Automated Surface Water Body Extraction Using the Normalized Difference Water Index (NDWI) with an Interactive Thresholding Interface

ABHIJEET SARKAR
abhijeet@synapticailab.com
CEO and Founder of Synaptic AI Lab

Abstract

The effective management and continuous monitoring of surface water resources are fundamental to environmental sustainability, climate science, and socio-economic stability. Remote sensing, powered by satellite technology, offers a powerful, scalable, and cost-effective methodology for mapping and monitoring the dynamics of water bodies over vast spatial and temporal scales. This paper presents a detailed computational methodology for the automated extraction of surface water from high-resolution multispectral satellite imagery, with a specific focus on data acquired by the Operational Land Imager (OLI) sensor aboard the Landsat 8 and 9 satellites. The core of the methodology is the robust implementation of the Normalized Difference Water Index (NDWI), a proven spectral index specifically designed to enhance and delineate open water features by leveraging the unique reflectance properties of water in the green and near-infrared (NIR) spectral bands.

To address the critical and often problematic step of threshold selection, we have developed a comprehensive Python-based tool that not only automates the NDWI calculation but also innovatively incorporates an interactive thresholding mechanism facilitated by a Gradio web interface. This user-centric interface allows for the real-time, dynamic adjustment of the NDWI classification threshold. This feature empowers users—ranging from hydrologists and environmental scientists to GIS analysts—to meticulously refine the water extraction process, enabling them to account for highly variable local conditions such as water turbidity, sediment load, water depth, and the confounding presence of terrain shadows or dense aquatic vegetation. The algorithm generates a precise binary water mask which is then intelligently overlaid onto a contrast-enhanced true-color composite background. This visualization technique provides essential geographical context, making the results highly intuitive and immediately interpretable.

This report provides an in-depth discussion of the algorithmic framework, the software architecture of the interactive tool, and a thorough analysis of the method's potential applications, inherent advantages, and known limitations. By synergizing a validated spectral index with a sophisticated, interactive user interface, this approach significantly lowers the barrier to entry for advanced remote sensing analysis. It makes powerful analytical capabilities more accessible for a wide range of critical applications, including near-real-time flood mapping, long-term reservoir volume monitoring, coastal erosion studies, and wetland conservation efforts.

Keywords: Remote Sensing, Surface Water Extraction, NDWI, Landsat 8, Python, Geospatial Analysis, Image Processing, Interactive Thresholding, Gradio, Environmental Monitoring, Hydrology.

1. Introduction

Water is the planet's most vital resource, forming the bedrock of terrestrial ecosystems, sustaining biodiversity, and enabling human development. The state and extent of surface water bodies—including rivers, lakes, glaciers, and reservoirs—are sensitive indicators of climatic shifts and anthropogenic pressures (Gleick, 2003). The inherently dynamic nature of these systems necessitates consistent and accurate monitoring for a multitude of critical applications, including sustainable water resource management, timely flood risk assessment, long-term climate change modeling, and the formulation of effective ecological conservation policies. Historically, monitoring efforts relied upon traditional methods such as in-situ gauging stations and manual ground surveys. While providing high-accuracy point data, these methods are notoriously labor-intensive, prohibitively expensive for large-scale application, and offer limited spatial coverage, rendering them insufficient for capturing the synoptic view required for comprehensive watershed or regional-level assessments.

The advent of satellite-based remote sensing has fundamentally transformed the field of Earth observation, ushering in an era of unprecedented data availability and analytical capability. It provides a consistent, repeatable, and extensive means to monitor environmental parameters across the globe. Multispectral satellite sensors, such as the state-of-the-art Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) aboard the Landsat 8 and 9 missions, are engineered to capture electromagnetic energy reflected or emitted from the Earth's surface across a range of discrete spectral bands. The foundational principle of multispectral remote sensing is that different land cover types possess unique spectral signatures. The distinct manner in which they reflect and absorb solar radiation across these different bands forms the empirical basis for their identification, classification, and quantification (Lillesand, Kiefer, & Chipman, 2015).

