50 YEARS OF WATER BODY MONITORING: THE CASE OF QARAAOUN RESERVOIR, LEBANON

Ali Ahmad Faour[1], Nabil Amacha[2], Ali J. Ghandour[3]

[1] American University of Beirut, Beirut, Lebanon
[2] Lebanese University, Beirut, Lebanon
[3] National Center for Remote Sensing, CNRS-L, Beirut, Lebanon - aghandour@cnrs.edu.lb

Keywords: Water Index, Volume Estimation, Water Reservoir Time-series, Remote Sensing.

Abstract

The sustainable management of the Qaraaoun Reservoir, the largest surface water body in Lebanon, located in the Bekaa Plain, hinges on reliable monitoring of its storage volume despite frequent sensor malfunctions and limited maintenance capacity. This study introduces a sensor-free approach that integrates open-source satellite imagery, advanced water-extent segmentation, and machine learning to estimate the reservoir surface area and then the volume in near real-time. Sentinel-2 and Landsat 1-9 images are processed where surface water is delineated using a newly proposed water segmentation index. A machine learning model based on Support Vector Regression (SVR) is trained on a curated dataset that includes water surface, water level and water volume calculation using reservoir bathymetry survey. The model is then able to estimated waterbody volume relying solely on water surface, extracted from satellite imagery without the need of any ground measurements. Water segmentation using the proposed index aligns with ground truth over 95% of the shoreline. Hyperparameter tuning with GridSearchCV yields an optimized SVR performance with error under 1.5% of full reservoir capacity and coefficients of determination exceeding 0.98. These results demonstrate the method's robustness and cost-effectiveness, offering a practical solution for continuous, sensor-independent monitoring of reservoir storage. The proposed methodology can be replicated to other water bodies, and the resulting 50+ years of time-series data is crucial for researchers studying climate change and environmental patterns.

1. INTRODUCTION

The Qaraaoun Reservoir (QR) is located at an average altitude of approximately 850 meters in the central Bekaa Valley, between the Lebanon Mountains and the Anti-Lebanon range. The reservoir lies near the Qaraaoun village, and specifically between the following geographic coordinates: 33°35'37"N, 33°32'53"N and 35°40'56"E, 35°42'26"E and was formed in 1959 by the construction of the Qaraaoun Dam across the Litani River. With a storage capacity exceeding 220 million cubic meters, the reservoir receives water from snow and rainfall and a number of existing springs and serves as a vital water resource for the region and supplies water to about 1 million people. This large volume buffers seasonal rainfall variability and provides a stable water supply that supports extensive irrigation systems, enhancing agricultural productivity in the valley. During extreme droughts, the reservoir plays a critical role in meeting domestic and industrial water demands. Additionally, its storage capacity contributes to hydroelectric power generation, about 22% of Lebanon's electricity, making it an essential component of regional economic development. These factors underscore the importance of consistently monitoring and managing the reservoir's water volume to ensure long-term sustainability and resilience amid declining water availability and climate change [1, 2].

The Litani River Authority (LRA), a governmental agency, is tasked with the integrated management and sustainable development of the Litani River Basin. Its responsibilities include monitoring water quality and quantity, regulating water distribution, planning water resource projects, and ensuring equitable access for agricultural, industrial, and domestic use. Specifically, for the QR, the LRA employs an integrated network of hydrometric stations to continuously monitor the volume of water. However, challenges such as sensor malfunction and human error pose significant risks to accurate data collection, potentially compromising effective water management. In many developing countries, the maintenance of sensor-based technology often falls short of established standards and best practices due to financial constraints and a shortage of experienced personnel in government institutions. As a result, accurate recording of water volume is often compromised.

Figure 1: Interactive dashboard showing time-series of water surface and volume in near real-time in the Qaraaoun Reservoir.

To address these challenges, this article proposes an efficient sensor-free approach that leverages remote sensing technology and machine learning to generate accurate weekly estimates of water volume in the Qaraaoun Reservoir. The contribution of this paper is three-folds as follows:

- Machine learning model was developed to infer the reservoir's water volume in near real-time from Sentinel-2 and LandSat 1-9 imagery. The model takes the extracted water surface area as an input and estimated the corresponding water volume without relying on any ground-based sensor readings.

