accuracy and reduces sensitivity to noise in the dataset. RF has demonstrated strong capability in capturing nonlinear relationships and complex interactions among variables and has been applied successfully in numerous hydrological forecasting studies.

XGBoost is a boosting-based machine learning algorithm that sequentially improves prediction performance by minimizing model errors through iterative tree construction. Owing to its high predictive capability, computational efficiency, and effective overfitting control, XGBoost has become one of the most widely used algorithms for regression and forecasting applications. The dataset was ordered chronologically and divided into training and testing subsets. Approximately 80 % of the time-series data were used for model training and the remaining 20 % were reserved for testing. This partitioning strategy was adopted to preserve the temporal structure of the dataset, to simulate realistic forecasting conditions, and to prevent information leakage between past and future observations.

Four statistical indicators were employed to evaluate model performance: mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and the coefficient of determination (R^2). MAE measures the average magnitude of prediction errors regardless of direction and provides a direct indication of model accuracy. MSE gives greater weight to large prediction errors through the squaring process. RMSE, the square root of MSE, expresses prediction error in the same unit as the target variable and is therefore easier to interpret. The R^2 metric indicates the proportion of variance explained by the model, with values closer to one representing stronger predictive performance. Models with lower MAE, MSE, and RMSE values and higher R^2 values were considered to have superior predictive performance. The three models were compared on these criteria to identify the most suitable approach for predicting evaporation from the reservoir.

4. Results and Discussion

The water storage volume of the Amir Kabir Dam reservoir during 2020–2024 exhibited considerable monthly and interannual variation. As illustrated in Figure 2, reservoir volume did not follow a uniform trend throughout the study period but instead showed alternating phases of increase and decline influenced by seasonal fluctuations, inflow conditions, water withdrawals, and climatic factors. These variations indicate that the hydrological behavior of the reservoir is highly sensitive to changes in inflow, outflow, and evaporative water loss, particularly during the warmer months of the year. Increases in reservoir volume during certain periods were likely associated with precipitation

events, upstream snowmelt, and enhanced inflow conditions. Reductions in reservoir storage, in contrast, may be attributed to decreased inflows, increased downstream water demand, controlled water releases, and intensified evaporation during warm seasons. Fluctuations in reservoir volume therefore cannot be linked to a single controlling factor but result from the combined influence of climatic, hydrological, and operational conditions.

Analysis of reservoir storage trends matters because variations in stored water volume directly affect the water–atmosphere interaction area, the reservoir surface extent, and consequently the evaporation rate. Although increases in storage volume do not necessarily produce a linear increase in evaporation, they generally lead to larger exposed surface areas and altered surface conditions that can enhance evaporative losses. Evaluating reservoir volume variations together with meteorological parameters such as temperature, solar radiation, relative humidity, and wind speed therefore provides a more comprehensive understanding of the mechanisms controlling reservoir evaporation. The results presented in Figure 2 also emphasize the importance of seasonally adaptive reservoir management strategies. Declining reservoir levels during periods of elevated temperature and intensified evaporation may increase water stress and reduce the reliability of stored water resources. Continuous monitoring of reservoir storage fluctuations can therefore support operational decision-making, water allocation planning, and vulnerability assessment under drought and seasonal warming conditions.

The distinct temporal variations shown in Figure 2 highlight the importance of reservoir dynamics in understanding evaporation behavior and in evaluating the sustainability of water resources. They provide the basis for examining the relationships among climatic variables, reservoir conditions, and the predictive capability of the developed models in the following sections. The observed variability also reflects the seasonal behavior of hydroclimatic parameters and the evaporation patterns associated with annual climatic cycles, which plays a key role in interpreting model performance.

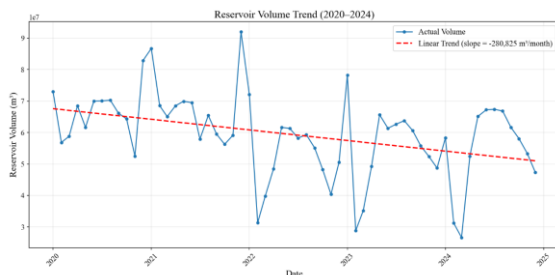


Figure 2. Temporal variations in the storage volume of the Amir Kabir Dam reservoir during 2020–2024.