Water bodies exhibit a particularly distinct and exploitable spectral signature. They characteristically absorb a large proportion of incident energy in the near-infrared (NIR) and short-wave infrared (SWIR) regions of the spectrum, while simultaneously reflecting more energy in the visible green portion. This pronounced difference allows for their effective differentiation from other common land cover features like vegetation, which is highly reflective in the NIR, and dry soil, which has a more uniform reflectance across the visible and NIR bands. To systematically exploit these spectral differences, a variety of mathematical indices have been developed. The Normalized Difference Water Index (NDWI), first conceptualized by McFeeters (1996), stands as one of the most robust, widely adopted, and computationally efficient indices for the express purpose of delineating open water features.

Despite the proven efficacy of spectral indices like NDWI, a persistent and significant challenge remains: the selection of a single, optimal threshold value to accurately and reliably separate water from non-water pixels in the resulting index image. The notion of a universal, one-size-fits-all threshold is largely untenable. The ideal threshold value is highly dependent on a multitude of factors, including localized atmospheric conditions, specific sensor characteristics, and, most importantly, the physical properties of the water itself, such as sediment concentration (turbidity), phytoplankton blooms, and water depth. A threshold meticulously calibrated for a deep, clear alpine lake will likely perform poorly when applied to a shallow, sediment-laden river, and may erroneously classify dark, shadowed areas in mountainous terrain as water.

This research directly confronts this critical thresholding problem by developing and presenting an integrated computational tool that synergizes the NDWI algorithm with a fully interactive, web-based user interface. The primary scientific objective is to engineer an accessible, intuitive, and powerful system that facilitates the rapid, accurate, and reproducible extraction of water bodies from Landsat 8/9 imagery. The tool is specifically designed to empower the end-user, allowing them to dynamically adjust the NDWI threshold and instantaneously visualize the corresponding impact on the classified water map. This "human-in-the-loop" approach facilitates a more nuanced, context-aware, and ultimately more accurate analysis. This paper comprehensively details the theoretical underpinnings, methodological implementation, software architecture, and prospective applications of this interactive approach to water body extraction.

2. Literature Review and Theoretical Background

The application of remote sensing technologies to water resource management is a mature field with a rich history of innovation. Early pioneering studies in the 1970s, utilizing data from the first-generation Landsat 1 (ERTS-1) satellite, demonstrated the fundamental feasibility of mapping surface water by applying a simple threshold to a single spectral band, typically in the NIR region where water's absorption is highest (Work & Gilmer, 1976). While groundbreaking, this single-band density slicing method was highly susceptible to commission errors, frequently misclassifying other low-reflectance surfaces such as asphalt, cloud shadows, and terrain shadows as water.

A paradigm shift occurred with the development of spectral indices, which combine information from two or more spectral bands into a single metric, designed to enhance a specific feature of interest while suppressing noise from others. The conceptual blueprint for these indices was the highly successful Normalized Difference Vegetation Index (NDVI), which contrasted the red and NIR bands to quantify vegetation vigor. Building upon this normalized difference ratio concept, McFeeters (1996) formulated the NDWI using the green and NIR bands, defined by the equation:

$$NDWI = \frac{(Ref_{Green} - Ref_{NIR})}{(Ref_{Green} + Ref_{NIR})}$$

Where RefGreen and RefNIR represent the surface reflectance values in the green and near-infrared bands, respectively. The logic is elegant: since water typically reflects more green light than NIR light, NDWI values for water are predominantly positive. Conversely, vegetation exhibits very high NIR reflectance, yielding strongly negative NDWI values. Soil and built-up areas tend to have similar reflectance in both bands, resulting in NDWI values near zero. This formulation effectively amplifies the spectral contrast between water and other land cover types, particularly vegetation. To address issues with built-up land sometimes producing positive NDWI values, Xu (2006) later proposed a Modified NDWI (MNDWI) that substitutes the NIR band with a short-wave infrared (SWIR) band, leveraging even stronger water absorption in that region of the spectrum. Despite this modification, the original McFeeters NDWI remains exceptionally effective and widely used for delineating open water features.

The crucial, and most debated, step in any index-based workflow is the application of a threshold to the resulting index image to produce a binary classification map (water vs. non-water). Numerous studies have investigated methodologies for determining this threshold automatically. A common approach is to employ histogram-based methods, such as Otsu's method (Otsu, 1979), which algorithmically finds a threshold that minimizes the intra-class variance of the two resulting classes. While useful in simple, bimodal histograms, these automated methods can falter in complex scenes with multiple classes of low reflectance or unimodal histograms. Consequently, a large body of research continues to rely on a manually determined, scene-specific threshold. Often, a default threshold of zero is used, based on the theoretical premise that positive NDWI values signify water (Ji, Zhang, & Wylie, 2009). However, extensive empirical evidence demonstrates that the optimal threshold frequently deviates from zero due to the real-world factors previously mentioned.