- Because model performance is highly sensitive to the accuracy of the water surface extraction, we introduce a novel water segmentation index that combines two existing indices from the literature using a weighted sum.
- Finally, we developed an interactive, web-based platform to visualize volume trends and segmentation results, as shown in Figure 1. The dashboard hosts a time series of over 50 years of reservoir statistics and serves as a valuable tool for researchers and stakeholders to explore environmental patterns and study the impacts of climate change.

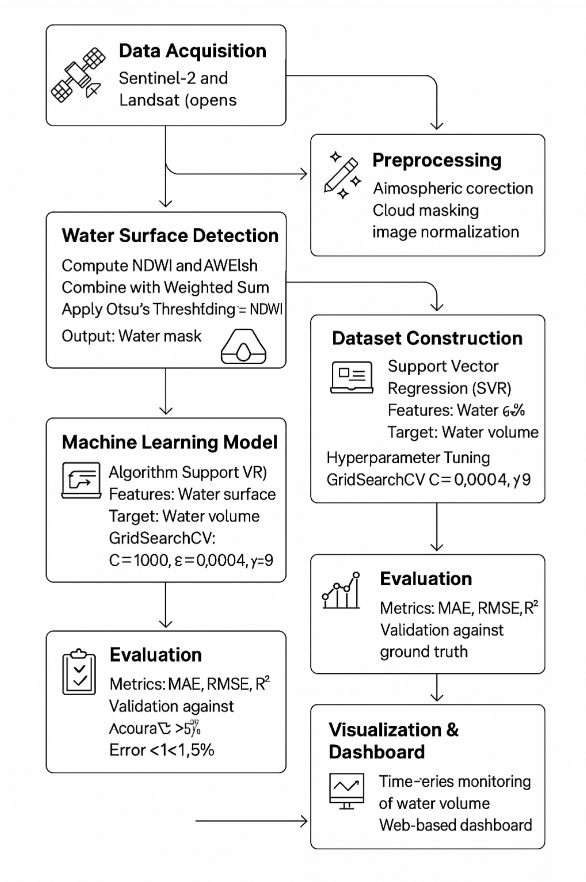


Figure 2: Workflow of the proposed pipeline, from data acquisition and preprocessing to water surface detection and volume estimation.

The proposed architecture is illustrated in Figure 2, outlining the complete workflow—from data acquisition and preprocessing to water surface detection, machine learning-based volume estimation, and dashboard visualization. Although applied here to the QR, this architecture is adaptable and can be replicated for any other water body.

The rest of this paper is organized as follows: Section 2 discusses water spectral indices and introduces a very accurate water segmentation index. Section 3 outlines the implementation of the Support Vector Regression (SVR) model for volume estimation and the data preparation process using bathymetric surveys. Section 4 presents both qualitative and quantitative findings, emphasizing segmentation accuracy and the model's performance based on various metrics. The Conclusion in Section 5 underscores the effectiveness of the proposed sensor-free monitoring approach while acknowledging potential areas for improvement.

2. WATER INDEX

The detection of the water surface is performed using a set of spectral indices supplemented with Otsu's adaptive thresholding. The most widely used index in the literature is the normalized difference in water index ($NDWI$), which is calculated as shown in Equation 1:

$$NDWI = \frac{Green - NIR}{Green + NIR} \tag{1}$$

where $NDWI$ varies from –1 to +1 and the existence of water will yield a positive value between zero and one. The primary use of $NDWI$ is to refine water pixel values to an extreme yielding a bimodal distribution; subsequently, because of the way new pixel values are distributed, Otsu's thresholding technique will be able to successfully capture an optimal threshold for separating water from non-water pixels.

