Taylor & Francis
Taylor & Francis Group

GIScience & Remote Sensing

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tgrs20

Unsupervised surface water mapping with airborne LiDAR data by leveraging physical properties of water

Hunsoo Song & Jinha Jung

To cite this article: Hunsoo Song & Jinha Jung (2025) Unsupervised surface water mapping with airborne LiDAR data by leveraging physical properties of water, GIScience & Remote Sensing, 62:1, 2437252, DOI: 10.1080/15481603.2024.2437252

To link to this article: https://doi.org/10.1080/15481603.2024.2437252

GISCIENCE & REMOTE SENSING
2025, VOL. 62, NO. 1, 2437252
https://doi.org/10.1080/15481603.2024.2437252

Unsupervised surface water mapping with airborne LiDAR data by leveraging physical properties of water

Hunsoo Song[a,b] and Jinha Jung[a]

[a]Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, USA; [b]Yale School of the Environment, Yale University, New Haven, CT, USA

ABSTRACT

Comprehensive mapping of surface water, especially smaller bodies of water (<1 ha), remains challenging due to the lack of robust and scalable extraction methods. Traditional methods require the use of either training procedures or the repetitive tuning of site-specific parameters, which present hurdles to automated mapping and introduce biases tied to training data and parameters. The dependence on water's reflectance properties, including LiDAR intensity, further complicates the issue, as higher-resolution images inherently introduce increased noise. In response to these challenges, we propose a unique, unsupervised method that focuses on the geometric characteristics of water instead of its variable reflectance properties. Unlike existing approaches, our method relies exclusively on 3D coordinate observations from airborne LiDAR data, taking advantage of the presumption that connected surface water remains flat due to surface tension. Leveraging this physical constraint and spatial connectivity, our method precisely extracts water bodies of diverse sizes and reflectance without the need for training procedures or intensive parameter tuning. Notably, by relying solely on 3D coordinate observations, our approach significantly facilitates the fully automated generation of comprehensive 3D topographical maps of both water and terrain, eliminating the need for human intervention or supplementary optical imagery. We validated the robustness and scalability of this method across diverse terrains, including urban, coastal, and mountainous areas. Overall, the proposed method achieved a 11% higher accuracy, measured by intersection over union, compared to the highly competitive NDWI-based method. Moreover, it proved its effectiveness in both accuracy and scalability compared to supervised machine learning and deep learning methods.

ARTICLE HISTORY
Received 8 April 2024
Accepted 28 November 2024

KEYWORDS
Water body extraction; airborne laser scanning; unsupervised; small water bodies; hydrography; digital elevation model

1. Introduction

Surface water mapping in remote sensing provides vital insights fundamental to diverse studies. It facilitates understanding the location, dynamics, and relationship of water with its surroundings, which is critical for water resource management (Hoekstra, Buurman, and Van Ginkel 2018) and human welfare (Sanders et al. 2022). The evolution of remote sensing has advanced the quest for a comprehensive surface water map, now covering even minor, small water bodies. Historically, large water bodies took precedence in mapping endeavors (Bijeesh and Narasimhamurthy 2020; Ji, Zhang, and Wylie 2009; Khandelwal et al. 2017; Pekel et al. 2016), with recent emphasis shifting to smaller water entities due to their societal relevance and the rise of high-resolution imaging (Hoekstra, Buurman, and Van Ginkel 2018; Kelly-Quinn et al. 2022; Kyzivat and Smith 2023; Xu et al. 2020).

However, high-resolution surface water mapping presents challenges. These primarily arise from the variability in reflectance of small water bodies in such images. This challenge is further compounded by the need for multiple image acquisitions to cover larger areas, which significantly elevates spectral variability (F. Chen et al. 2020a; Ogilvie et al. 2018). Addressing these challenges has inspired methods integrating multi-temporal data (Pickens et al. 2020), sensor fusion (Liu et al. 2022; Tayer et al. 2023), and machine learning (Ko, Kim, and Nam 2015; Wang et al. 2023). Yet, inherent spectral variations in water, influenced by acquisition conditions (Martins et al. 2017) and surface conditions (Ogilvie et al. 2018), pose limitations.

Furthermore, current methodologies largely offer only 2D information, whereas understanding water's elevation is essential for various applications (Arrighi and Campo 2019; Musa, Popescu, and Mynett 2015).

CONTACT Jinha Jung jinha@purdue.edu

This 3D perspective, possible with LiDAR data, can enhance our grasp of water dynamics. Though airborne laser scanning (ALS) has been utilized for adding elevation for water bodies (Moore et al. 2019) and holds significant potential, automated surface water mapping using ALS is still in nascent stages.

In our research, we present a scalable solution for surface water mapping that harnesses topographic ALS. By “scalable,”we refer to a method capable of delivering consistent and reliable results across a wide range of diverse and expansive environments, adapting to both aquatic surfaces and their surrounding terrestrial landscapes without the need for site-specific parameter adjustments or human intervention. Our method introduces a novel water elevation-based region merging, leveraging the assumption that surface water remains relatively flat. It recognizes minimal elevation variations caused by wind or flow but considers them negligible compared to changes occurring with adjacent terrestrial bodies. It operates unsupervised, relying solely on LiDAR’s coordinate information. Notably, our method’s unique independence from training procedures facilitates the automatic generation of a comprehensive 3D topographic map using only ALS.

