www.gi.sanu.ac.rs, www.doiserbia.nb.rs
J. Geogr. Inst. Cvijic. 2025, 75(2), pp. 217–232

Original scientific paper

UDC: 528.88:004.62(574)
https://doi.org/10.2298/IJGI241025005T

Received: October 25, 2024
Reviewed: February 24, 2025
Accepted: March 11, 2025

IDENTIFICATION OF SMALL LAKES IN KAZAKHSTAN USING REMOTE SENSING DATA

Akmaral Tolepbayeva[1,2]*, *Dinara Abiyeva*[1], *Roza Karagulova*[1], *Aigerim Tanbayeva*[1,3], *Zhanerke Sharapkhanova*[1,2]

[1]Institute of Geography and Water Security, Laboratory of Geographic Information Systems, Almaty, Kazakhstan; e-mail:s tolepbayevaakmaral@gmail.com; dinara.abiyeva5@outlook.com; rozakaragulova@hotmail.com
[2]Al-Farabi Kazakh National University, Department of Geography and Environmental Sciences, Almaty, Kazakhstan; e-mail: zhanerke-sharap@hotmail.com
[3]Kazakh National Agrarian Research University, Almaty, Kazakhstan; e-mail: ai_tanbayeva@outlook.com

Abstract: The purpose of this study is to develop a methodology for automated determination of water surfaces and identification of small lakes in Kazakhstan using cartographic methods and an array of multi-time remote sensing (RS) data. The methodology involved automated surface water classification using multi-temporal Sentinel-2 satellite imagery (spanning the period 2016–2021, focusing on the warm months from May to September), Python-based processing on the Google Earth Engine platform, geographic information system (GIS) based morphometric analysis, and field validation to accurately identify and characterize small lakes in Kazakhstan. The study applied the Normalized Difference Water Index (NDWI), Modified Normalized Difference Water Index (MNDWI), and Normalized Difference Vegetation Index (NDVI) to enhance surface water detection, reduce noise from vegetation, and improve the accuracy of lake boundary delineation from multi-temporal Sentinel-2 imagery. A technique for automated extraction of morphometric characteristics of small lakes has been developed, data on lake morphometry have been obtained. Verification against field measurements demonstrated a high degree of accuracy, with relative error rates of 12% for lake lengths and 13% for the widths. However, challenges such as dense vegetation, high salinity, and the color of shallow lake bottoms led to some classification errors, highlighting the need for further refinement of automated algorithms. As a result, a list of small lakes in Kazakhstan with a surface area from 1 to 10 km^2 was identified.

Keywords: surface water, small lakes, remote sensing data, Sentinel-2, Kazakhstan

1. Introduction

The study of lake surface dynamics offers valuable insights into regional environmental changes, such as climate shifts and anthropogenic impacts. Several studies have explored the use of remote sensing (RS) for water body assessment. For instance, Gourgouletis and Baltas (2023) examined hydroclimatic variables in Western Greece using Earth observation

*Corresponding author, e-mail: tolepbayevaakmaral@gmail.com

data, highlighting RS's ability to provide regular assessments of natural lakes under varying climates. Albarqouni et al. (2022) analyzed spatiotemporal changes in water surface extent in Turkey using Google Earth Engine, showcasing the use of cloud-based technologies for managing large datasets. Zhang et al. (2016) improved water extraction using Landsat TM/ETM+ imagery in Ebinur Lake, China, demonstrating advanced techniques for refining water body assessments. These studies emphasize the increasing reliance on multi-temporal satellite data for water body monitoring.

Despite these advances, there remains a significant gap in the identification and monitoring of small lakes, particularly in regions with vast, inaccessible or under-monitored areas. The existing literature largely focuses on large, well-studied water bodies, leaving smaller lakes under-researched. This gap is critical because small lakes, although often overlooked, can undergo dramatic changes due to climatic variability and human activities, providing crucial early indicators of broader environmental changes (Bai et al., 2025). Amani et al. (2022) highlight this issue by discussing the importance of capturing long-term trends in water surface dynamics, particularly in regions prone to drastic environmental change. Studies by Abiyeva et al. (2021), Jin-Ming et al. (2019), and Li et al. (2019) further emphasize the importance of RS techniques in filling gaps in water monitoring. These studies highlight the role of satellite data in improving the accuracy and timeliness of assessments, particularly in regions with limited ground-based observations.

Kazakhstan faces unique challenges for water monitoring due to its vast and remote areas. The use of remotely sensed datasets, as highlighted by Rylov and Pestunov (2019), has significantly improved the assessment of water bodies, particularly in inaccessible regions. Studies by Peng et al. (2021) and Little et al. (2021) emphasize the importance of RS for monitoring hard-to-reach water bodies. Recent work by Abiyeva et al. (2020), using the Global Surface Water (GSW) dataset, has provided a foundation for more comprehensive studies of Kazakhstan’s water bodies, with Tolepbayeva et al. (2020) further demonstrating the value of satellite imagery in tracking surface water dynamics.

The availability of optical RS data with medium to high spatial resolution, such as WorldView-3 (31 cm), GeoEye-1 (41 cm), SPOT 6/7 (2 m), and Pleiades (30 cm), has advanced water body assessments (Dyldaev et al., 2021; Latipov & Komilova, 2024; Mustafayeva & Tagiyev, 2023). Additionally, Kazakh satellites KazEOSat-1 and KazEOSat-2 provide valuable resources, though their limited observation periods restrict their use for long-term studies (Abiyeva et al., 2020). Operational and archived data from Landsat and Sentinel-2 satellites have proven crucial in monitoring water surface dynamics in Kazakhstan and beyond (Copernicus: Sentinel-2, 2023; Li et al., 2022; Manilyuk & Maslova, 2017; Mueller et al., 2016).

This paper aims to develop an automated method for identifying and monitoring small lakes in Kazakhstan based on Sentinel-2 data (Sentinel-2 Satellite Imagery, 2025). The automated approach will ensure regular monitoring and analysis of surface water dynamics, contributing to effective water resource management. The study can contribute to the research on climate change adaptation and sustainable water resource management, as current methods (such as traditional field surveys, manual mapping, and in-situ measurements) are inadequate for small lakes in remote or difficult-to-access areas (Fedonyuk et al., 2020).

2. Study area

The study primarily focuses on lakes that are relatively small in size, with surface areas between 1 and 10 km^2 (Physical-geographic characteristics of the region, n.d.). These lakes are scattered across different water-economic basins (WEBs) within Kazakhstan (Figure 1). The WEBs are regions that combine natural boundaries with socio-economic factors like water usage, economic activities, and resource management to ensure sustainable water distribution and development (Abiyeva et al., 2021).