The predictive performance of the developed models was evaluated using mean absolute error (MAE), mean

squared error (MSE), root mean squared error (RMSE), and the coefficient of determination (R^2). The results of the three approaches are summarized in Table 1.

Table 1. Performance comparison of Linear Regression, Random Forest, and XGBoost models for reservoir evaporation prediction.

Model	MAE	MSE	RMSE	R^2
Linear Regression	256.2536	94695.52	307.7264	0.515737
Random Forest	115.3502	24340.84	156.0155	0.875524
XGBoost	133.5934	26380.34	162.4202	0.865094

The Random Forest model achieved the highest predictive performance among the evaluated approaches. It produced an MAE of 115.35, an MSE of 24,340.84, an RMSE of 156.02, and an R^2 value of 0.8755. This R^2 value indicates that approximately 87.5 % of the variability in reservoir evaporation was explained by the selected input variables, which demonstrates the strong capability of the model in representing evaporation dynamics.

The XGBoost model also showed strong predictive capability, with an R^2 value of 0.8651 and an RMSE of 162.42, and therefore performance comparable to that of the Random Forest model. The small difference between the two machine learning approaches suggests that both algorithms captured the nonlinear interactions among meteorological variables, reservoir conditions, and evaporation processes effectively. The lower MAE, MSE, and RMSE values obtained by the Random Forest model nevertheless indicate slightly greater prediction stability and overall accuracy for the dataset used in this study.

The Linear Regression model, in contrast, demonstrated the weakest predictive performance, with an MAE of 256.25, an RMSE of 307.73, and an R^2 value of 0.5157. The low coefficient of determination and the higher prediction errors suggest that the relationships governing reservoir evaporation are highly nonlinear and involve complex interactions that cannot be represented adequately by a purely linear approach. Although the linear regression model captured part of the variability in evaporation, its predictive capability remained substantially lower than that of the machine learning models.

Table 1 therefore shows that the machine learning models, and particularly Random Forest and XGBoost, improved evaporation prediction accuracy substantially compared with the linear regression model. The Random Forest model achieved the best overall predictive performance and was identified as the most suitable model for reservoir evaporation prediction in this study.

Figure 3 presents the relationship between the observed evaporation values and those predicted by the Multivariate Linear Regression (MLR) model. Comparing predicted and observed values provides an effective means of evaluating the capability of the model to reproduce evaporation behavior under varying hydroclimatic conditions. In the scatter plot, the 1:1

reference line ($y = x$) represents ideal agreement between predicted and observed evaporation values. Points located close to this line indicate accurate model predictions, whereas deviations from the line reflect prediction errors. The distribution of points indicates that the MLR model was generally capable of representing evaporation behavior within moderate evaporation ranges, where many observations were clustered near the reference line.

Larger deviations from the ideal line were observed at higher evaporation values, however, which indicates a reduction in predictive accuracy under conditions of intensified evaporation. This behavior suggests that the linear regression model was unable to capture the nonlinear interactions governing reservoir evaporation fully. Evaporation is influenced by complex relationships among meteorological variables, including air temperature, solar radiation, relative humidity, and wind speed, whose combined effects are often nonlinear and dynamically interdependent. Noticeable dispersion of several points was also observed at lower evaporation levels, which suggests that abrupt meteorological fluctuations and transitional seasonal conditions may introduce additional prediction uncertainty. Such variability is likely associated with short-term atmospheric instability and rapid changes in climatic conditions.

Although the MLR model reproduced the general trend of evaporation variability, considerable prediction scatter therefore remained across certain evaporation ranges. These findings highlight the limitations of linear approaches in modeling complex hydroclimatic processes and support the application of more advanced machine learning techniques for improving prediction accuracy. The following sections accordingly investigate the relationships between climatic variables and evaporation through correlation analysis and feature importance evaluation, in order to obtain a more comprehensive understanding of the dominant factors controlling reservoir evaporation.