2. Related works

2.1. Optical and radar-based water mapping

Water mapping has largely been achieved through optical imaging with various satellite missions, offering resolutions from 1-km to 1-m, paving the way for decades of Earth’s water body mapping (Khandelwal et al. 2017; Pekel et al. 2016). Advancements in imaging technology have sparked a surge in mapping smaller water bodies over extensive areas (Becker et al. 2019; Chen et al. 2020a).

Although high-resolution imagery captures small water bodies, large-area mapping of these is more intricate than using low-resolution images for vast areas. Challenges include varied spectral reflectance in smaller bodies (Ogilvie et al. 2018) and disruptions like shadows and urban structure occlusions (Yang et al. 2018). Moreover, high-resolution imaging demands varied methods for different image acquisitions (Ji, Zhang, and Wylie 2009).

To counter inaccuracies in high-resolution mapping, advanced decision rules have emerged. Studies have utilized multiple indices to reduce urban area errors (Yang et al. 2018) or introduced water index roughness (Dong et al. 2022). Data-driven techniques, such as machine learning-based classifications (e.g. support vector machines (Sun et al. 2015) and random forest classifiers (Ko, Kim, and Nam 2015)) and deep learning-based semantic segmentation methods (Y. Chen et al. 2020b; Isikdogan, Bovik, and Passalacqua 2019; Luo, Tong, and Hu 2021; Ma et al. 2023; Wieland et al. 2023), have been explored. However, sourcing high-quality training samples for expansive areas, particularly in high-resolution images, is a significant hurdle (Pekel et al. 2016; Pickens et al. 2020). Furthermore, the variability in water’s optical properties and the data distribution shift inherent in machine learning approaches complicate the reuse of trained models (Tuia, Persello, and Bruzzone 2016; Wieland et al. 2023).

Radar-based water mapping, combined with optical imagery, is gaining prominence due to its cloud-penetrating capabilities (Liu et al. 2022). However, its limited high spatial resolution data renders it less cost-effective for large-scale mapping of small water bodies.

2.2. Airborne LiDAR-based water mapping

For surface water mapping, topographic ALS is less commonly used than other methods, excluding bathymetric LiDAR (Szafarczyk and To 2023). The essence of topographic LiDAR in this context is to leverage the intensity values of airborne LiDAR and the fact that LiDAR point density within water areas is often lower than in non-water areas. Specifically, when an ALS flight strip passes over a water body, the number of reflective returns can decrease sharply as the angle of the laser deviates from nadir. This significant dropout occurs due to the specular, or directional, properties of light reflection on water surfaces, in contrast to the more diffuse reflection observed on non-water surfaces (Höfle et al. 2009).

The foundational work in this area was conducted by Brzank et al., utilizing a supervised fuzzy classification method using height, intensity, and point density (Brzank et al. 2008). Similarly, HÖfle et al. employed both intensity and point density, modeled them with sensing geometry, and conducted seeded region growing segmentation (Höfle et al. 2009).

Smeeckaert et al. utilized support vector machines using historical coastline data (Smeeckaert et al. 2013). Likewise, Crasto et al. used a decision tree approach, leveraging LiDAR's point density, elevation, and intensity (Crasto et al. 2015). The machine learning trend continued with studies such as those by Malinowski et al., integrating radiometric and geometric laser point features of full-waveform LiDAR into supervised classifiers for flood mapping (Malinowski et al. 2016).

The fusion of LiDAR with other datasets also emerged. For instance, Irwin et al. merged airborne LiDAR data with SAR imagery to reduce surface water overestimation in vegetated areas (Irwin et al. 2017). Another study by ProÅ¡ek et al. used support vector machine by combining multiple LiDAR-based variables with hyperspectral imagery (Prošek et al. 2020). Beyond water body-specific mappings, other machine learning methods, specifically employing richer datasets such as multi-spectral LiDAR (Pan et al. 2019; Yu et al. 2021, 2022; Zhao et al. 2021) or hyperspectral data (Akwensi, Kang, and Wang 2023; Guo et al. 2023), have exhibited promising results.

Machine learning methods, particularly when integrated with other remote sensing data, offer the potential for high accuracy. However, these methods necessitate significant human intervention. To address, a notable advancement in this domain is the introduction of the scan line intensity – elevation ratio (SLIER) (Shaker, Yan, and LaRocque 2019; Yan 2023; Yan, Shaker, and LaRocque 2019). This is designed to automatically produce reliable training labels based on the observation that water surfaces generally display a lower elevation variance but exhibit heightened intensity fluctuations in LiDAR datasets.