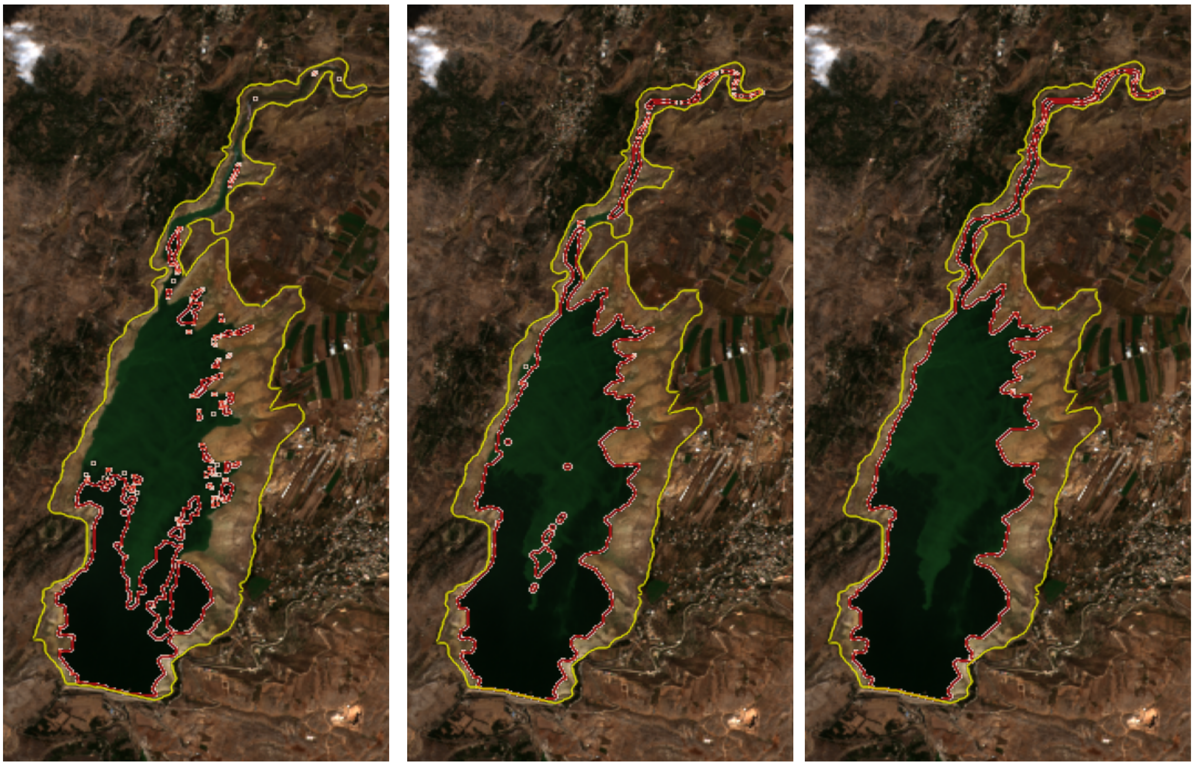
Qaraaoun reservoir has an abundant presence of vegetation and soil, which can be mistaken for water. $NDWI$ is effective in distinguishing areas of vegetation and soil from water. Water typically exhibits low reflectance in both Green and NIR bands, leading to high values of $NDWI$, making it easier to distinguish it from other land covers.

While $NDWI$ is widely used in the literature, multiple studies have shown that spectral indices like $MNDWI$ [9] and $ANDWI$ [10] might outperform $NDWI$ in faithfully separating water pixels from noise especially in turbid and built-up areas. A careful choice of spectral indices is indeed pivotal for accurate water detection.

The effectiveness of water indices for detecting surface water extent varies depending on seasonal and geographic factors. The Qaraaoun Reservoir experiences significant shadowing effects from surrounding mountain chains that alter the optical properties of satellite images. This leads to inaccurate classification, as pixels in shadowed regions exhibit low spectral reflectance. These non-water properties can reduce the accuracy of existing water indices, leading to segmentation errors at different times of the year. Therefore, selecting the most appropriate water index is crucial to ensuring reliable surface water extent detection.

In other words, non-water pixels and water pixels may have similar spectral reflectance. However, while shadowing can pose a problem, its impact is relatively minimal compared to that of soil and vegetation, which significantly hinders accurate segmentation. This observation makes the Automated Water Extraction Index – Non-Shadow ($AWEInsh$) a natural choice, as it is specifically designed to enhance the detection of water features in satellite images. It leverages multiple spectral bands, including the Near Infrared (NIR) and Shortwave Infrared (SWIR) bands, to maximize contrast between water and non-water features [10]. The mathematical formulation for calculating $AWEInsh$ is shown in Equation: 2:

$$AWEInsh = 4 \times (Green - \text{SWIR}_1) - (0.25 \times NIR + 2.75 \times \text{SWIR}_2) \tag{2}$$



(a) $NDWI$ water segmentation results. (b) $AWEInsh$ water segmentation results. (c) $WCWI$ water segmentation results.