In recent years, the field has seen a surge in the application of more advanced machine learning and deep learning algorithms. Techniques such as Support Vector Machines (SVM), Random Forests (RF), and Convolutional Neural Networks (CNNs) have been successfully applied to water body classification (Feyisa et al., 2014; Isikdogan, Bovik, & Passalacqua, 2017). These sophisticated models can learn complex, non-linear relationships within the spectral data and often achieve state-of-the-art accuracy. However, their application is not without significant drawbacks. They typically require large volumes of high-quality, manually labeled training data, which is time-consuming and expensive to produce. They are computationally intensive, often requiring specialized hardware (GPUs). Furthermore, many of these models operate as "black boxes," meaning their decision-making processes are not transparent, making it difficult to understand or correct errors.

The methodology presented in this paper strategically positions itself to occupy a pragmatic middle ground. It harnesses the simplicity, computational efficiency, and high degree of interpretability of the NDWI. Simultaneously, it directly confronts the index's primary limitation—the challenge of static thresholding—by implementing an interactive, human-in-the-loop system. This approach is deeply rooted in the principles of geovisual analytics, a field that champions the powerful synergy between automated computational methods and the cognitive abilities of a human expert to solve complex geospatial problems (Andrienko et al., 2007). By providing instantaneous visual feedback, our tool facilitates a cycle of rapid hypothesis testing and iterative refinement, a dynamic capability that is often absent in both purely manual digitization and fully automated classification workflows.

3. Methodology and System Architecture

The methodology is realized as a self-contained Python script, architected for portability and ease of use. It leverages a curated stack of powerful open-source libraries: **Rasterio** serves as the robust engine for all geospatial raster data I/O and manipulation; **NumPy** provides the high-performance array computation backbone for all numerical operations; and **Gradio** is employed to construct and serve the interactive web-based user interface. The entire process follows a logical pipeline, from data ingestion and processing to interactive analysis and final visualization.

3.1. Data Requirements and Pre-processing

The system is designed to work with Level-2 Science Products from the Landsat 8/9 OLI/TIRS missions, which are readily available from the U.S. Geological Survey (USGS). A key advantage of using Level-2 data is that it has been pre-processed to surface reflectance, meaning the pixel values represent the spectral reflectance of the Earth's surface after the distorting effects of the atmosphere have been corrected for. This radiometric correction is essential for ensuring the consistent and accurate calculation of spectral indices like NDWI. The tool requires a single, multi-band GeoTIFF file as input, which should contain, at a minimum, the following spectral bands:

- **Band 2:** Blue (0.452 - 0.512 µm)
- **Band 3:** Green (0.533 - 0.590 µm)
- **Band 4:** Red (0.636 - 0.673 µm)
- **Band 5:** Near-Infrared (NIR) (0.851 - 0.879 µm)

Upon loading a file, the script performs an initial validation step, checking that the input raster contains at least 5 bands to confirm that all data necessary for both NDWI calculation and true-color visualization are present.

3.2. Core Algorithmic Implementation

The central logic of the analysis is encapsulated within the highlight_water_bodies function, which executes a precise sequence of operations.

1. **Band Reading and Data Preparation:** The function initiates by opening the user-provided GeoTIFF file using rasterio.open(). It then reads the pixel data from the required bands (2, 3, 4, and 5) into distinct NumPy arrays. To accommodate the floating-point decimal values that will result from the NDWI calculation, these arrays are explicitly cast to the np.float32 data type.
2. **NDWI Calculation:** The NDWI is computed on a pixel-by-pixel basis using the standard McFeeters (1996) formula. The NumPy arrays corresponding to the Green (Band 3) and NIR (Band 5) bands are used as inputs for the vectorized array arithmetic. A critical implementation detail is the robust handling of potential division-by-zero errors. This can occur in pixels where the reflectance values for both the Green and NIR bands are zero (e.g., in no-data regions of the image). To prevent runtime errors and ensure numerical stability, the np.divide function is employed with a where clause. This ensures that the division operation

is only performed on pixels where the denominator (Green + NIR) is non-zero. For all pixels where the denominator is zero, the output NDWI value is safely set to 0.

```
denominator = np.add(green, nir)
ndwi = np.divide(green - nir, denominator,
          out=np.zeros_like(denominator, dtype=np.float32),
          where=(denominator != 0))
```