Despite advancements, the complexity of large-scale LiDAR data combined with machine learning techniques highlights that a universally reliable and scalable solution for surface water mapping has yet to be achieved. While SLIER (Yan 2023) can autonomously generate training labels, its line-by-line processing amplifies computational demands and redundancy, particularly as ALS often results in overlapping and convoluted scan lines due to repetitive scans and diverse swaths. Such complexities further complicate the quest for consistent and trustworthy performance in large-area mapping projects. Beyond SLIER's unique challenges, ensuring highly reliable performance, especially in machine learning-driven methods, remains a persistent challenge in remote sensing, regardless of the use of LiDAR or other technologies. The issue mirrors challenges in mapping with optical and radar imagery due to data distribution shifts (Tuia, Persello, and Bruzzone 2016; Wieland et al. 2023).

2.3. Scalable surface water mapping: challenges and opportunities

A primary challenge in scalable surface water mapping, for both optical and LiDAR methods, is the variable reflectance of water in remote sensing data. This variability is heightened in high-resolution mappings which often target smaller water bodies with varied spectral properties, making them more prone to signal noise and environmental factors. Moreover, large-area mapping often requires multiple image captures due to the limited swath width, especially in high-resolution imaging. These varied captures introduce more variability in water reflectance due to differing atmospheric conditions or sensor types, complicating water mapping scalability (Ogilvie et al. 2018).Given the variations in remote sensing data features, coupled with the multitude of water body types, existing methods necessitate repetitive training dataset creation and validation. This not only complicates large-scale applications but requires rigorous human-validation for machine learning-based outputs (Bijeesh and Narasimhamurthy 2020).

In the meantime, the topographic map is a composite of terrain and water, each shaping the other. Capturing this dynamic interaction requires both terrain and water maps, along with their respective elevations, underscoring the pivotal role of elevation in augmenting the utility of surface water maps (Musa, Popescu, and Mynett 2015). ALS is, therefore, indispensable not only for creating high-resolution terrain maps but also for embedding elevation information into surface water maps. Its potential increases with the expanding collection of national-scale airborne LiDAR data (Moudr et al. 2023; Stoker and Miller 2022). Yet, while airborne LiDAR has been predominantly used for terrain mapping, generating a holistic 3D topographic map – encompassing a surface water map – has often necessitated separate optical images (Moore et al. 2019). This is primarily due to the absence of an ALS-based fully automated and scalable water mapping algorithm.

Recent advancements in LiDAR technology, such as the development of topo-bathymetric LiDAR (Mandlburger et al. 2020) and systems integrating LiDAR with optical sensors (Janowski et al. 2024), show promise for simultaneous surface water and topographic mapping. Nevertheless, the differences in ideal laser bandwidth for bathymetric and topographic data collection limit their effectiveness when compared to systems dedicated solely to one type of mapping (Awadallah, Juárez, and Alfredsen 2022; Kastdalen et al. 2024). Additionally, the higher costs associated with these advanced systems cannot be overlooked. In this context, the advantages of developing scalable surface water mapping techniques using traditional topographic laser scanning are evident. Such developments would not only enhance the utility of current systems but also take advantage of the vast amount of data already collected.

In this paper, we present an unsupervised, scalable water mapping approach that only requires the 3D coordinates from topographic ALS. This method capitalizes on an intrinsic physical property of water: the surface of connected water bodies is flat. By introducing a novel water elevation-based region merging technique, we offer scalable mapping without the necessity for training procedures or repetitive parameter tuning. Since it doesn't utilize intensity data, it eliminates the need for sensor- and scan-specific calibrations and mitigates uncertainties linked to surface reflectance. Thus, it paves the way for fine-scale, comprehensive surface water mapping and facilitates the creation of a full 3D topographic map using only airborne LiDAR data.

3. Proposed water mapping method

3.1. Overall strategy

A typical LiDAR sensor emits near-infrared pulses, which are often absorbed or largely undergo specular reflection at higher incidence angles when interacting with water. This leads to significant laser point dropouts over water bodies in airborne LiDAR data (Höfle et al. 2009). Thus, local LiDAR point density can potentially differentiate water from non-water areas. However, challenges arise due to obstructions and varied incidence angles affecting point density, making direct point density use for water body mapping complex. While intensity information was previously employed to counteract this, inconsistencies in sensing conditions limit its scalability for water mapping.

Our approach uniquely hinges on the point cloud's 3D coordinates, eschewing intensity, and incorporates the robust physical constraint that connected surface water remains flat. Figure 1 offers a visual representation of standard topographic ALS.

The marked red dots are the points detected by the receiver. Only the 3D coordinates of these dots are considered. As shown, sparser points might indicate water bodies or areas obscured by structures. Our method first labels low-density areas as potential water regions and then filters out false positives