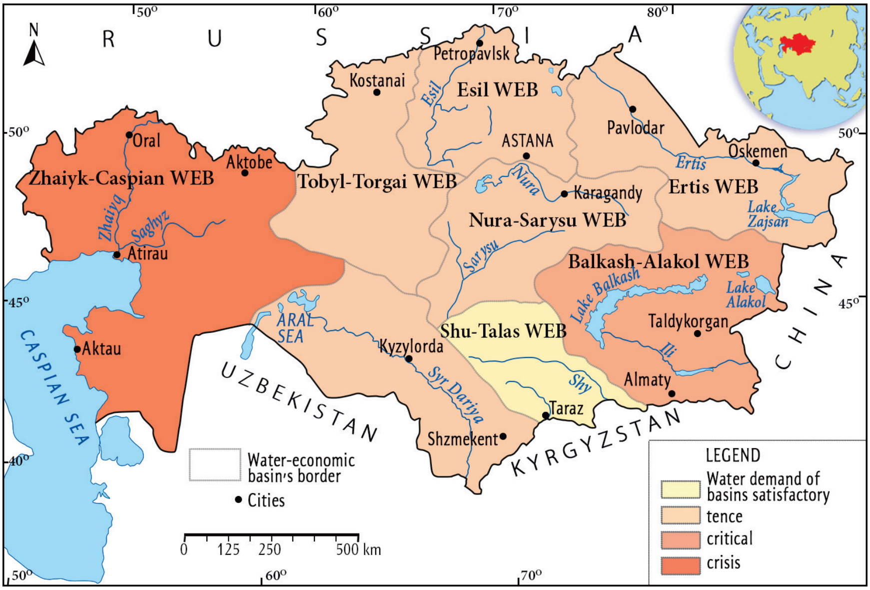

Figure 1. The water basins of Kazakhstan.

Table 1 delineates the principal metric indicators of various WEBs, offering insights into their extent, water reserves, and contributions from water sources. It emphasizes the regional disparities in water availability, the dependence on diverse water sources, and the comparative allocation of water among the basins.

Table 1. Key metric indicators of WEBs in Kazakhstan

WEBs	Area (km^2)	Water reserve (km^3)	Source of water	Percentage from different sources
Aral-Syrdariya basin	345,000	37.9	Mainly from upper basin, tributaries, rivers	70% from upper basin, 20-23% from tributaries, 7-9% from rivers
Balkash-Alakol basin	353,000	149.4	High plains, Balkash Lake, river flow	14% from river flow
Ertis basin	1,643,000	43.8	River flow and lakes	59% from river flow, 16% from lakes, 18% from main water reserves
Zhaiyk-Caspian basin	415,00	28.0	River flows, groundwater, reservoirs	94% from river flows, 3% from groundwater and reservoirs
Esil basin	45,000	5.34	Lakes, river flows, reservoirs	55% from lakes, 34% from river flow, 7% from reservoirs

Table 1. Key metric indicators of WEBs in Kazakhstan *(continued)*

WEBs	Area (km^2)	Water reserve (km^3)	Source of water	Percentage from different sources
Nura-Sarysu basin	60,800	4.59	Rivers, lakes, groundwater, reservoirs	20% from surface sources, 33% from riverbeds, 4% from reservoirs, 25% from groundwater
Shu-Talas basin	64,300	6.11	Rivers, lakes, reservoirs, groundwater	Predominantly surface sources
Tobyl-Torgay basin	214,000	2.9	Groundwater, lakes, water reservoirs, rivers	15% from groundwater, 33% from lakes, 17% from reservoirs, 35% from rivers

Note. Compiled by the authors based on Physical-geographic characteristics of the region (n.d.).

3. Materials and methods

To enhance the clarity and structure of the methodology, a workflow chart has been developed to visually represent the sequence of steps followed in the study (Figure 2). This chart outlines the main stages of the process, which are further broken down into specific substeps.

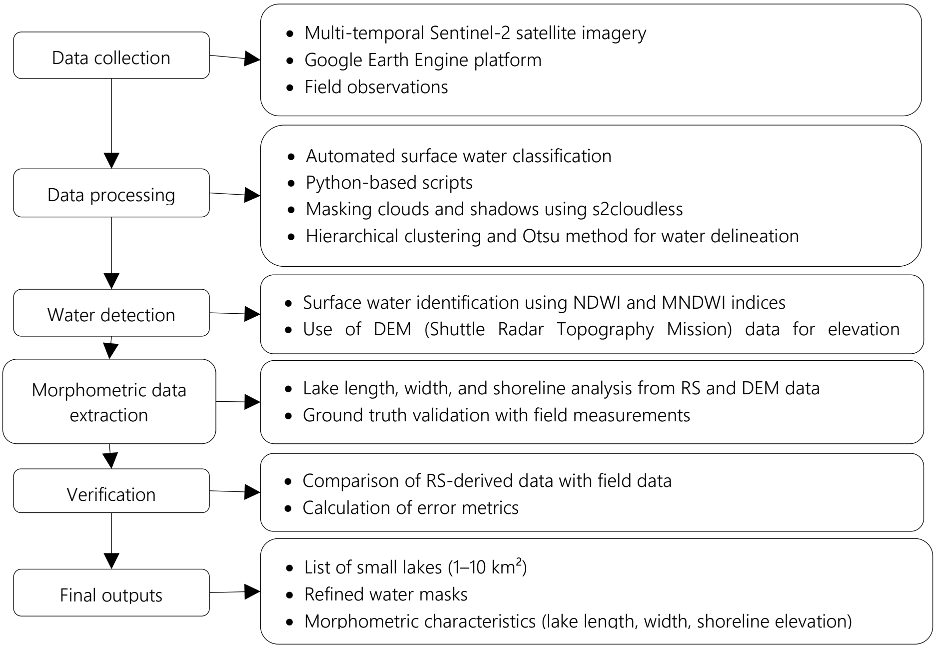

Figure 2. Workflow for identifying small lakes using remote sensing data.