Figure 3: Comparison of water segmentation results using different water indices: (a) $NDWI$, (b) $AWEInsh$ and (c) $WCWI$ composite index on a Sentinel-2 imagery from 17 October 2023 where yellow outlines represent the nominal lake contour and red indicate detected water extent.

To make use of both $NDWI$ and $AWEInsh$ indices, a weighted sum referred to as Weighted Composite Water Index ($WCWI$) is proposed here.The newly proposed water index is illustrated in Equation 3:

$$WCWI = 0.8 \times AWEInsh + 0.2 \times NDWI \tag{3}$$

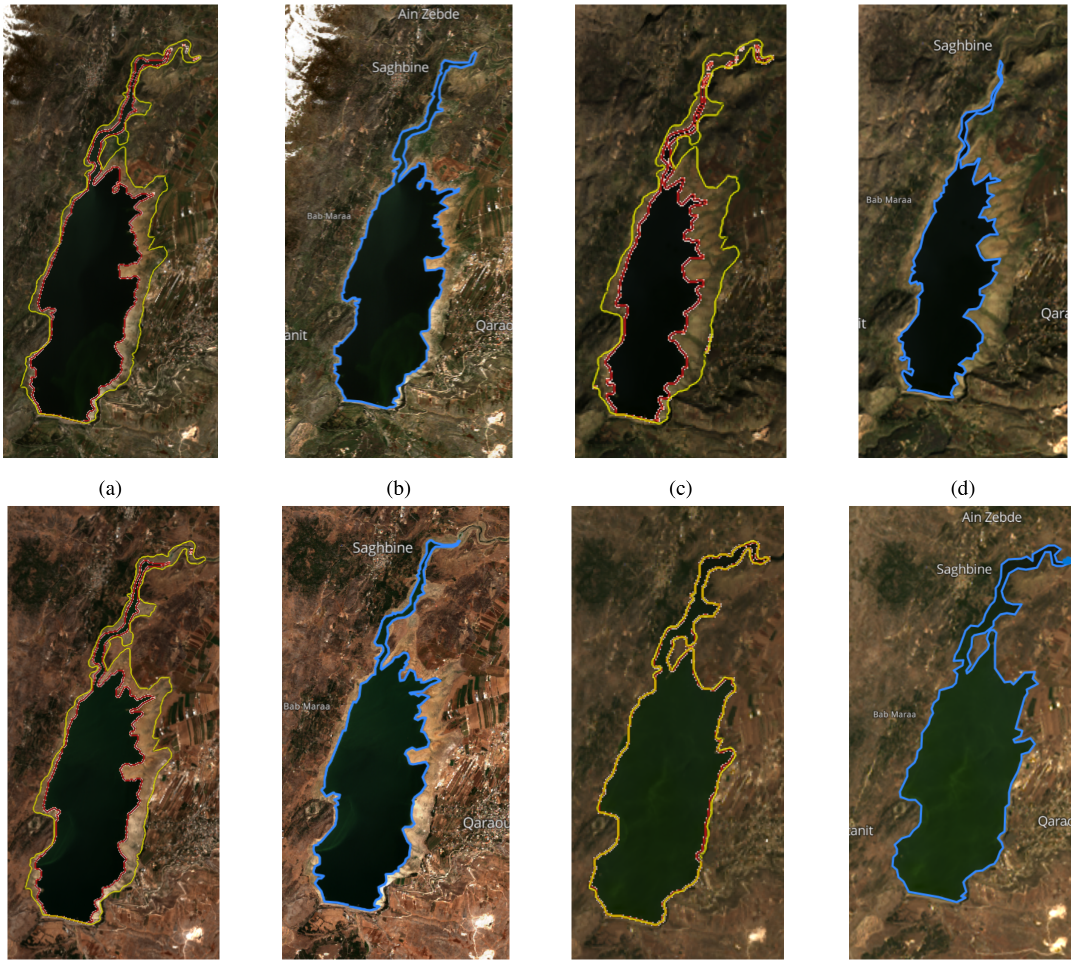
Results in Figure 3 show the performance of the three indices $NDWI$, $AWEInsh$ and $WCWI$ on a Sentinel-2 for the QC dated on 17 October 2024. It is clear in Figure 3(a) that $NDWI$ struggles to segment water extent when color varies sharply in the reservoir. $AWEInsh$ on the other hand results in few segmentation errors towards the center of the reservoir. It also suffers from under-segmentation at the upper narrow upstream channel of the lake within shallow water and minimal width although not easily visible to the naked eye. The under-segmentation in Figure 3(b) reduces the total water surface area, consequently leading to an underestimation of water volume. in contrast, the proposed $WCWI$ produces an accurate water segmentation mask, as demonstrated in Figure 3(c). This analysis was performed on dozens of images across the time-series; however, due to space limitations, we present only one representative example here and additional results are provided in the Results section.