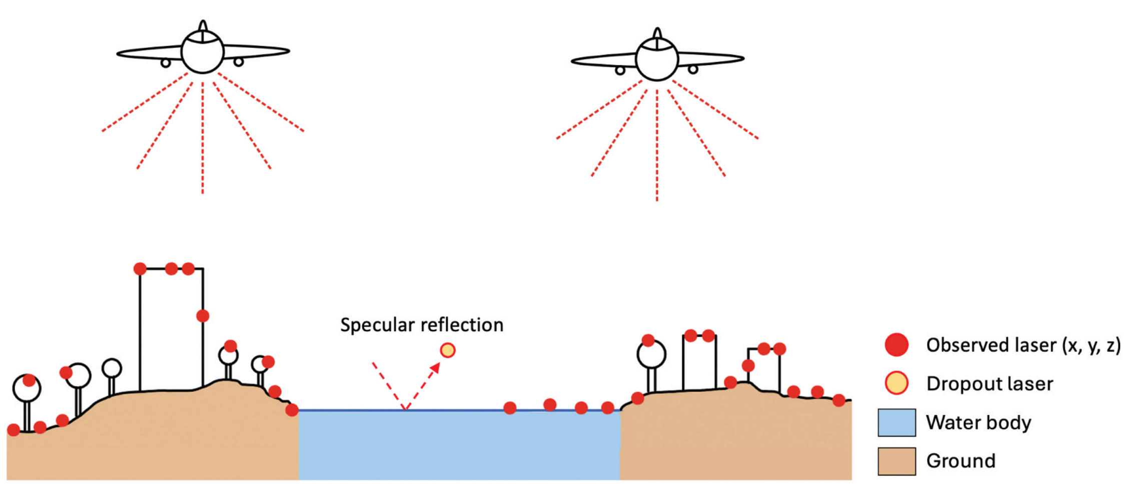


Figure 1. Profile view of an airborne laser scanning with red dots indicating detected points.

arising from tall structures. Also, due to lower incidence angles or surface water conditions, some portions of a water body can have high point density, resulting in incomplete, initial water segments. To address this, we expand the initial water segment by merging it with others of similar elevations that connect to the original segment. In sum, our approach consists of (1) initial water segment extraction and (2) water elevation-based region merging (WERM).

3.2. Initial water segments extraction

Our approach extracts initial water segments based on lower point density over water bodies, due to water's specular reflection and absorption of infrared light. Specular reflection often directs laser light away from the sensor, particularly at oblique angles, while absorption reduces return signals. For an illustration of how acquired airborne LiDAR data points vary in density over areas with water bodies, and to see examples of initial water segments extracted using this data, please refer to "1. Initial water segments extraction"of Figure 2.

First, raw airborne LiDAR data is rasterized into a high-resolution 2D gridded space, G, forming a digital surface model (DSM). This converts LiDAR point clouds into an $m \times n$ grid, where m and n represent pixel counts in rows and columns, respectively. Pixels without any laser point registration are termed as E. Areas with higher non-registered pixel concentrations are likely water bodies. The method determines the average cell density (P) of a DSM using:

$$P = \frac{\text{Total cell sin}\,G - \text{Total cells of}\,E}{\text{Total cells in}G} \quad (1)$$

Then, a sliding window W, sized $N \times N$ pixels, traverses over G. For each position of W, the count of non-empty cells within it follows a $B(N^2, P)$ binomial distribution. We adjust this to $B(N^2, P')$ where $P' = P/2$ to account for commission errors arising from point density imbalances due to scanning overlaps. A center pixel in the window is labeled as water if its point count falls below the confidence interval's lower bound:

$$W_{\text{center}} = \begin{cases} \text{Water} & \text{if the number of non} - \text{empty cells in}\, W < Z \times \sqrt{N^2 \times P' \times (1 - P')} \\ \text{Non} - \text{Water} & \text{otherwise} \end{cases} \quad (2)$$

Finally, connected pixels labeled as water become initial water segments S. By default, the point clouds are rasterized into a 0.5m grid, employing a 9×9 window and a default confidence level of 2 (the critical z-score, Z). This setup aligns with typical optical imagery used for ground truthing and is sufficiently detailed for city-level and national mapping applications, especially considering the dynamic nature of water bodies. The choice of resolution can vary based on project needs and resource availability. However, caution is necessary to maintain the inherent, relative point density gap between water and non-water bodies without excessively distorting the original distribution. For example, a 1m resolution with 100 points$/m^2$ of point cloud data could significantly increase the likelihood of capturing laser points per cell over water bodies. If higher resolution is not needed, random downsampling of the original point clouds would be beneficial.

In densely urbanized areas with numerous high-rises, an optional "building buffer"operation can be implemented. This strategy aims to prevent false-positive water segments that arise from occlusions created by tall buildings. For this, we utilize a 3D building mapping algorithm (Song and Jung 2022).

3.3. Water elevation-based region merging

Initial water segments derived previously are generally incomplete due to varying point densities within water regions. Thus, our method expands these segments, assuming that connected water bodies share similar elevations. For a visual representation of how WERM operates and its impact on improving the initial water segments, please see "2. Water Elevation-based Region Merging (WERM)"in Figure 2. This figure displays the transformation from initial water segments to final water bodies, illustrating the principle of the elevation-based merging strategy in capturing the full extent of water surfaces.