Research on the identification of small lakes with a surface area from 1 to 10 km^2 was conducted in 2 stages: Stage 1 (2020–2021)—Balkash-Alakol, Ertis, Esil, and Nura-Sarysu WEBs; Stage 2 (2021–2022)—Aral-Syrdariya, Shu-Talas, Tobyl-Torgay, and Zhaiyk-Caspian WEBs. The research was conducted in two temporal phases to allow for targeted analysis of different WEBs, accommodating logistical challenges and the unique ecological conditions of each region. This phased approach also provided flexibility in refining methodologies and addressing regional environmental variations, ensuring more precise and context-specific data collection. This study utilized open geospatial and RS datasets to detect surface water and compare lake masks. A combined technique for automated surface water classification using multi-temporal images was developed, employing Sentinel-2 satellite image collections (2A and 2B; Table 2) and the Global Surface Water (GWS) dataset (Global Surface Water Explorer, 2025) for lake detection. Sentinel-2 (Sentinel-2 Satellite Imagery, 2025) Level-1C orthoimage products, with Top-of-Atmosphere reflectance and radiometric and geometric corrections, served as source data. The GSW dataset was used to validate Sentinel-2 imagery by providing long-term water occurrence data, helping to verify lake presence and assess the accuracy of water masks, particularly in areas with limited field data.

Table 2. Key characteristics of Sentinel-2 satellite data

Characteristic	Details
Data source	European Space Agency (ESA)
Resolution	10–60 meters (depending on band)
Spectral bands	13 bands: Visible (blue, green, red), Near-Infrared (NIR), Shortwave Infrared (SWIR)
Band details	Band 1 (443 nm), Band 2 (490 nm), Band 3 (560 nm), Band 4 (665 nm), Band 5 (705 nm), Band 6 (740 nm), Band 7 (783 nm), Band 8 (842 nm), Band 8A (865 nm), Band 11 (1,610 nm), Band 12 (2,190 nm), Band 9 (945 nm, atmospheric correction), Band 10 (1,375 nm, cirrus detection)
Spatial coverage	Global, with specific focus on Kazakhstan's small lakes (1–10 km²)
Temporal resolution	5 days (for paired Sentinel-2 satellites in orbit)
Purpose	Monitoring small lakes, land use/land cover changes, vegetation health, water quality, and environmental change
Data format	JPEG2000, GeoTIFF, and Sentinel-2 products in standard data formats
Advantages	High-resolution imagery, frequent revisit time, multi-spectral data
Disadvantages	Cloud cover can obstruct data collection, lower accuracy in shallow or highly turbid waters, data processing can be time-intensive

Note. Compiled by the authors based on Sentinel-2: Color vision for Copernicus (n.d.).

The methodology included sampling, processing Sentinel-2 data, and automated surface water delineation. Python3 scripts (Python Software Foundation, n.d.) and Google Earth Engine Application Programming Interface (API; n.d.) functionality were implemented. Sentinel-2 Level-2A data (Sentinel-2 Satellite Imagery, 2025), processed using Sen2Cor from the Copernicus Open Access Hub, were utilized. Sen2Cor's five modules handled data reading, processing, and format conversion, performing atmospheric, terrain, and cloud corrections to generate Bottom-of-Atmosphere products with additional maps (Sentinel-2 Satellite Imagery, 2025). The s2cloudless package (Sentinel-2 Satellite Imagery, 2025), a machine learning tool, provided Sentinel-2 image collections and masked clouds and shadows.

Several Python libraries were used for data analysis. NumPy (Numpy, n.d.) and pandas facilitated data manipulation, while matplotlib and seaborn were used for visualizations. Fiona (Fiona, n.d.) handled geospatial data, enabling the integration of shapefiles for grid creation. These libraries together enabled efficient and accurate processing of remote sensing data. Image collections spanned April to September 2015–2020, created separately for each grid (scale ≈ 1:1,000,000) of each water-economic basin (WEB) using shapefiles. The grid was constructed using the Fiona Python library, which imported shapefiles and divided the study area into smaller, standardized units, allowing efficient data processing across basins. The grid played a key role in organizing and structuring the satellite image data, ensuring that the analysis could be performed consistently and efficiently across diverse geographical regions. The s2cloudless algorithm provided cloud presence probabilities

(0–100%) to adjust cloud masking, requiring testing to optimize cloud threshold settings. Different cloud cover thresholds were applied to each grid to ensure comprehensive area coverage, increasing thresholds in regions with frequent cloud cover. Images with over 50% cloud cover were excluded, and grid pixels without sufficient image coverage were marked as No Data values.

The abovementioned method of surface water delineation and lake identification allowed the determination of the preliminary number of small lakes verified by field survey data. The main purpose of the verification was to evaluate the efficiency of this method and to formulate recommendations to solve the encountered problems. To verify the measurements of lake morphometry obtained from RS and field studies, absolute and relative differences between the values were calculated, including the mean absolute error (MAE; Willmott & Matsuura, 2005), root mean square error (RMSE; Chai & Draxler, 2014), correlation coefficient (r; Ratner, 2009), percentage bias (PBIAS; Cole, 1981), and Nash-Sutcliffe efficiency ratio (NSE; Nash & Sutcliffe, 1970) for the final list of lakes (Figure 3).

R	RMSE	MAE	PBIAS	NSE
$\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i-\hat{y}_i\rvert$	$\frac{\sum_{i=1}^{n}(y_i^{obs}-y_i^{sim})}{\sum_{i=1}^{n}(y_i^{obs})}$	$1-\frac{\sum_{t=1}^{T}(Q_m^t-Q_o^t)^2}{\sum_{t=1}^{T}(Q_0^t-\bar{Q}_0)^2}$

Figure 3. Determination of indices for efficiency assessment of the method of automated extraction of morphometric data from RS.

Field observations were conducted to validate the accuracy of the RS data and refine the morphometric analysis of the small lakes in the study area. It was conducted during summers 2020 and 2021 for Stage 1 and summers 2021 and 2022 for Stage 2. Using Global Navigation Satellite System (GNSS) devices, precise geospatial data, including lake boundaries, shoreline heights, and dimensions, were collected directly from the field. In this study, specifically Trimble R8s and Leica GS18 GNSS devices, with precision of up to 2–3 cm, were used. The observations focused on a representative sample of lakes across various water-economic basins, ensuring coverage of diverse geographic and hydrological conditions. Number of lakes chosen for the verification was 183. The process involved selecting small lakes from the initial list based on their characteristics, such as water availability, lake size, and location across different WEBs in Kazakhstan. Only 139 lakes were verified. Some lakes were excluded from field studies due to issues such as high vegetation coverage, salinity, or being dried up. These lakes were not included in the final verification process, ensuring that only those with reliable, accurate measurements were considered.