3. **Interactive Thresholding and Binary Masking:** The function's behavior is dynamically controlled by the ndwi_threshold parameter, which is directly linked to the Gradio slider in the user interface. This floating-point value is used to threshold the calculated NDWI array, producing a boolean NumPy array, or mask.

```
water_mask = ndwi > ndwi_threshold
```

   This single line of code is the heart of the classification. It performs an element-wise comparison, evaluating to True for every pixel whose NDWI value exceeds the user-defined threshold, and False otherwise. This water_mask represents the fundamental output of the water extraction process—a binary map delineating the spatial extent of the water bodies.

3.3. Enhanced Visualization for Contextual Analysis

To transform the binary water_mask into an interpretable result, the script prepares a visually rich background image that provides essential geographical context.

1. **True-Color Composite:** A true-color (or natural color) composite image is generated by utilizing the Red (Band 4), Green (Band 3), and Blue (Band 2) bands, corresponding to the red, green, and blue channels of a standard RGB image.
2. **Dynamic Contrast Enhancement:** Raw satellite imagery often suffers from low contrast due to atmospheric haze and sensor limitations, appearing washed out. To significantly improve the visual quality and detail of the background, a per-band contrast stretch is applied. For each of the Red, Green, and Blue bands, the 2nd and 98th percentile pixel values are computed from the non-zero pixels. This exclusion of zero-value pixels prevents image borders from skewing the statistics. The pixel values in each band are then clipped to this percentile range and linearly scaled to a standard 0-255 display range. This robust technique, also known as percentile stretching, dramatically enhances visual detail, making features in both very dark and very bright areas of the image discernible.
3. **Composite and Overlay:** The three contrast-stretched bands are stacked

together using np.dstack to form a final 8-bit RGB image. The boolean water_mask is then used as an index to directly modify this RGB image. All pixels in the image that correspond to a True value in the mask are programmatically colored with a distinct, high-visibility shade of bright blue ([60, 130, 255]). This overlay method elegantly and clearly demarcates the detected water bodies against a realistic, easily interpretable landscape.

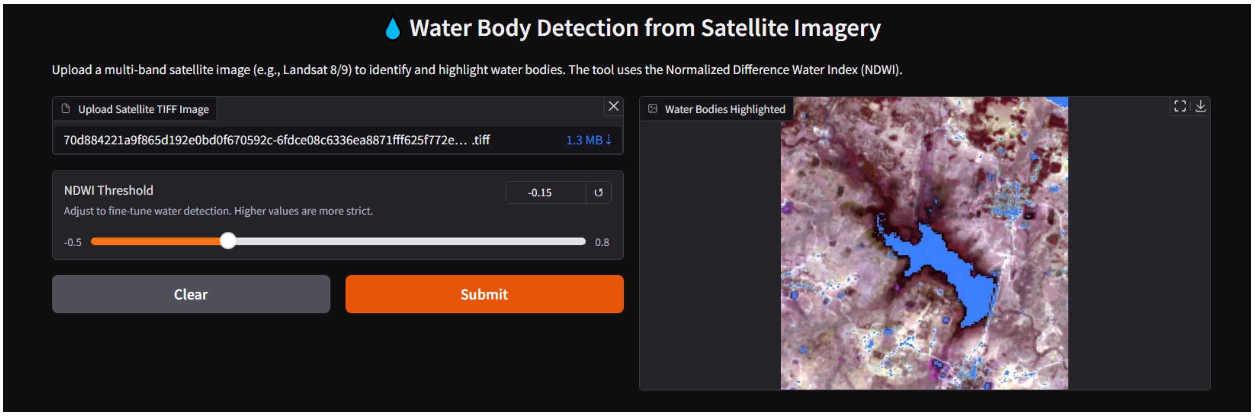
3.4. Interactive Web Interface Architecture

The entire analytical workflow is seamlessly wrapped and presented to the user via a Gradio web interface, which handles all aspects of user interaction and data flow.