For every initial segment S_i, regions of the similar elevation are extracted by "slicing"the DSM perpendicular to the elevation axis. Specifically, for each S_i, we determine its elevation E_i by computing the 10^{th} percentile of the corresponding DSM's elevation E_{D_i}:

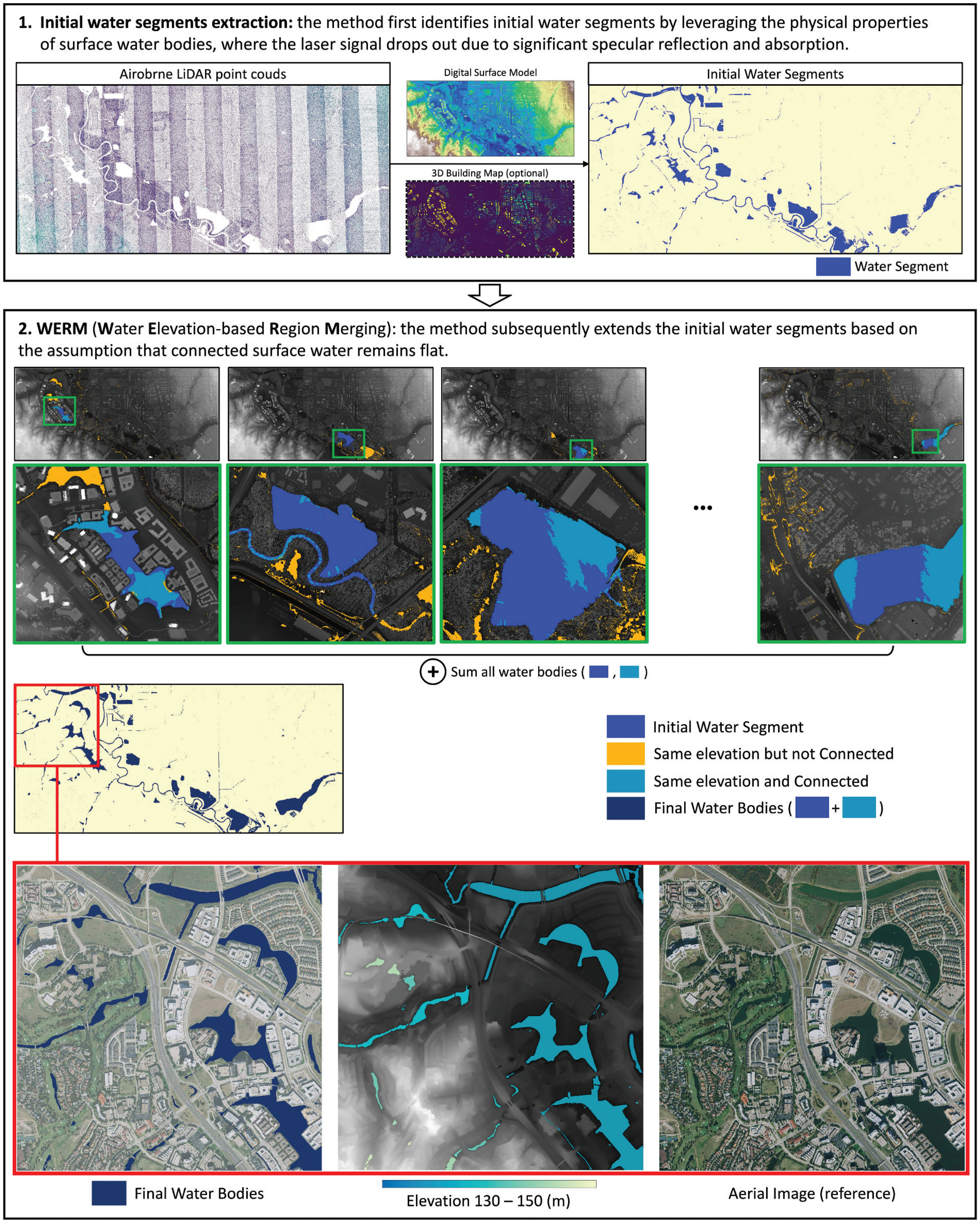


Figure 2. Overview of the method: (1) initial water segments extraction, utilizing the specular reflection and absorption properties of surface water bodies. (2) water elevation-based region merging, leveraging the principle that connected surface water remains flat.

$$E_i = \text{percentile}(E_{D_i}, 10) \tag{3}$$

Next, for each S_i, we identify regions of nearly identical elevation by slicing the DSM within a $\pm\, 0.1m$ interval around E_i, reflecting the typical vertical accuracy of airborne LiDAR data.

$$R_i = r \in D_i : E_i - 0.1 \leq r \leq E_i + 0.1 \tag{4}$$

Then, a connected component labeling algorithm identifies subsets C_j in the binary image formed by S_i and R_i. Intersecting components with S_i are merged, creating refined water segments $S_{i'}$:

$$W_{center} = \begin{cases} S_i \cup C_j & \text{if } S_i \cap C_j \neq \emptyset \\ S_i & \text{otherwise} \end{cases} \tag{5}$$

In such manner, WERM expands these segments into nearby regions of similar elevation. For enhanced precision in elevation and boundary, the process iterates twice by default on each $S_{i'}$ to derive $S_{i''}$:

$$S_i'' = \text{iterate}(S_i', E_i, R_i) \tag{6}$$

The final surface water map includes both the initial and merged segments. Since these water body boundaries are delineated based on the elevations captured by LiDAR, they demonstrate high precision. WERM is applied to initial segments of water bodies larger than $500m^2$, deliberately excluding minor elements like puddles to prevent unnecessary extension of their boundaries. It's important to note that our algorithm does extract water bodies smaller than $500m^2$, but does not apply WERM to these smaller entities to avoid their undue expansion. The parameters for WERM, such as the 10^{th} percentile, a range of $\pm 0.1m$, and the $500m^2$ threshold, are empirical defaults proven to be effective across diverse landscapes with different airborne LiDAR datasets. A detailed discussion on parameters is provided in Section 5.