3.1. Normalized Difference Water Index and Modified Normalized Difference Water Index

The first processing step resulted in a grid for each WEB, as well as a collection of images where cloud cover and cloud shadows were masked in each image. At the second stage, the method of surface water detection from Sentinel-2 multispectral multitemporal images based on hierarchical clustering with machine learning classifier (Hierarchical Clustering in Machine Learning, 2025) using water indices—Normalized Difference Water Index (NDWI; Gao, 1996) and Modified Normalized Difference Water Index (MNDWI; Xu, 2006) was applied. The NDWI index is calculated using the following first equation (Shuka et al., 2011):

$$NDWI = \frac{Green - NIR}{Green + NIR}, \quad (1)$$

where, Green is the green band: 560 nm (S2A)/559 nm (S2B), which corresponds to Band 3 Sentinel-2, and NIR is the near-infrared band: 835.1 nm (S2A)/833 nm (S2B) – Band 8.

This index is designed to maximize the reflectivity of water by using green wavelengths, minimizing the low reflectance of near-infrared radiation from water bodies, and taking advantage of the high reflectivity of NIR from vegetation and soil elements (Shults & Annenkov, 2018). As a result, water bodies have positive values, while vegetation and soil usually have zero or negative values (Albarqouni et al., 2022; Romanchuk et al., 2018).

However, the application of NDWI in regions with built-up areas is problematic. The information on surface water in these regions is often mixed with noise from built-up areas, meaning that many built-up lands have positive values in the NDWI image. Therefore, a modified NDWI was used to solve this problem. MNDWI (Xu, 2006) is calculated as follows (2):

$$MNDWI = \frac{Green - SWIR}{Green + SWIR}. \quad (2)$$

The NDWI was modified by replacing the near-infrared band with Band 12 (SWIR) to create MNDWI, enhancing open water detection while reducing noise from built-up areas, vegetation, and soil. NDWI alone often misestimates water areas due to urban noise, so a combination of NDWI and MNDWI was used for improved surface water information (Skarbøvik et al., 2010; Xu, 2006). An unsupervised clustering algorithm identified water pixels by setting a threshold for water-land separation. Given the variability in water reflectivity across regions and conditions, a fixed threshold was impractical. Instead, the Otsu method (Otsu, 1979) automatically determined optimal thresholds for each grid by maximizing interclass variance in Sentinel-2's near-infrared band (Liao et al., 2001). This approach produced rasters with water mask attributes and clustering results, culminating in a final water mask raster with values of 0 = no water, 1 = water, and 255 = no data/cloud.

3.2. Normalized Difference Vegetation Index

To eliminate errors associated with the recognition of dense vegetation as surface water, average Normalized Difference Vegetation Index (NDVI; Rouse et al., 1974) was calculated for the selected Sentinel-2 collection (equation 3; Han & Niu, 2020; Rusho et al., 2024):

$$NDVI = \frac{NIR - Red}{NIR + Red}. \quad (3)$$

NDVI for Sentinel-2 is calculated using Band 8 (NIR) and Band 4 (Red). When superimposing NDVI data and surface water mask, pixels with values higher than 0.25 are excluded from the mask. The Watershed tool available in Google Earth Engine was used to remove shadows caused by terrain features. The Shuttle Radar Topography Mission (SRTM,

2025) provided the input DEM to aid in shadow removal in areas with elevation differences, notably during water surface detection. Based on the differences in surface water elevation and nearby slopes, shaded areas were highlighted and then excluded when generating surface water output masks for each grid.

4. Results

As a result of automatic classification of a set of multi-temporal Sentinel-2 images, monthly (May–September) and annual surface water masks were obtained. Annual surface water masks provide information on permanent water (pixels classified as water in all images for the period May–September of each year) and seasonal water. The results of automated surface water detection were used to obtain maximum masks of lakes and are one of the main input data for the subsequent process of lake identification and remote retrieval of lake morphometric characteristics.

Based on the results of surface water identification, the masks of water bodies with surface area of more than 1.0 km^2 were derived. These datasets revealed lakes that were not included in the list of surveyed small lakes. The type of water bodies was verified using the GSW thematic datasets and Sentinel Hub services.

The analysis of data accuracy of the relative differences of values allowed the compilation of a list of lakes with a high value of discrepancies between measured and extracted from RS data (modeled) morphometric indicators. Visual analysis of satellite images of lakes, the results of surface water classification, superimposed on the data of field studies, revealed that large discrepancies are associated with the following groups of errors:

- Initial incorrect delineation of the maximum masks of lakes;
- Errors in automated classification of RS data associated with changes in the spectral characteristics of surface water due to the presence of vegetation on the surface of lakes, high salinity, the color of the bottom of shallow lakes (Figures 4);
- Shifting from the lines of morphometric indicators during field studies;
- The length and width of the surveyed lakes were determined by the boundaries of lake basins, and not by water masks, due to the drying up of lakes (Figure 5).

The different line colors in Figures 4 and 5 represent variations in water body classifications, delineation corrections, and shoreline adjustments derived from RS data and ground truth validation.

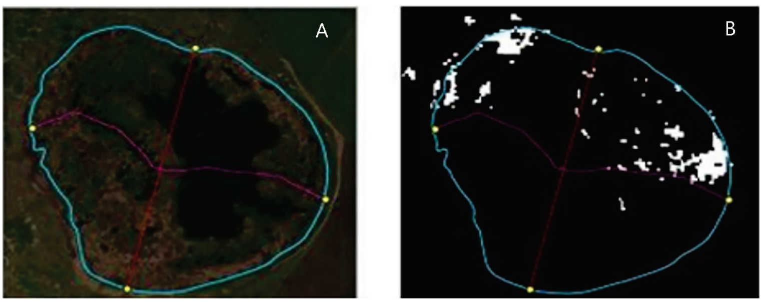

Figure 4. Classification errors associated with changes in the spectral characteristics of surface water due to the presence of vegetation. Lake Kashykbai (Tobyl-Torgay WEB): A) satellite image; B) water mask.

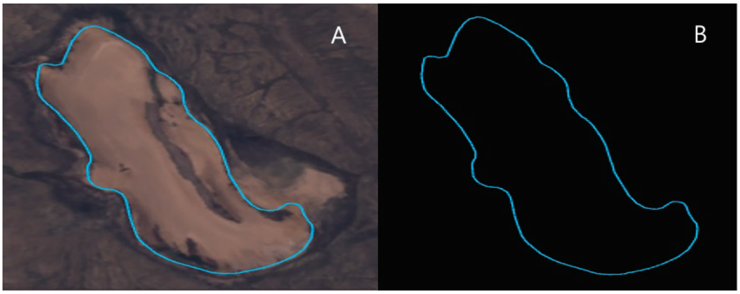

Figure 5. Basin of a dried lake. Lake Shandakkol (Ertis WEB): A) satellite image; B) water mask.