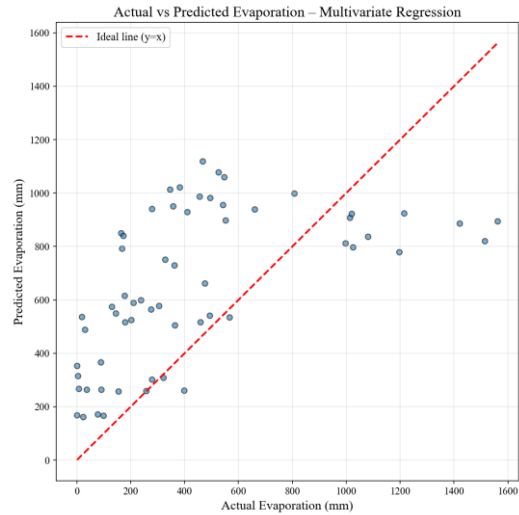


Figure 3. Comparison between observed and predicted evaporation values obtained using the Multivariate Linear Regression model.

Figure 4 presents the time-series comparison between observed evaporation values and those predicted by the Random Forest (RF) and XGBoost models for the year 2024. Both machine learning models reproduced the general seasonal pattern of reservoir evaporation, including the progressive increase during the warm season and the subsequent decline toward the end of the year. During the first quarter of the year (January to March), observed evaporation values remained relatively low. Both models captured this low-evaporation period reasonably well, although slight overestimations were observed in March, which indicates that the models responded somewhat earlier to the onset of the increasing evaporation trend.

A pronounced increase in evaporation was observed between April and June, with peak evaporation occurring in May. Both RF and XGBoost represented the seasonal intensification of evaporation successfully, but both underestimated the peak value in May. XGBoost produced predictions closer to the observed peak, whereas RF exhibited slightly larger deviations. The underestimation of peak evaporation may reflect the inherent difficulty of machine learning models in reproducing extreme hydroclimatic conditions and abrupt nonlinear responses in evaporation behavior. Observed evaporation declined after June, and both models generally reproduced the decreasing trend from July to September. Model performance varied across individual months during this period: RF slightly overestimated evaporation in August, for example, while XGBoost predictions remained closer to the observed values. Both approaches nevertheless maintained acceptable agreement with the observed seasonal trend.

During the final months of the year (October to December), evaporation exhibited moderate short-term variability. Both models captured the increase in evaporation observed in October relative to September, although the magnitude of this increase was

underestimated. In November, observed evaporation values exceeded the predictions of both models, with XGBoost producing the lowest estimates. XGBoost predictions in December, in contrast, showed closer agreement with observations, whereas RF slightly underestimated evaporation.

Both RF and XGBoost therefore possessed strong capability in reproducing the annual seasonal variability of reservoir evaporation. Although both approaches captured the overall increasing and decreasing trends throughout the year effectively, prediction accuracy decreased during periods of peak evaporation. XGBoost showed comparatively better agreement with observations during certain months, including May, August, and December, whereas RF provided more accurate estimates during other periods, particularly in June and July. The predictive performances of the two machine learning models were thus generally comparable, with each approach exhibiting relative advantages under different hydroclimatic conditions. These results further suggest that machine learning approaches can provide valuable support for reservoir evaporation forecasting and water resource management applications. Improving prediction accuracy during extreme evaporation periods, however, may require additional model optimization and the incorporation of further influential climatic or hydrological variables.

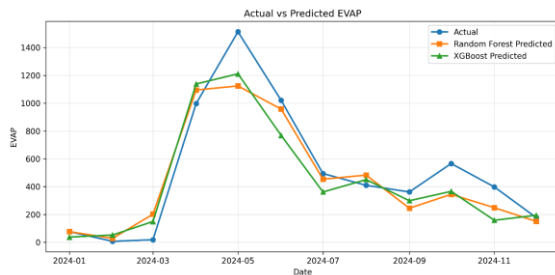


Figure 4. Comparison between observed and predicted evaporation time series generated by the Random Forest and XGBoost models for 2024.

Correlation analysis between the input variables and reservoir evaporation provides valuable insight into the hydroclimatic mechanisms governing evaporation dynamics and helps to explain the predictive behavior of the developed models. Figure 5 presents the correlation matrix between the climatic variables and the physical characteristics of the reservoir. The matrix illustrates both the strength and the direction of the linear relationships among variables and serves as an important basis for interpreting model performance. As shown in Figure 5, evaporation (EVAP) exhibited positive correlations with several meteorological parameters. Solar radiation showed the strongest positive correlation with evaporation ($r = 0.58$), which indicates that increased radiative energy at the reservoir surface enhances evaporation intensity. Wind speed and air temperature also demonstrated moderate positive correlations with evaporation ($r = 0.52$ and $r = 0.50$,

respectively). These relationships are physically consistent with evaporation processes, since increased temperature and solar radiation supply the energy required for water vaporization, while stronger wind conditions enhance vapour transport from the water surface to the atmosphere.