3. MODEL DESIGN

In an effort to improve water management in the Qaraaoun Reservoir, the Litani River Basin Management Support (LRBMS) program was initiated in 2013 and conducted a bathymetric survey of the reservoir. The survey aimed to assess sedimentation by comparing recent depth data with a topographic map from 1950. The survey was conducted through a series of east-west and west-east crossings using a boat equipped with a Doppler flow meter (River Surveyor). This process produced high-resolution depth profiles used to establish an updated level–volume curve that serve as the definitive ground-truth for all subsequent volume estimations [5].

Existing bathymetry data was first digitized and geo-referenced and transformed it into a digital elevation model (DEM). Discrete depth measurements were then interpolated using the "Nearest Neighbor" method to create a continuous surface representing the lake elevation. Using the DEM, a simulation of various water levels that the reservoir might experience was performed. We then utilized ground truth data collected from the hydrometric station on the QR to construct a dataset comprising: *(i)* water level measurements obtained from on-site sensor readings, *(ii)* water surface area extracted from satellite imagery using the proposed accurate water index, and *(iii)* water volume estimates computed using the water level, surface area, and reservoir bathymetry. This dataset was then used to train a machine learning model capable of inferring water volume solely from water surface input, thereby eliminating the need for water level measurements and enabling a sensor-free estimation approach.

Water bodies tend to fill in a nonlinear manner because they possess extremely complex geometry, and thus Support Vector Regression (SVR) appears to be a suitable model of choice for describing the correlation between water surface percent and water volume. In supervised learning, the SVR model was trained from the built dataset with the water surface percentage as the input feature and relative water volumes as the output



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4: Water segmentation results in red versus ground truth values in blue; (a) and (b) Sentinel-2 imagery for 1 March 2023, (c) and (d) LandSat 8 imagery for 21 January 2023, (e) and (f) Sentinel-2 imagery for 26 September 2024 and (g) and (h) Landsat 8 imagery for 8 June 2024. Nominal lake contour is shown in yellow.

target. The data was divided into training and test sets, where 80% of the data was utilized for training and the other 20% for testing. To enhance model convergence as well as accuracy, a Min-Max scaler was applied to both sets to normalize the values of the data to a range of [0,1]. First, an SVR model with an RBF kernel and standard hyper-parameters was instantiated.

Subsequently, a careful hyper-parameter optimization process was conducted with GridSearchCV and 10-fold cross-validation. This vigilant process tuned the critical parameters relative to the adverse mean squared error criterion and calculated optimal hyperparameter values of C = 1000, epsilon = 0.0004, and gamma = 9. The model with these settings gave an average Mean Absolute Error (MAE) of approximately 0.0122 and a Root Mean Squared Error ($RMSE$) of approximately 0.0216 over the test set. The MAE provides an average estimate of the absolute difference between the predicted and actual values, demonstrating high precision, while the $RMSE$, penalizing larger errors more severely, enforces the model to pick up on the underlying nonlinear relationship. This process of training and tuning produced a robust inference model for accurate water volume estimation in the Qaraaoun Reservoir. Additional results are provided in the next section.

4. RESULTS

Results were gathered and analyzed to assess the segmentation and volume estimations obtained from the rigorous solution presented in this paper. Assessing segmentation accuracy was carried out through comparing water segmented imagery obtained from the algorithm described earlier and ground truth imagery. In Figure 4, Sentinel-2 and Landsat-8 imagery were used for qualitative analysis. Figures 4(a), 4(c), 4(e) and 4(g) show that the proposed approach was able to faithfully capture the lake's water extent with high accuracy (more than 95% along the shoreline), in comparison to the ground truth in Figures 4(b), 4(d), 4(f) and 4(h); respectively. The yellow contour represents the lake's nominal boundaries, while the red outline indicates the water segmentation result, and the blue contour shows the manually labeled ground truth water mask. Small under-segmented gaps appear in the narrow upstream channel north of Saghbine and along the eastern shore. Shallow turbid water and wet soil occasionally leads to under-segmentation, which was extremely minimized in our case by relying on the proposed $WCWI$ index.