Based on the initial inspection results, the following adjustments were made to the verification process, methodology for retrieving morphometric indicators from RS data, and field data processing workflow:

- Corrected water mask boundaries using field survey data;
- Truncated lines for morphometric measurements at water mask boundaries if they exceeded the maximum water masks;
- Excluded lakes with classification errors due to vegetation and salt crusts;
- Removed dried lakes, using morphometric indicators of lake basins instead.

After these adjustments, data accuracy and method efficiency for extracted morphometric data (RS data) improved significantly. The accuracy of water levels at lake edges from Airbus DEM also showed good results (Tables 2 and 3).

Table 2. An excerpt from the accuracy evaluation of morphometric indicators extracted from RS data (Ertis WEB)

Indicators		Kabantakyr	Kalatuz	Aydarsha	Kuboldy	Sharbukty
Accuracy evaluation of RS-derived values of lake lengths	RS data (m)	2,543	3,710	2,764	2,891	3,160
	Field data (m)	2,539	3,726	2,762	2,922	3,165
	Absolute difference (m)	3	–16	2	–30	–5
	Relative difference (%)	0.1	0.4	0.1	1.1	0.2
Accuracy evaluation of RS-derived values of lake widths	RS data (m)	2,427	2,652	1,706	1,402	2,447
	Field data (m)	2,441	2,674	1,751	1,409	2,469
	Absolute difference (m)	–14	–22	–45	–7	–22
	Relative difference (%)	–1	–1	–3	–1	–1
Accuracy evaluation of lake shorelines (height above the sea level) extracted from	RS data (m)	125.52	103.93	122.19	282.58	130.75
	Field data (m)	125.76	103.6	122.72	283.33	131.61
	Absolute difference (m)	–0.23	0.33	–0.53	–0.75	–0.86
Indicators		Yeskeldy	Tuz	Zhasybay	Dogabassor	Rahman
Accuracy evaluation of RS-derived values of lake lengths	RS data (m)	1,871	5,305	3,475	2,171	2,635
	Field data (m)	1,913	5,315	3,547	2,340	2,723
	Absolute difference (m)	–42	–10	–72	–169	–88
	Relative difference (%)	2.3	0.2	2.1	7.8	3.3
Accuracy evaluation of RS-derived values of lake widths	RS data (m)	1,356	2,701	1,906	1,147	566
	Field data (m)	1,378	2,710	2,004	1,156	620
	Absolute difference (m)	–22	–9	–98	–8	–54
	Relative difference (%)	–2	0	–5	–1	–10

Table 2. An excerpt from the accuracy evaluation of morphometric indicators extracted from RS data (Ertis WEB) *(continued)*

Indicators		Yeskeldy	Tuz	Zhasybay	Dogabassor	Rahman
Accuracy evaluation of lake shorelines (height above the sea level) extracted from	RS data (m)	333.6	134.72	395.38	167.55	1,799.52
	Field data (m)	334.51	135.43	395.59	168.53	1,800.31
	Absolute difference (m)	–0.91	–0.71	–0.2	–0.98	–0.78
Indicators		Kemerkol	Bolshoye	Tuzdykol	Bakshoky	Yrysay
Accuracy evaluation of RS-derived values of lake lengths	RS data (m)	3,011	11,657	5,413	2,073	1,822
	Field data (m)	3,133	11,831	5,551	2,057	2,147
	Absolute difference (m)	–122	–174	–138	17	–325
	Relative difference (%)	4	1.5	2.5	–0.8	17.8
Accuracy evaluation of RS-derived values of lake widths	RS data (m)	1,609	2,453	677	1,472	1,653
	Field data (m)	1,680	2,484	711	1,446	1,757
	Absolute difference (m)	–71	–31	–34	27	–105
	Relative difference (%)	–4	–1	–5	2	–6
Accuracy evaluation of lake shorelines (height above the sea level) extracted from	RS data (m)	855.44	312.09	325.44	523.47	465.85
	Field data (m)	858.19	312.84	325.46	522.14	464.69
	Absolute difference (m)	–2.75	–0.74	–0.01	1.33	1.16
Indicators		Karakol	Muzkol	Shoptikol	Sabyrbay	Karasor
Accuracy evaluation of RS-derived values of lake lengths	RS data (m)	2,714	2,017	1,464	3,020	4,094
	Field data (m)	2,765	2,111	1,522	3,041	4,435
	Absolute difference (m)	–51	–94	–58	–21	–341
	Relative difference (%)	1.9	4.6	4	0.7	8.3
Accuracy evaluation of RS-derived values of lake widths	RS data (m)	1,084	1,167	1,543	986	1,712
	Field data (m)	1,140	1,190	1,566	1,013	1,684
	Absolute difference (m)	–56	–23	–22	–26	28
	Relative difference (%)	–5	–2	–1	–3	2
Accuracy evaluation of lake shorelines (height above the sea level) extracted from	RS data (m)	773.75	124.01	118.03	452.72	348.04
	Field data (m)	774.58	124.36	117.9	451.87	347.59
	Absolute difference (m)	–0.83	–0.35	0.13	0.85	0.45

The analysis of the accuracy indices of lake shoreline heights extracted from the DEM indicates a high level of accuracy of the DEM used. In total, RS data for 139 lakes were verified, where only dried lakes were excluded. More than 70 lakes have shoreline elevation errors resulting in shifts up to 1 m, 37% of lakes have shifts from 1 to 2 m, 6% are within 2–3 m, and only 2 lakes (2.5%) have shifts of 3–4 m. Thus, since high-precision GNSS was used to determine elevations in the field studies, the data extracted from the DEM can be evaluated as sufficiently accurate (Table 3).

Table 3. Indicators of the method efficiency (features with errors related to the detection of the maximum mask were excluded)

Morphometric indicators	r	RMSE	PBIAS	MAE	NSE
Length	0.99	253.48	4.11	117.2	0.97
Width	0.99	136.21	3.49	39.5	0.97
Shoreline height	0.99	3.05	0.17	0.53	0.99

To interpret the results of the efficiency assessment, the efficiency assessment values for hydrological modeling were used (Table 4).