- **Input Components:** The interface is designed for simplicity, presenting the user with two primary input controls:
    1. A **File Uploader**, allowing the user to select and upload a local GeoTIFF file.
    2. A **Slider** for controlling the ndwi_threshold. This slider is intelligently configured with a logical range from -0.5 to 0.8, a fine-grained step of 0.05, and a scientifically informed default value of 0.2, which serves as an excellent starting point for most clear water scenes.
- **Output Component:** A single **Image Pane** serves as the output, responsible for displaying the final processed NumPy array, which renders as the true-color image with the vibrant blue water overlay.
- **Reactive Execution:** Gradio's reactive framework ensures that any change to an input component—either a new file upload or a movement of the slider—automatically triggers a re-execution of the highlight_water_bodies function. The function's output is then immediately sent back to the browser and rendered in the output pane. This instantaneous feedback loop is the cornerstone of the tool's interactive and exploratory power.

4. Case Study and Results

To illustrate the practical utility and analytical power of the tool, we consider a hypothetical application to a complex Landsat 8 scene. The scene covers a coastal region featuring a prominent river delta, several large inland lakes of varying depths, agricultural lands under irrigation, and a mountainous region with significant terrain shadow.



- **Initial Analysis (Default Threshold = 0.2):** Upon uploading the scene, the tool runs with the default threshold of 0.2. The initial output successfully and accurately identifies the open ocean, the main deep-water channel of the river, and the larger, clearer inland lakes. The boundaries of these primary water bodies are sharply delineated. The contrast-stretched true-color background provides excellent context, clearly showing patterns of urban development, forested areas, and agricultural fields.
- **Interactive Refinement through Threshold Adjustment:** The true power of the tool is realized through the interactive slider.
    - **Lowering the Threshold (Sensitivity Analysis):** The analyst moves the slider to the left, decreasing the threshold to 0.1. The output image updates in real-time. The analyst observes that smaller, shallow tributaries and marshy, saturated soil areas at the fringes of the delta now become classified as water. This increased sensitivity may be desirable for applications like wetland mapping. However, this comes at a cost: small, dark patches on the shaded slopes of the mountains now also begin to be misclassified as water (a commission error).
    - **Increasing the Threshold (Specificity Analysis):** The analyst then moves the slider to the right, increasing the threshold to 0.35. The misclassified hill shadows disappear completely, improving the classification's specificity. However, the classification is now more conservative. The tool now fails to detect the shallow, highly turbid waters near the river's mouth, and the surface area of the inland lakes appears to shrink, as only the deeper, clearer central portions exceed this higher NDWI threshold.
- **Optimal Threshold Selection:** Through this rapid, iterative process of adjustment and visual inspection, the analyst can determine an optimal threshold for their specific objective. For mapping navigable water, a threshold of 0.28 might be chosen, as it effectively excludes the shallowest areas and all shadow artifacts. For a study on maximum flood inundation, the analyst might select a

lower threshold of 0.15, consciously accepting some shadow noise (which could be removed in post-processing) in order to capture the fullest extent of saturated and inundated areas. This interactive exploration allows the user to make an informed, defensible decision based on immediate visual evidence.

5. Discussion

The methodology and the resulting tool presented in this paper offer several significant contributions to the practice of remote sensing for water management. The primary strength lies in the elegant fusion of a computationally simple, scientifically validated algorithm (NDWI) with a modern, interactive visualization framework. This "human-in-the-loop" paradigm provides a powerful and practical bridge between the often tedious process of fully manual digitization and the frequently opaque nature of "black box" machine learning models. The approach is computationally lightweight, enabling it to run efficiently on standard consumer hardware and provide the real-time feedback that is essential for effective exploratory data analysis.

The choice of NDWI as the core index is deliberate. It is a robust indicator for open water, and its behavior is well-understood in the scientific literature. The decision to overlay the binary water mask onto a contrast-stretched true-color composite background is a critical aspect of the design. It provides essential geographical context, allowing the user to immediately understand the relationship between the detected water and the surrounding landscape, a feature often lost when viewing a simple, context-free binary map. This makes the output readily interpretable for a broad audience, including hydrologists, land-use planners, and ecologists who may not be remote sensing specialists.