Datasets and experimental design

We aimed to develop a scalable water mapping algorithm, so our evaluation utilized data spanning diverse and extensive landscapes. We assessed five datasets (Figure 3) that cover three metropolitan regions, featuring diverse types of rivers and lakes, a mountainous territory, and a coastal urban area. These datasets encompass seas, lakes, rivers, streams, ponds, wetlands, ditches, and snows, totaling an area of approximately $2{,}500km^2$ with 10 billion pixels at 0.5 m resolution.

For quantitative assessment, we utilized the water body layer from the U.S. Geological Survey (USGS) High Resolution National Hydrography Dataset Plus

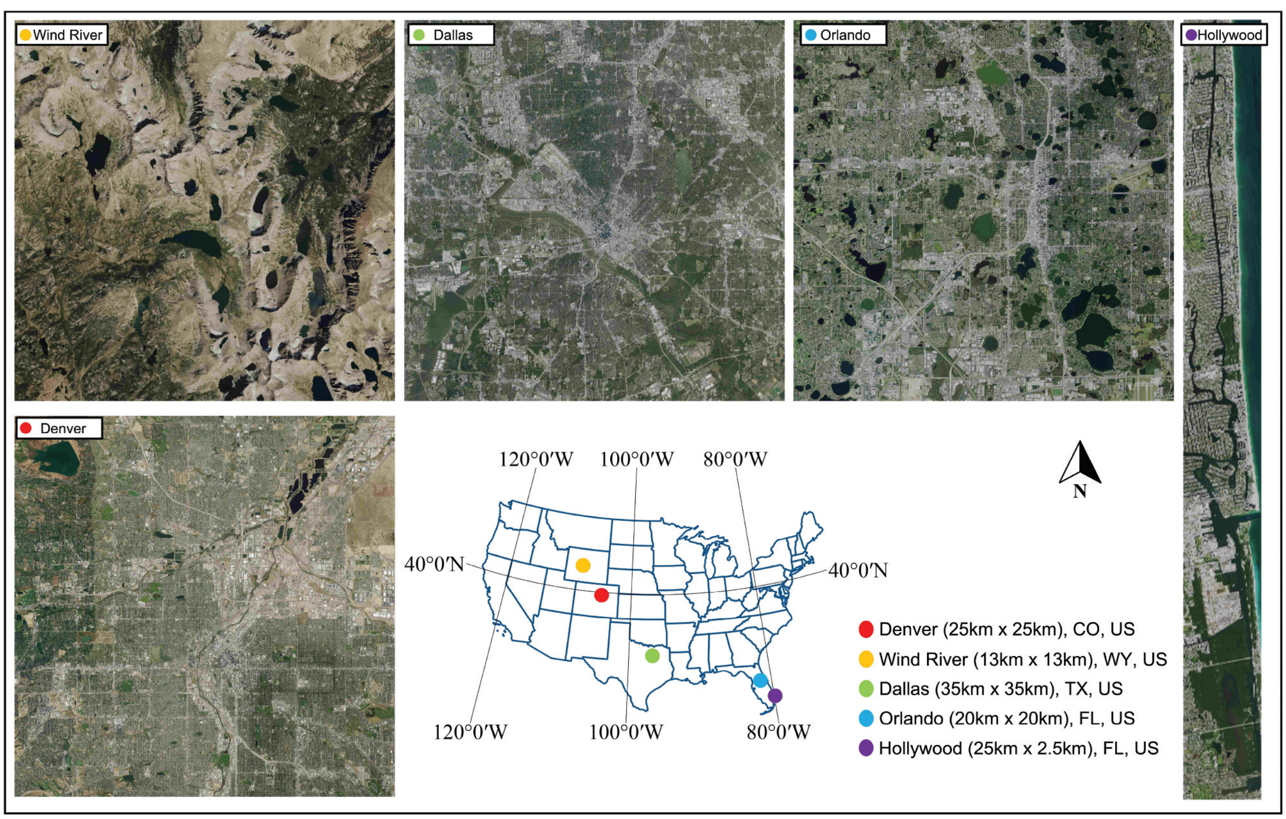


Figure 3. Datasets used for evaluation: three metropolitans (Denver, Dallas, Orlando), a mountainous area (Wind River), and a coastal zone (Hollywood), totaling approximately $2{,}500km^2$.