Table 4. Assessment of the efficiency of the method used for the automated determination of morphometric characteristics from RS data

PBIAS	NSE	Efficiency assessment
0–10%	0.75–1	Very good
10–15%	0.65–0.75	Good
15–25%	0.5–0.75	Satisfactory
>25%	<0.5	Unsatisfactory

The assessment of the effectiveness of the method indicates high reliability in terms of extracting information about the dynamics of morphometric characteristics (lake length, lake width), and also points to the suitability of the Airbus WorldDEM4Ortho (WorldDEM4Ortho, 2020) DEM for deriving water level data. However, as mentioned above, this method is not suitable for extracting morphometric characteristics of overgrown and turbid lakes. Based on the analysis of RS data verification results and visual analysis of satellite images of lakes, the contours of lake boundaries were corrected. These modifications allowed the compilation of a list of small lakes of Kazakhstan (with a surface area from 1 to 10 km^2). The total number of lakes across all basins is 2,396, with the highest number found in the Tobyl-Torgay basin (756 lakes) and the lowest in the Shu-Talas basin (40 lakes). The distribution of small lakes in other regions is as follows: Esil basin—583, Nura-Sarysu basin—229, Ertis basin—472, Balkash-Alakol basin—64, Aral-Syrdariya basin—99, Zhaiyk-Caspian basin—153. The observed variation is reflective of the disparate geographical and hydrological conditions that prevail across these basins. This comprehensive enumeration of minor lacustrine bodies constitutes a substantial contribution to the comprehension of Kazakhstan's hydrological resources, particularly in the context of climate change and the associated challenges of regional water management.

5. Discussion

The findings of this study illustrate the efficacy of employing RS data for the identification and categorization of minor lakes in Kazakhstan. The development and implementation of a method integrating Sentinel-2 and GSW datasets has made a significant contribution to the automated detection and morphometric analysis of water bodies with areas ranging from 1 to 10 km^2. The capacity to delineate surface water areas and detect seasonal and permanent water bodies using multi-temporal satellite data exemplifies the potential of this method to facilitate continuous and precise monitoring of lakes, particularly in remote and inaccessible locations (Annenkov et al., 2023; Shults & Annenkov, 2023).

The study's methodology, which integrates long-term, multi-temporal remote sensing data from 2016 to 2021 with field observations, allows for the identification of 2,396 small lakes across several WEBs in Kazakhstan, accounting for the variability of these water bodies. However, the full impact of human-induced climate change requires a longer timeframe to distinguish between natural variability and anthropogenic effects. The identification of such a high number of lakes across various WEBs demonstrates their sensitivity to environmental changes, but it is important to note that this is a snapshot over a limited period. This finding corroborates the findings of previous research conducted by Gourgouletis and Baltas (2023), who studied the hydroclimatic variables affecting lakes in Greece. The utilization of RS data revealed considerable year-to-year fluctuations in lake surface areas, driven by environmental conditions, which is consistent with the high variability observed in this study. Furthermore, the research by Albarqouni et al. (2022) also demonstrated the dynamic nature of small lakes, particularly in response to climatic conditions such as dry winters and low precipitation years. The researchers employed the Google Earth Engine to assess lake surface temperatures and water extent in Turkey. As in the present study, their findings emphasize the usefulness of RS data in capturing temporal changes in water bodies across different geographic regions.

The veracity of the morphometric data obtained through RS was corroborated by field observations, particularly regarding the dimensions of the lakes. The high correlation values between RS-derived morphometric parameters and field measurements serve to reinforce the reliability of the developed method. In this context, the results are comparable to those reported by Abiyeva et al. (2020), who evaluated the suitability of GSW datasets for studying lake dynamics in Kazakhstan. Similarly, their study emphasized the value of RS data in providing accurate and up-to-date information on water bodies in regions where ground-based observation posts are limited or non-existent. This result is consistent with the findings of Peng et al. (2021), who utilized multisource RS images to monitor water surface changes in Dongting Lake, China. The aforementioned study demonstrated the efficacy of integrating DEM data with satellite imagery to obtain precise measurements of water body characteristics, thereby further substantiating the methodology employed in the present research.

This study had limitations, particularly in classifying lakes with dense vegetation or high salinity, which hindered the spectral characteristics used for automated water identification, leading to misclassification in some cases. This issue was also noted by Zhang et al. (2016), who faced similar challenges in Xinjiang, China, where vegetation and soil noise impacted the accuracy of water indices, requiring modified indices like MNDWI for better detection. The findings have important implications for water resource management in Kazakhstan. As noted by Amani et al. (2022), RS technologies are essential for understanding long-term water body dynamics. The methodology developed here offers a scalable, cost-effective solution for continuous monitoring of small lakes, adaptable to regions with similar conditions.

This study could also inform water management policies and strategies (Xu, 2006). Moreover, it contributes to the growing field of environmental monitoring using satellite data. In arid regions like Kazakhstan, where water bodies are vulnerable to climate fluctuations, multi-temporal Sentinel-2 data enhances understanding of lake dynamics and supports further climate change impact studies on freshwater resources (Liao et al., 2001; Karches, 2023). In areas where ground-based monitoring is impractical, RS data provide high-frequency, high-resolution insights to inform sustainable water management (Fedoniuk & Skydan, 2023; Shumka, Kalogianni, et al., 2020; Shumka, Sumka, et al., 2020).

The integration of Sentinel-2 data, GIS tools, and field validation has proven to be an effective approach for mapping and analyzing water bodies, despite challenges such as vegetation interference and shallow water issues. The study also underscores the importance of refining classification techniques to address these challenges and improve accuracy. The insights gained from the accuracy assessment and error analysis provide valuable directions for future improvements in remote sensing-based water detection and morphometric analysis.

6. Conclusion

This research established an automated approach for the identification and analysis of tiny lakes in Kazakhstan utilizing multi-temporal Sentinel-2 data, resulting in an updated inventory of 2,396 small lakes. The method demonstrated significant accuracy in defining lake boundaries and extracting essential morphometric properties; nevertheless, obstacles such as dense vegetation, elevated salinity, and lake desiccation resulted in classification errors. These findings underscore the efficacy of remote sensing data in the surveillance of small lakes, sometimes neglected in hydrological studies, and illustrate the promise of this methodology for ongoing and extensive environmental monitoring.

This study's theoretical implications contribute to the expanding field of remote sensing for water body monitoring, providing a significant methodology for areas with restricted access to terrestrial observations. The concept can be utilized to enhance water resource management and climate change adaption strategies, particularly in arid countries such as Kazakhstan. Future research must concentrate on optimizing algorithms to alleviate classification challenges posed by vegetation, salinity, and desiccated lakes, while integrating supplementary datasets to improve the precision of lake identification and morphometric assessment.

References

Abiyeva, D. K., Karagulova, R. K., Ayupov, K. A., Tolepbayeva, A. K., & Urazbayeva, G. M. (2020). Analysis of suitability of global surface water data sets for studying the dynamics of the lake areas in Kazakhstan. *Issues of Geography and Geoecology*, *3*, 33–42.