According to volume time-series shown in Figure 5, the highest recorded surface water coverage occurred on **23 May 1992** with an estimated water volume of **182, 802, 172 m^3**, while the

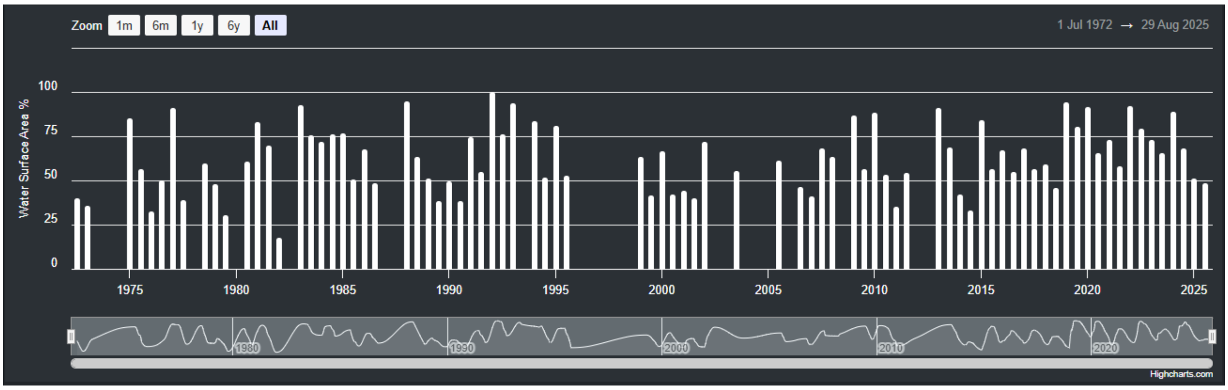


Figure 5: Time series (1973–2025) of water surface area and storage volume in the Qaraaoun Reservoir. July 2025 recorded an exceptionally low volume (49.2×10^6 m³), representing a 66% decline from July 2024 and 56% from July 2023, highlighting an emerging drought signal.

lowest recorded coverage was on **18 January 1982**, with a water volume of $\mathbf{12,491,132\ m^3}$. This time series illustrates the percentage of water surface area and volume estimation over a 50+ year period for the Qaraaoun Reservoir.

Quantitative analysis is grounded upon evaluating different error metrics to highlight different aspects of the SVR model's overall performance against the LRBMS volume measurements. Employed error metrics include Mean Absolute Error (MAE), Root Mean Standard Deviation Ratio (RSR), Mean Absolute Percentage Errors ($MAPE$), Root Mean Square Error ($RMSE$), Coefficients of determination (R^2) and Percent Bias $PBIAS$, chosen in accordance with their relevance to the proposed near real-time water volume monitoring application. MAE and $RMSE$ quantify prediction errors, $MAPE$ represents relative accuracy, while RSR contextualizes error spread against natural variability. R^2 values explain the model's ability to explain observed fluctuations and $PBIAS$ is for assessing tendencies in prediction bias.

As depicted in Table 1, all metrics meet or exceed scholarly thresholds for SVR volume estimations using LandSat imagery only. In 2023, MAE was 5.10 and $RMSE$ 5.70, while in 2024 MAE rose modestly to 7.50 and $RMSE$ to 10.50. These values remain well within acceptable bounds, underscoring practically negligible absolute deviations relative to reservoir storage. Both years achieved $MAPE$ under 6% (4.72% in 2023 and 5.69% in 2024), classifying the forecasts as "highly accurate" [6]. RSR stayed well under 0.7 (0.216 and 0.203 in 2023 and 2024, respectively), indicating statistically robust residual distributions [7]. Coefficients of determination (R^2) of 0.989 and 0.985 far exceed the 0.7 benchmark for satisfactory environmental models, explaining over 98% of volume variance [8]. Finally, $PBIAS$ of –4.57% and –5.14% lies comfortably within ±25%, demonstrating no significant systematic error [6].