Evaporation displayed a negative correlation with relative humidity ($r = -0.24$). Although weaker than the relationships observed for temperature, radiation, and wind speed, this negative correlation remains consistent with the physical principles governing evaporation, since higher atmospheric humidity reduces the vapour pressure gradient between the water surface and the surrounding air and thereby limits evaporation potential. The correlation structure further indicates that energy-transfer and atmospheric transport mechanisms exert stronger control over reservoir evaporation than humidity conditions alone. Significant interrelationships were also observed among the climatic variables themselves. Air temperature showed a strong positive correlation with solar radiation ($r = 0.76$), while solar radiation and wind speed were also strongly correlated ($r = 0.80$). Relative humidity exhibited negative correlations with temperature ($r = -0.76$), solar radiation ($r = -0.70$), and wind speed ($r = -0.68$). These patterns demonstrate the presence of substantial covariance and interdependence among the predictor variables.

Such dependencies suggest that the influence of individual climatic variables on evaporation cannot be considered independently, but rather through complex interactions among atmospheric processes. This covariance structure may partly explain the limited performance of the simple linear regression model, since linear approaches may not capture nonlinear interactions and coupled hydroclimatic responses adequately. As regards the morphological characteristics of the reservoir, surface area and storage volume exhibited a nearly perfect positive correlation ($r = 1.00$), which indicates that both variables changed simultaneously within the analyzed dataset. Both parameters showed relatively weak correlations with evaporation ($r = 0.15$), however. This finding suggests that although reservoir hydrological conditions influence evaporation behavior, short-term evaporation variability in the studied system is controlled more strongly by atmospheric and climatic conditions than by direct geometric changes in reservoir area or volume.

Figure 5 therefore confirms that reservoir evaporation is a multifactorial process governed primarily by solar radiation, air temperature, and wind speed, while relative humidity exerts a moderating effect. The strong interdependence among predictor variables highlights the complexity of the evaporation system and emphasizes the need for predictive approaches capable of capturing nonlinear relationships and variable interactions. The correlation analysis thus provides a strong foundation for the subsequent

investigation of variable importance within the machine learning models.

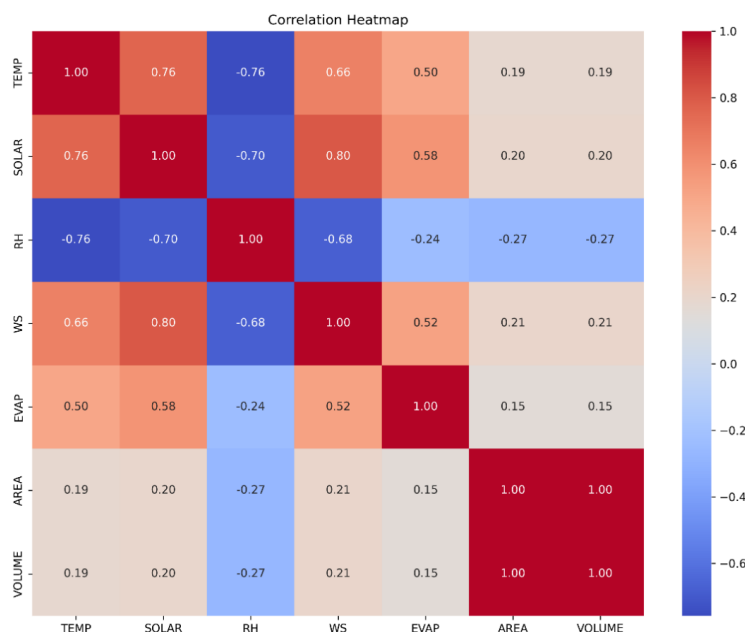


Figure 5. Pearson correlation matrix of hydroclimatic variables and reservoir characteristics.

Figure 6 shows the relative importance of the input variables used for reservoir evaporation prediction in the Random Forest (RF) and XGBoost models. The predictors were month, year, air temperature (TEMP), solar radiation (SOLAR), relative humidity (RH), and wind speed (WS). Feature importance analysis was conducted to identify the dominant variables influencing model predictions and to improve the interpretability of the machine learning approaches. Solar radiation was identified as the most influential variable in both the RF and XGBoost models. This result is physically consistent with the evaporation process, since solar radiation represents the primary energy source driving water vaporization from the reservoir surface, and higher radiative energy increases the available surface energy and consequently intensifies evaporation rates. The strong importance assigned to SOLAR by both machine learning models indicates that the algorithms captured the dominant role of radiative forcing in controlling evaporation dynamics.