Abiyeva, D. K., Karagulova, P. K., Nysanbaeva, A. S., Abayev, N. N., Urazbayeva, G. M., Radusnova, O. V., Sharapkhanova, Z. M., & Tolepbayeva, A. K. (2021). Geospatial web application to support climate change research in Kazakhstan. *InterCarto InterGIS*, *27*, 108–119. https://doi.org/10.35595/2414-9179-2021-3-27-108-119

Albarqouni, M. M. Y., Yagmur, N., Bektas Balcik, F., & Sekertekin, A. (2022). Assessment of Spatio-Temporal Changes in Water Surface Extents and Lake Surface Temperatures Using Google Earth Engine for Lakes Region, Türkiye. *ISPRS International Journal of Geo-Information*, *11*(7), Article 407. https://doi.org/10.3390/ijgi11070407

Amani, M., Kakooei, M., Ghorbanian, A., Warren, R., Mahdavi, S., Brisco, B., Moghimi, A., Bourgeau-Chavez, L., Toure, S., Paudel, A., Sulaiman, A., & Post, R. (2022). Forty Years of Wetland Status and Trends Analyses in the Great Lakes Using Landsat Archive Imagery and Google Earth Engine. *Remote Sensing*, *14*(15), Article 3778. https://doi.org/10.3390/rs14153778

Annenkov, A., Medvedskyi, Y., Demianenko, R., Adamenko, O., & Soroka, V. (2023). Preliminary accuracy assessment of low-cost UAV data processing results. *International Conference of Young Professionals "GeoTerrace 2023"*, *2023*, 1–5. Lviv, Ukraine. https://doi.org/10.3997/2214-4609.2023510014

Bai, B., Mu, L., & Tan, Y. (2025). A Global Lakes/Reservoirs Surface Extent Dataset (GLRSED): An integration of multi-source data. *Geoscience Data Journal*, *12*(1), Article e285. https://doi.org/10.1002/gdj3.285

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development Discussions*, *7*, 1247–1250. http://dx.doi.org/10.5194/gmdd-7-1525-2014

Cole, N. S. (1981). Bias in testing. *American Psychologist*, *36*(10), 1067–1077. https://doi.org/10.1037/0003-066X.36.10.1067

Copernicus: Sentinel-2. (2023). https://directory.eoportal.org/web/eoportal/satellite-missions/c-missions/copernicus-sentinel-2

Dyldaev, M., Chymyrov, A., Mukabaev, A., & Omurzak Uulu, O. (2021). Investigation of the population area of snow leopard in the Central Tian-Shan Mountains. *E3S Web of Conferences, 227*, Article 02002. https://doi.org/10.1051/e3sconf/202122702002

Fedoniuk, T. P., & Skydan, O. V. (2023). Incorporating geographic information technologies into a framework for biological diversity conservation and preventing biological threats to landscapes. *Space Science and Technology*, *29*(2), 10–21. https://doi.org/10.15407/knit2023.02.010

Fedonyuk, T. P., Galushchenko, O. M., Melnichuk, T. V., Zhukov, O. V., Vishnevskiy, D. O., Zymaroieva, A. A., & Hurelia, V. V. (2020). Prospects and main aspects of the GIS-technologies application for monitoring of biodiversity (on the example of the Chornobyl Radiation-Ecological Biosphere Reserve). *Space Science and Technology, 26*(6), 75–93. https://doi.org/10.15407/knit2020.06.075

Fiona. (n.d.). Fiona (Version 1.8.18) [Computer Software]. https://pypi.org/project/fiona/

Gao, B. (1996). NDWI – A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Global Surface Water Explorer. (2025). https://global-surface-water.appspot.com/

Google Erath Engine API. (n.d.). https://developers.google.com/earth-engine/reference/rest

Gourgouletis, N., & Baltas, E. (2023). Investigating Hydroclimatic Variables Trends on the Natural Lakes of Western Greece Using Earth Observation Data. *Sensors*, *23*(4), Article 2056. https://doi.org/10.3390/s23042056

Han, Q., & Niu, Z. (2020). Construction of the Long-Term Global Surface Water Extent Dataset Based on Water-NDVI Spatio-Temporal Parameter Set. *Remote Sensing*, *12*(17), Article 2675. https://doi.org/10.3390/rs12172675

Hierarchical Clustering in Machine Learning. (2025). https://www.geeksforgeeks.org/hierarchical-clustering/

Jin-Ming, Y., Li-Gang, M., Cheng-Zhi, L., Yang, L., Jian-li, D., & Sheng-Tian, Y. (2019). Temporal-spatial variations and influencing factors of Lakes in inland arid areas from 2000 to 2017: a case study in Xinjiang. *Geomatics, Natural Hazards and Risk*, *10*(1), 519–543. https://doi.org/10.1080/19475705.2018.1531942

Karches, T. (2023). Towards a Dynamic Compartmental Model of a Lamellar Settler. *Symmetry*, *15*(4), Article 864. https://doi.org/10.3390/sym15040864

Latipov, N. F. O., & Komilova, N. K. (2024). The Importance of Anthropogenic Influence in Assessing the Ecological Condition of Cities in Uzbekistan (in the Case of Navoi). *Baghdad Science Journal, 21*(4), 1214–1219. https://doi.org/10.21123/bsj.2023.8892

Li, B., Zhang, Y.-C., Wang, P., Du, C.-Y., & Yu, J.-J. (2019). Estimating Dynamics of Terminal Lakes in the Second Largest Endorheic River Basin of Northwestern China from 2000 to 2017 with Landsat Imagery. *Remote Sensing*, *11*(10), Article 1164. https://doi.org/10.3390/rs11101164

Li, M., Hong, L., Guo, J., & Zhu, A. (2022). Automated Extraction of Lake Water Bodies in Complex Geographical Environments by Fusing Sentinel-1/2 Data. *Water*, *14*(1), Article 30. https://doi.org/10.3390/w14010030

Liao, P. S., Chen, T. S., & Chung, P. C. (2001). A Fast Algorithm for Multilevel Thresholding. *Journal of Information Science and Engineering*, *17*(5), 713–727. https://doi.org/10.6688/JISE.2001.17.5.1

Little, S., Pavelsky, T. M., Hossain, F., Ghafoor, S., Parkins, G. M., Yelton, S. K., Rodgers, M., Yang, X., Crétaux, J.-F., Hein, C., Ullah, M. A., Lina, D. H., Thiede, H., Kelly, D., Wilson, D., & Topp, S. N. (2021). Monitoring Variations in Lake Water Storage with Satellite Imagery and Citizen Science. *Water*, *13*(7), Article 949. https://doi.org/10.3390/w13070949