For SVR estimations on Sentinel-2-based imagery, across both 2023 and 2024, Table 2 demonstrates again that the model is consistent and maintains strong performance against scholarly benchmarks. In 2023, the model achieved an MAE of 4.53 and a $RMSE$ of 5.22, rising in 2024 to MAE of 7.08 and $RMSE$ of 9.85; absolute deviations that remain negligible in practical terms. Both years produced $MAPE$ below 6% (4.28% in 2023 and 5.22% in 2024), classifying them as "highly accurate" forecasts [7]. RSR remained well under the 0.7 threshold (0.194 and 0.200 in 2023 and 2024 respectively), indicating statistically robust residual distributions [6]. R^2 of 0.988 and 0.983 far exceed the 0.7 benchmark for satisfactory environmental models, explaining over 98% of volume variance [8]. Finally, $PBIAS$ values of –4.11% and –4.74% lie well within the ±25% "satisfactory" range, confirming no significant systematic error [6].

Metric	2023	2024
MAE (10^6 m^3)	5.1	7.5
RMSE (10^6 m^3)	5.7	10.5
MAPE (%)	4.72	5.69
RSR	0.216	0.203
R^2	0.989	0.985
PBIAS (%)	-4.57	-5.14

Table 1: Volume estimation accuracy metrics for **Landsat** Imagery in years 2023 and 2024

Metric	2023	2024
MAE (10^6 m³)	4.53	7.08
RMSE (10^6 m³)	5.22	9.85
MAPE (%)	4.28	5.22
RSR	0.194	0.2
R^2	0.988	0.983
PBIAS (%)	-4.11	-4.74

Table 2: Volume estimation accuracy metrics for **Sentinel-2** Imagery in years 2023 and 2024

The qualitative and quantitative results demonstrate that our solution pipeline reliably segments and predicts Qaraaoun Reservoir volumes across sensors and years. Segmentation aligns with ground truth > 95% of the shoreline, and volume estimates exhibit < 6% relative error, RSR $\approx$ 0.2, R^2 > 0.98, and negligible bias; exceeding established hydrologic and forecasting benchmarks. This confirms the tool's suitability for cost-effective, sensor-free operational monitoring of reservoir storage. Furthermore, a multi-year water persistence analysis (2016–2024) was performed to evaluate the spatial stability of surface water coverage in the reservoir as depicted in Fig. 6. The persistence map reveals a stable central core that remains water covered for more than 300 days annually, while shoreline regions exhibit high variability, with coverage durations often below 150 days. These fluctuating margins correspond to seasonal inflows and interannual droughts, reinforcing the temporal variability observed in the long-term volume record.

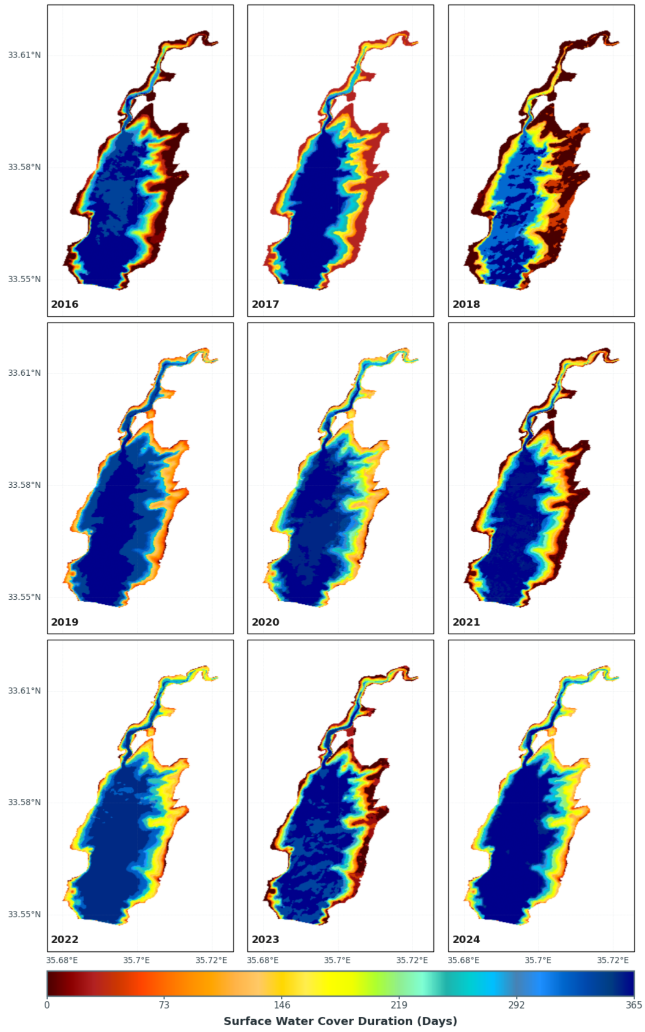


Figure 6: Water persistence map of the Qaraaoun Reservoir (2016–2024), derived from Sentinel-2 imagery. Blue areas indicate permanent water coverage (>300 days), while yellow to red areas represent ephemeral or seasonal water presence (<150 days). The map reveals a stable reservoir core alongside variable shoreline zones, consistent with documented hydrological variability.