In the Random Forest model, solar radiation, air temperature (TEMP), and relative humidity (RH) were identified as the most significant predictors. The high importance of temperature reflects its direct influence on evaporation intensity through increased thermal energy availability. Relative humidity also contributed substantially to model performance, because atmospheric moisture conditions regulate the vapour pressure gradient between the reservoir surface and the surrounding air. The RF model was therefore particularly sensitive to both the energy-related and the atmospheric humidity controls governing evaporation behavior. For the XGBoost model, solar radiation again exhibited the highest importance, followed by the month

variable. The elevated importance of month indicates that XGBoost captured the seasonal and temporal variability of evaporation processes effectively. This behavior is consistent with the seasonal nature of reservoir evaporation, which generally increases during warmer periods characterized by higher radiative input and decreases during cooler seasons.

Comparison of the two models reveals that although the ranking of the secondary variables differed, both approaches emphasized the dominant influence of solar radiation consistently. Agreement between two independent machine learning algorithms strengthens the reliability of the result and confirms that solar radiation is a stable and critical predictor for reservoir evaporation modeling.

Variables such as year and wind speed (WS) exhibited comparatively lower importance values. Lower importance does not imply that these variables are irrelevant, however; it indicates that their contribution to reducing prediction error was less significant within the context of the studied dataset and model structures.

Figure 6 therefore indicates that energy-related and seasonal variables exert the strongest influence on reservoir evaporation variability. The high importance of solar radiation, temperature, and the temporal factors confirms that evaporation dynamics are controlled primarily by thermal and seasonal conditions. These findings are also consistent with the correlation analysis presented in Figure 5, where solar radiation, temperature, and wind speed exhibited the strongest positive relationships with evaporation. The feature importance analysis thus not only enhances the interpretability of the machine learning models but also

supports the physical understanding of the mechanisms governing reservoir evaporation.

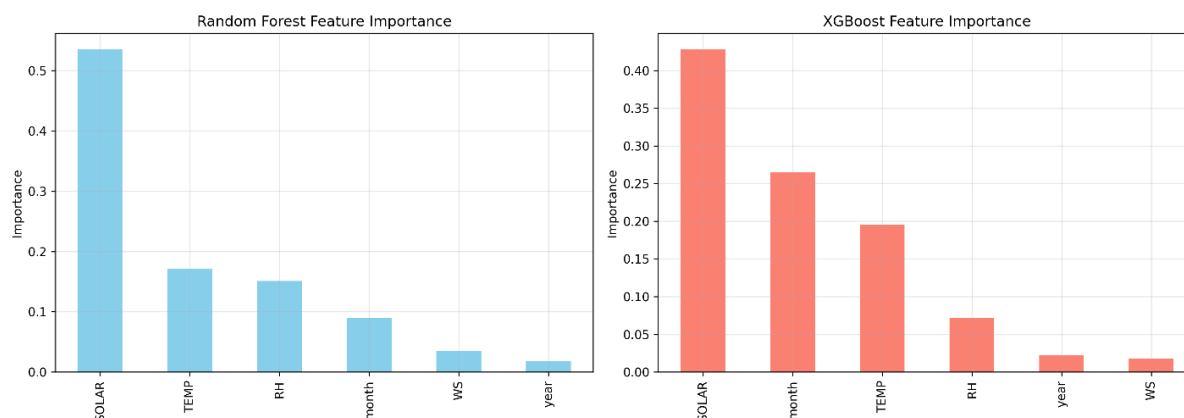


Figure 6. Feature importance of the input variables for reservoir evaporation prediction using the Random Forest and XGBoost models.

5. Conclusions

The present study investigated the prediction of evaporation from the Amir Kabir Dam reservoir during the period 2020–2024 using statistical and machine learning approaches. Multivariate Linear Regression (MLR), Random Forest (RF), and XGBoost models were developed using month, year, air temperature, solar radiation, relative humidity, and wind speed as predictor variables, while evaporation was the target variable. The results demonstrated that reservoir evaporation is a multifactorial, seasonal, and highly nonlinear process controlled primarily by hydroclimatic conditions. The main findings are summarized as follows.