Manilyuk, T. L., & Maslova, L. V. (2017). The Research of Water Surfaces of Lakes by Applying of Water Index Based on the Remote Sensing Data. *Transbaikal State University Journal, 23*(3), 4–11. https://www.researchgate.net/publication/319468079_The_Research_of_Water_Surfaces_of_Lakes_by_Applying_of_Water_Index_Based_on_the_Remote_Sensing_Data

Mueller, N., Lewis, A., Roberts, D., Ring, S., Melrose, R., Sixsmith, J., Lymburner, L., McIntyre, A., Tan, P., Curnow, S., & Ip, A. (2016). Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. *Remote Sensing of Environment*, *174*, 341–352. https://doi.org/10.1016/j.rse.2015.11.003

Mustafayeva, E., & Tagiyev, A. (2023). Perspective of using groundwater in the Ganikh-Ayrichay foothills. *Reliability: Theory and Applications, 18*(5), 136–141. https://doi.org/10.24412/1932-2321-2023-575-136-141

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I – A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

NumPy. (n.d.). NumPy (Version 1.18) [Computer Software]. NumPy team. https://numpy.org/

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Peng, Y., He, G., Wang, G., & Cao, H. (2021). Surface Water Changes in Dongting Lake from 1975 to 2019 Based on Multisource Remote-Sensing Images. *Remote Sensing*, *13*(9), Article 1827. https://doi.org/10.3390/rs13091827

Physical-geographic characteristics of the region. (n.d.). http://www.cawater-info.net/bk/water_land_resources_use/english/docs/fiziko_geog_kharack_kazakhstan.html

Python Software Foundation. (n.d.). Python (Version 3.0) [Computer Software]. Python Software Foundation. https://www.python.org/

Ratner, B. (2009). The correlation coefficient: Its values range between +1/−1, or do they? *Journal of Targeting, Measurement and Analysis for Marketing*, *17*, 139–142. https://doi.org/10.1057/jt.2009.5

Romanchuk, L., Fedonyuk, T., Pazych, V., Fedonyuk, R., Khant, G., & Petruk, A. (2018). Assessment of the stability of aquatic ecosystems development on the basis of indicators of the macrophytes fluctuating asymmetry. *Eastern-European Journal of Enterprise Technologies, 4*(10(94)), 54–61. https://doi.org/10.15587/1729-4061.2018.141055

Rouse, J. W., Haas, R. H., Scheel, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the great plains with ERTS. In *Proceedings, 3rd Earth Resource Technology Satelite (ERTS) Symposium, vol. 1* (pp. 48–62). Washigton: NASA, Goddard Space Flight Center. https://ntrs.nasa.gov/citations/19740022614

Rusho, M. A., Azizova, R., Mykhalevskiy, D., Karyonov, M., & Hasanova, H. (2024). Advanced earthquake prediction: unifying networks, algorithms, and attention-driven LSTM modelling. *International Journal of GEOMATE*, *27*(119), 135–142. https://doi.org/10.21660/2024.119.m2415

Rylov, S. A., & Pestunov, I. A. (2019). Assessment of lakes areas by Sentinel-2 satellite data. *Journal of Siberian Federal University. Engineering & Technologies*, *12*(5), 526–535. https://doi.org/10.17516/1999-494X-0108

Sentinel-2 Satellite Imagery. (2025). https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-2

Sentinel-2: Color vision for Copernicus. (n.d.). https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2

Shuka, L., Çullaj, A., Shumka, S., Miho, A., Duka, S., & Bachofen, R. (2011). The Spatial and Temporal Variability of Limnological Properties of Bovilla Reservoir (Albania). *Water Resources Management, 25*(12), 3027–3039. https://doi.org/10.1007/s11269-011-9788-z

Shults, R., & Annenkov, A. (2018). Investigation of the different weight models in Kalman filter: A case study of GNSS monitoring results. *Geodesy and Geodynamics, 9*(3), 220–228. https://doi.org/10.1016/j.geog.2017.09.003

Shults, R., & Annenkov, A. (2023). BIM and UAV photogrammetry for spatial structures sustainability inventory. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 48*(5/W2-2023), 99–104. https://doi.org/10.5194/isprs-archives-XLVIII-5-W2-2023-99-2023

Shumka, S., Kalogianni, E., Šanda, R., Vukić, J., Shumka, L., & Zimmerman, B. (2020). Ecological particularities of the critically endangered killifish Valencia letourneuxi and its spring-fed habitats: A long-lost endemic species of south Albania. *Knowledge and Management of Aquatic Ecosystems, 421*, Article 45. https://doi.org/10.1051/kmae/2020036

Shumka, S., Shumka, L., Trajce, K., & Ceci, S. (2020). First record of the Western Greece goby - Economidichthys pygmaeus (Holly, 1929), in Greater Prespa Lake (Albania). *Ecologica Montenegrina, 35*, 78–81. https://doi.org/10.37828/EM.2020.35.6

Skarbøvik, E., Shumka, S., Mukaetov, D., & Nagothu, U. S. (2010). Harmonised monitoring of Lake Macro Prespa as a basis for integrated water resources management. *Irrigation and Drainage Systems, 24*(3–4), 223–238. https://doi.org/10.1007/s10795-010-9099-1

SRTM Shuttle Radar Topography Mission. (2025). https://www.earthdata.nasa.gov/data/instruments/srtm

Tolepbayeva, A. K., Tanbayeva, A. A., Karagulova, R. K., Iskaliyeva, G. M., Zhakupova, A. A., Urazbayeva, G. M., & Lentschke, J. (2020). Estimation o25.10f the changes in water surfase area based on the use of archives satellite images of Landsat satellites (on the example of the Ertis river). *News of the Academy of Sciences of the Republic of Kazakhstan*, *3*(441), 79–87. https://doi.org/10.32014/2020.2518-170X.57

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*(1), 79–82. https://doi.org/10.3354%2Fcr030079

WorldDEM4Ortho (Version 2) Technical Product Specification. (2020). https://earth.esa.int/eogateway/documents/20142/37627/WorldDEM4Ortho-Technical-Specification.pdf

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*(14), 3025–3033. https://doi.org/10.1080/01431160600589179

Zhang, F., Tiyip, T., Johnson, V. C., Wang, J., & Nurmemet, I. (2016). Improved water extraction using Landsat TM/ETM+ images in Ebinur Lake, Xinjiang, China. *Remote Sensing Applications: Society and Environment*, *4*, 109–118. https://doi.org/10.1016/j.rsase.2016.08.001