5. CONCLUSION

This study proposes an innovative approach to monitor water volume of the Qaraaoun Reservoir in real-time through the utilization of open-source satellite imagery, along with the use of a well-documented solid pipeline for water extent detection and advanced machine learning algorithms. Through thorough examination, highly accurate segmentation and reliable volume estimations were achieved. The results demonstrated how our methods contributed to building a cost-effective, sensor-free solution with minimal error and resilience against environmental challenges. Despite its promising outcomes, this solution presents certain challenges and limitations that come with the reliance on satellite data, such as atmospheric interference, which has a direct impact on segmentation accuracy. In addition to that, the chosen water index must be well-tailored in accordance with the local geographic features and seasonal changes and mechanism of water inflow into the reservoir. However, by combining multiple indices and enhancing the inference model, we were able to pave the way for durable monitoring systems.

The Qaraaoun Reservoir is a vital source for Lebanon's agriculture (more than 40.000 ha) and water supply, which stands to benefit greatly from this tool, especially with limited infrastructure and sensor malfunctions. This research highlights the potential of remote sensing and machine learning in overcoming local challenges and advancing water management practices for developing countries. Moving forward, further optimization of the algorithm and model, alongside more extensive real-time data integration, could contribute to the long-term sustainability and resilience of the reservoir, ensuring its ability to support Lebanon's water needs amid changing climatic conditions. Also, the integration of water quality indices extracted from LandSat and Sentinel imagery would be a valuable addition to the proposed dashboard.

ACKNOWLEDGMENTS

This work was partially supported by the first Call for Proposals for Researchers - SEALACOM project.

REFERENCES

Almawla, S., Hage Chehade, F., & Kaddah, F. (2018). Response of a Lebanese Rock-Filled Dam to Seismic Excitation – The Case of Qaraoun Dam. *WIT Transactions on Engineering Sciences, 121*, 33–45.

Hammoud, N. A., Zervou, S.-K., Kaloudis, T., et al. (2021). Investigation of the Occurrence of Cyanotoxins in Lake Karaoun (Lebanon) by Mass Spectrometry, Bioassays and Molecular Methods. *Toxins, 13*(10), 716.

Zupanc, A., Zupanc, M., Peressutti, D., Aleksandrov, M., Batic, M., Lubej, M., Milcinski, G., & Kirac, K. (2019, May). Bluedot Water Observatory – Cost-effective near real-time monitoring of global water resources [Poster]. Living Planet Symposium. `https://www.researchgate.net/publication/333262569_Bluedot_water_observatory_-_Cost_effective_near_real_time_monitoring_of_global_water_resources`

Sentinel Hub. (n.d.). Sentinel Hub documentation. Retrieved June 19, 2025, from `https://docs.sentinel-hub.com`

International Resources Group. (2013). Litani River Basin Management Support Program: Qaraoun Reservoir Bathymetric Survey (Contract EPP-I-00-04-00024-00, Task Order No. 7). Washington, DC: United States Agency for International Development.

Moriasi, D. N., et al. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE, 50*(3), 885–900.

Hyndman, R. J., & Koehler, A. B. (2006). Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting, 22*(4), 679–688.

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research, 35*(1), 233–241.

Du, Y., Zhang, Y., Ling, X., & Li, W. (2012). Water Bodies’ Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-m Spatial Resolution. *Remote Sensing, 4*(11), 3383–3397. doi:`https://doi.org/10.3390/rs4113383`

Feyisa, G. L., Meilby, H., Fensholt, R., & Proud, S. R. (2014). Automated Water Extraction Index: A New Technique for Surface Water Mapping Using Landsat Imagery. *Remote Sensing of Environment, 140*, 23–35. doi:`https://doi.org/10.1016/j.rse.2013.08.029`