(NHDPlus HR) (Moore et al. 2019). Referencing NHDPlus HR, we calculated Intersection over Union (IoU), precision, recall, F1-score, and overall accuracy (OA) for the quantitative evaluation:

- Intersection over Union (IoU): Ratio of the intersection to the union of predicted and actual water bodies.
- Precision: Proportion of true positive predictions among all positive predictions.
- Recall: Proportion of actual water pixels correctly predicted among all actual water pixels.
- F1-score: Harmonic mean of precision and recall, balancing both metrics.
- Overall Accuracy (OA): Proportion of correctly identified pixels (both water and non-water) out of the total number of pixels.

To facilitate robust validation of our method's performance on a large scale, we benchmarked it against two highly competitive versions of established normalized difference water index (NDWI)-based methods. The first ("NDWI-G-Opt") determined an optimal threshold value for the highest overall accuracy (OA) relative to the NHDPlus HR reference per experimental dataset. The second ("NDWI-L-Opt") segmented datasets into tiles and optimized the threshold for each tile to achieve the highest performance relative to the reference. Hence, while the first method constructs a water map using a universally optimal threshold, the second stitches locally optimized tiles. Though these NDWI methods aren't entirely practical – as groundtruths are supposed to be unknown – we employed them to benchmark our approach stringently. Their thresholds were fine-tuned between 0 and 0.9 at 0.05 increments, mirroring (Chen et al. 2020a). In contrast, our technique consistently used one default parameter set across all datasets, proving its innate scalability and robustness without site-specific adjustments.

Table 1 provides a detailed summary of the datasets utilized in our assessment, which considered the diversity of both geographic and LiDAR data acquisition configurations. All LiDAR data were sourced from the USGS's 3D Elevation Program (3DEP) (Stoker and Miller 2022). For NDWI-based methods, we used the <1 meter resolution, near-infrared and green bands from the U.S. Department of Agriculture's National Agriculture Imagery Program (NAIP).

Importantly, the vastness of our dataset means that average results might not capture all performance nuances. For instance, statistics can be skewed by a few large water bodies. Moreover, the NHDPlus HR categorizes certain features like snow and wetlands as water bodies; these are temporally variable and are not the primary focus of either our proposed or NDWI-based methods. To gain better insights, we implemented a tile-based evaluation alongside the standard quantitative analysis. This involved dividing the entire experimental area into small tiles (optimized by the "NDWI-L-opt"method). We then excluded tiles that contained a large amount of snow, wetlands, or other types of label errors to refine our sample tiles for evaluation, termed "Reliable Tiles'.' Additionally, we identified "Challenging Tiles"based on the pronounced discrepancies between surface water maps. By doing so, we offer results from diverse subsets of the entire tile collection, rather than presenting a singular outcome.

Moreover, we conducted additional comparative analyses against supervised methods. This approach allows us to contrast our method with highly competitive alternatives, which, although requiring training labels, are utilized due to the limited availability of public benchmark datasets and algorithms for unsupervised LiDAR-based methods. Consequently, we employed both machine learning-based classifications and deep learning-based semantic segmentation methods, using two distinct types of input

Table 1. Summary of the experimental datasets.

Name	Denver	Wind River	Dallas	Orlando	Hollywood
Location	Denver, CO	Wind River, WY	Dallas, TX	Orlando, FL	Hollywood, FL
Dimension	25-km by 25-km	13-km by 13-km	35-km by 35-km	20-km by 20-km	25-km by 2.5-km
Geography	Metropolitan area	Mountainous area	Metropolitan area	Metropolitan area	Coastal area
Scanning details	Leica TerrianMapper 2.4 points/m^2 AGL: 2.7–3.1km scan angle: 40° overlap: 20%	Leica TerrianMapper 2.0 points/m^2 AGL: 3.0–3.2km scan angle: 40° overlap: 25%	Leica ALS 80 3.0 points/m^2 AGL: 1.8km scan angle: 35° overlap: 30%	Leica ALS 80 9.8 points/m^2 AGL: 1.4km scan angle: 40° overlap: 60%	Riegl VQ-1560i 8.2 points/m^2 AGL: 1.3km scan angle: 60° overlap: 30%
LiDAR acquisition	2020.05.–2020.06.	2019.08.–2019.09.	2019.03.–2019.07.	2018.12.–2019.03.	2018.06.
NAIP acquisition	2019.08.–2019.09.	2019.08.	2020.10.–2020.11.	2019.11.	2019.11.