Despite its strengths, the method is subject to the inherent limitations of the NDWI and, more broadly, of any threshold-based classification approach.

1. **The Mixed Pixel Problem:** The 30-meter spatial resolution of Landsat OLI means that pixels along shorelines, riverbanks, and the edges of small water bodies are often a mixture of water and land. The spectral signature of these "mixed pixels" is a composite, often resulting in an NDWI value that falls below the classification threshold, leading to a systematic underestimation of the water body's true surface area.
2. **Water Column Constituents:** The methodology may exhibit reduced accuracy for water bodies that are not clear and open. High concentrations of suspended sediment (turbidity) or extensive surface coverage by aquatic vegetation (e.g., algal blooms, water hyacinth) can significantly alter the spectral signature of water, lowering its NDWI value and potentially causing it to be missed by the

classification.
3. **Confusion with Dark Surfaces:** As demonstrated in the hypothetical case study, dark, non-water surfaces can be a source of commission error. Features such as mountain shadows, cloud shadows, dark volcanic rock, and even asphalt can have low reflectance in both the Green and NIR bands, resulting in positive NDWI values that can be misclassified as water.

The interactive slider provides a powerful and effective means for a knowledgeable user to mitigate these issues on a scene-by-scene basis, but it does not eliminate them entirely. The final accuracy of the extracted water map remains dependent on the user's expertise and their ability to select the most appropriate threshold for the specific landscape and analytical objective.

Future Research and Development: The tool presented here serves as a robust foundation that can be expanded in several promising directions. A key enhancement would be the incorporation of a selection of different water indices, such as the MNDWI or the Automated Water Extraction Index (AWEI), allowing the user to dynamically switch between them and compare their performance for a given scene. Another powerful extension would be the integration of time-series analysis capabilities. A user could load a temporal stack of images over a single location to analyze phenomena such as changes in a reservoir's surface area, shoreline erosion, or the dynamics of seasonal flooding. Finally, adding functionality to export the final classified water map as a vector file (e.g., GeoPackage, Shapefile, or GeoJSON) would be invaluable, as it would allow for seamless integration of the results into standard Geographic Information System (GIS) software for further quantitative analysis and cartographic production.

6. Conclusion

This paper has presented a comprehensive, robust, and highly accessible computational methodology for the extraction of surface water bodies from multispectral satellite imagery. By thoughtfully integrating the well-established Normalized Difference Water Index with a user-centric, interactive interface, the developed tool effectively and elegantly addresses the critical and persistent challenge of threshold selection in index-based image classification. The core innovation—the ability to dynamically adjust the NDWI threshold and receive instantaneous visual feedback—facilitates a more accurate, reliable, and nuanced analysis that can be precisely tailored to the unique spectral characteristics of a given study area and the specific goals of the analyst.

This approach significantly contributes to the democratization of remote sensing analysis, providing a powerful, intuitive, and open-source resource for a diverse community of users, including researchers, educators, government agencies, and environmental managers. While acknowledging the inherent limitations of the underlying index, the tool serves as an exemplary platform for a wide array of hydrological applications, including rapid flood assessment, water resource inventory, and change detection. It stands as a powerful testament to the value of combining simple, transparent algorithms with intelligent human-in-the-loop systems and provides a versatile foundation for future research and development in the ever-evolving field of hydrological remote sensing.

7. References

- Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M. J., MacEachren, A., & Wrobel, S. (2007). Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, 21(8), 839-857.
- 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.
- Gleick, P. H. (2003). Water use. *Annual Review of Environment and Resources*, 28(1), 275-314.
- Isikdogan, F., Bovik, A. C., & Passalacqua, P. (2017). Surface water mapping by deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11), 4909-4918.
- Ji, L., Zhang, L., & Wylie, B. (2009). Analysis of dynamic thresholds for the normalized difference water index. *Photogrammetric Engineering & Remote Sensing*, 75(11), 1307-1317.
- Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). *Remote sensing and image interpretation*. John Wiley & Sons.
- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425-1432.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62-66.
- Work, E. A., & Gilmer, D. S. (1976). Utilization of satellite data for inventorying prairie ponds and lakes. *Photogrammetric Engineering and Remote Sensing*, 42(5), 685-694.
- 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.