- Temporal analysis of reservoir storage volume revealed considerable monthly and interannual fluctuation associated with climatic conditions, inflow variability, reservoir releases, and operational factors. Reservoir dynamics and evaporation behavior are therefore governed by the combined influence of climatic, hydrological, and management-related processes.

- Model comparison demonstrated that the Random Forest model achieved the highest predictive performance among the evaluated approaches (MAE = 115.35, RMSE = 156.02, $R^2 = 0.8755$), while XGBoost also showed strong predictive capability ($R^2 = 0.8651$). The Multivariate Linear Regression model, in contrast, exhibited substantially lower performance ($R^2 = 0.5157$), which highlights the limitations of linear approaches in representing complex and nonlinear evaporation processes. Time-series analysis further showed that both RF and XGBoost reproduced the seasonal variability of evaporation successfully, although prediction accuracy decreased during peak evaporation periods and abrupt climatic transitions.

- Correlation analysis indicated that evaporation was positively associated with solar radiation, wind speed, and air temperature, whereas relative humidity showed a negative relationship. Significant interdependence among the climatic variables was also observed, which

suggests the presence of complex hydroclimatic interactions. Feature importance analysis identified solar radiation as the dominant predictor in both the RF and XGBoost models, which confirms the critical role of radiative energy in controlling reservoir evaporation. Temperature, relative humidity, and seasonal variables such as month also showed considerable influence on evaporation variability.

The findings overall confirm the strong capability of machine learning approaches, and particularly of the Random Forest model, for accurate reservoir evaporation prediction and hydroclimatic analysis. These models can support evaporative loss monitoring, drought-risk assessment, reservoir operation planning, and water resource management decision-making. Future studies should incorporate longer observational periods, additional hydrological and environmental variables, remote sensing datasets, and advanced machine learning-based decision-support systems to further improve prediction accuracy and sustainable reservoir management strategies.

References

- [1] Ghazvinian, H., et al., Experimental study of evaporation reduction using polystyrene coating, wood and wax and its estimation by intelligent algorithms. *Irrigation and Water Engineering*, 2020. 11(2): p. 147-165.
- [2] Zhang, H., et al., A remote sensing method for estimating regional reservoir area and evaporative loss. *Journal of Hydrology*, 2017. 555: p. 213-227.
- [3] Chowdhury, S., M. Al-Zahrani, and A. Abbas, Implications of climate change on crop water requirements in arid region: an example of Al-Jouf, Saudi Arabia. *Journal of King Saud University-Engineering Sciences*, 2016. 28(1): p. 21-31.
- [4] Hammadi, S.H., Integrated solar still with an underground heat exchanger for clean water production. *Journal of King Saud University-Engineering Sciences*, 2020. 32(5): p. 339-345.
- [5] Morsy, S.M., et al., Imposing water deficit on modern and wild wheat collections to identify drought-resilient

- genotypes. *Journal of Agronomy and Crop Science*, 2022. 208(4): p. 427-440.
- [6] Haghighi, E., K. Madani, and A.Y. Hoekstra, The water footprint of water conservation using shade balls in California. *Nature Sustainability*, 2018. 1(7): p. 358-360.
- [7] Chowdhury, S. and M. Al-Zahrani, Characterizing water resources and trends of sector wise water consumptions in Saudi Arabia. *Journal of King Saud University-Engineering Sciences*, 2015. 27(1): p. 68-82.
- [8] Abdallah, A.M., et al., Critical evaluation of functional aspects of evaporation barriers through environmental and economics lens for evaporation suppression - a review on milestones from improved technologies. *Science of the Total Environment*, 2022. 833: p. 155346.
- [9] Craig, I.P., Loss of storage water due to evaporation - a literature review. 2005.
- [10] Zhao, G. and H. Gao, Estimating reservoir evaporation losses for the United States: Fusing remote sensing and modeling approaches. *Remote Sensing of Environment*, 2019. 226: p. 109-124.
- [11] Shirazadi, A., et al., Comparison of erosion and sediment yield estimation in watersheds with emphasis on the effect of dam construction. *Journal of Water and Soil Conservation*, 2019. 26(3).
- [12] Sadeghi, A.R., Reducing evaporation from dam water reservoirs. 2018.
- [13] Nouri, A., Estimation of evaporation from Chitgar Lake and the effect of chemical coatings on its reduction. 2015.
- [14] Mozafari, A., B. Mansouri, and S.F. Chini, Effect of wind flow and solar radiation on functionality of water evaporation suppression monolayers. *Journal of Hydrology*, 2021. 598: p. 126259.
- [15] Madadi, E., et al., Comparison of the performance of two types of natural and artificial covers in reducing evaporation from water reservoirs. *Iranian Journal of Irrigation and Drainage*, 2021. 15(4): p. 831-841.
- [16] Hogeboom, R.J., L. Knook, and A.Y. Hoekstra, The blue water footprint of the world's artificial reservoirs for hydroelectricity, irrigation, residential and industrial water supply, flood protection, fishing and recreation. *Advances in Water Resources*, 2018. 113: p. 285-294.
- [17] Craig, I.P., Loss of storage water due to evaporation. A literature review, NCEA Publication, University of Southern Queensland, 2006.
- [18] Friedrich, K., et al., Reservoir evaporation in the Western United States: current science, challenges, and future needs. *Bulletin of the American Meteorological Society*, 2018. 99(1): p. 167-187.
- [19] Ourgani, K., et al., Estimation of losses due to conveyance, distribution and delivery of agricultural water. 2010.
- [20] Han, K.-W., et al., Water savings efficiency of counterweighted spheres covering a plain reservoir in an arid area. *Water Supply*, 2020. 20(3): p. 1088-1097.
- [21] Youssef, Y.W. and A. Khodzinskaya, A review of evaporation reduction methods from water surfaces. in *E3S Web of Conferences*. 2019. EDP Sciences.
- [22] Zhang, H., et al., Evaporation reduction by suspended and floating covers: overview, modelling and efficiency. *Urban Water Security Research Alliance Technical Report*, 2012. 28.
- [23] Ruskowitz, J.A., et al., Evaporation suppression and solar energy collection in a salt-gradient solar pond. *Solar Energy*, 2014. 99: p. 36-46.
- [24] Rezaazadeh, A., P. Akbarzadeh, and M. Aminzadeh, The effect of floating balls density on evaporation suppression of water reservoirs in the presence of wind flow. *Journal of Hydrology*, 2023. 617: p. 128944.
- [25] Gallego-Elvira, B., et al., Energy balance and evaporation loss of an irrigation reservoir equipped with a suspended cover in a semiarid climate (southeastern Spain). *Hydrological Processes*, 2011. 25(11): p. 1694-1703.
- [26] Maestre-Valero, J., et al., Effects of a suspended shade cloth cover on water quality of an agricultural reservoir for irrigation. *Agricultural Water Management*, 2013. 118: p. 70-78.
- [27] Farzin, S. and F. Alizadeh Sanami, Shadow ball, a strategy for reducing water evaporation from dam reservoirs. *Journal of Water and Sustainable Development*, 2020. 7(1): p. 105-110.
- [28] Arnell, N.W., Climate change and global water resources. *Global Environmental Change*, 1999. 9: p. S31-S49.
- [29] White, W., World water: resources, usage and the role of man-made reservoirs. 2010.
- [30] Alvarez, V.M., et al., Efficiency of shading materials in reducing evaporation from free water surfaces. *Agricultural Water Management*, 2006. 84(3): p. 229-239.
- [31] Rost, S., et al., Agricultural green and blue water consumption and its influence on the global water system. *Water Resources Research*, 2008. 44(9).
- [32] Mansfield, W., Effect of surface films on the evaporation of water. *Nature*, 1953. 172(4389): p. 1101-1102.
- [33] Cooley, K.R. and L.E. Myers, Evaporation reduction with reflective covers. *Journal of the Irrigation and Drainage Division*, 1973. 99(3): p. 353-363.
- [34] Fox, R., An oil-wax-surfactant system for retarding the evaporation of water. *Nature*, 1965. 205(4975): p. 1004-1005.
- [35] Cooley, K.R., Energy relationships in the design of floating covers for evaporation reduction. *Water Resources Research*, 1970. 6(3): p. 717-727.
- [36] Asmani, M., et al., Determining an equation for estimating the evaporation from water surfaces covered by floating spheres using dimensional analysis and artificial neural networks. 2024.