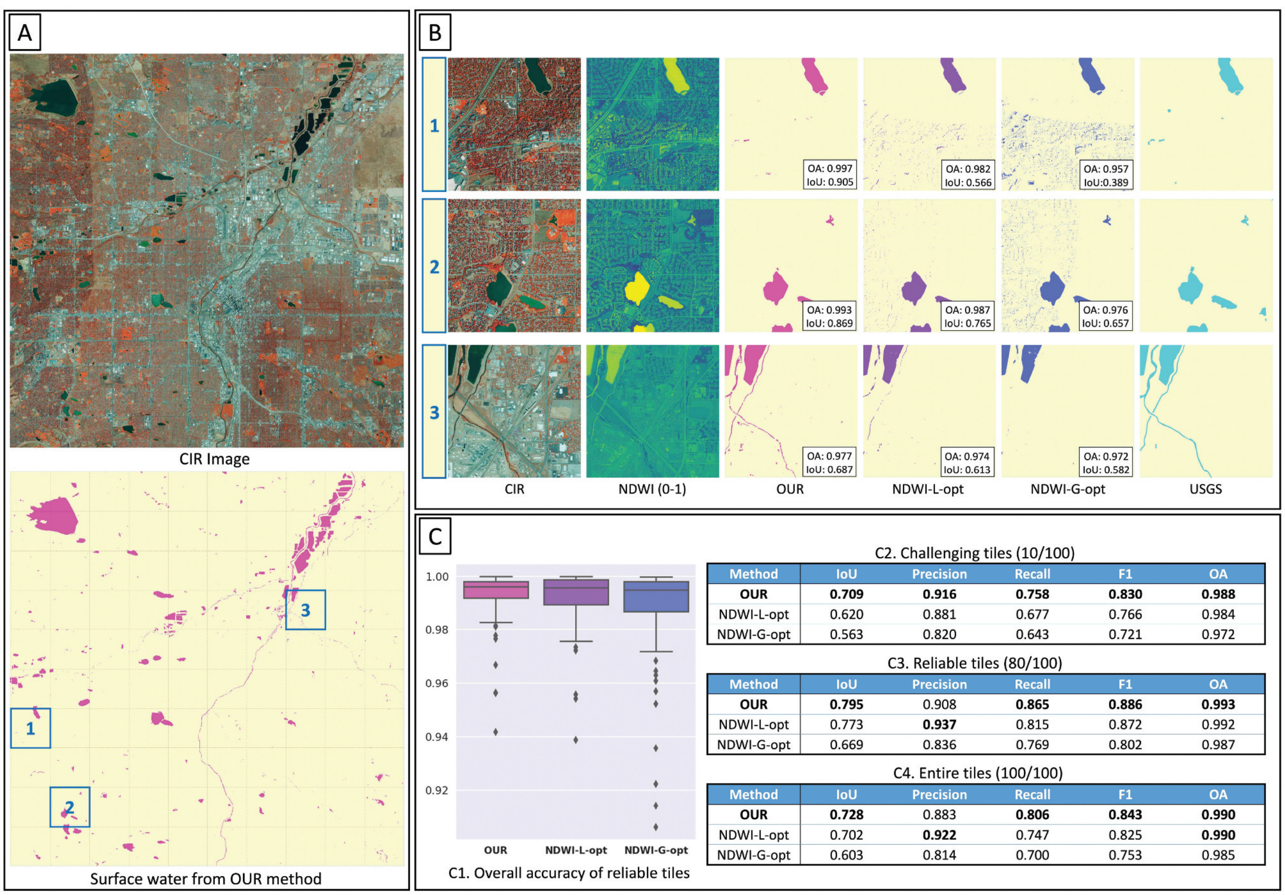
datasets: optical imagery and airborne LiDAR data. Specifically, the optical imagery input comprised stacked RGB-NIR imagery from NAIP. For the LiDAR input, we constructed four feature maps from the 3DEP airborne LiDAR data, including the DSM, intensity, standard deviation of elevation, and point density. We normalized DSM and intensity data after clipping the top and bottom 2% of values and calculated the standard deviation of elevation and point density using a 9-m by 9-m kernel. These features were selected for their proven effectiveness in supervised water mapping (Crasto et al. 2015; Höfle et al. 2009; Malinowski et al. 2016).

5. Experimental results

5.1. Detailed performance analysis across datasets

Throughout Figures 4–8, subfigure A shows the entire extent of each dataset, presenting the color-infrared (CIR) image alongside the surface water map generated by the proposed method, labeled as "OUR."Subfigure B highlights three subset regions with significant differences among surface water maps, along with their corresponding OA and IoU values against the reference. The "USGS"label represents the USGS NHDPlus HR surface water map, used as our benchmark for quantitative metrics. Subfigure C provides statistical insights: Boxplot C1 shows the distribution of OAs, while Tables C2 and C3 present results from different subsets of the tile collection. C2 shows results from the top 10 tiles (top 3 for Hollywood), referred to as Challenging Tiles, with the most pronounced IoU discrepancies. C3 presents outcomes from Reliable Tiles, excluding those with significant errors, and C4 shows results for all tiles without exclusion.

In the Denver dataset (Figure 4), the proposed method consistently achieved higher accuracy compared to NDWI-based methods. Notably, our



C2. Challenging tiles (10/100)

Method	IoU	Precision	Recall	F1	OA
OUR	0.709	0.916	0.758	0.830	0.988
NDWI-L-opt	0.620	0.881	0.677	0.766	0.984
NDWI-G-opt	0.563	0.820	0.643	0.721	0.972

C3. Reliable tiles (80/100)

Method	IoU	Precision	Recall	F1	OA
OUR	0.795	0.908	0.865	0.886	0.993
NDWI-L-opt	0.773	0.937	0.815	0.872	0.992
NDWI-G-opt	0.669	0.836	0.769	0.802	0.987

C4. Entire tiles (100/100)

Method	IoU	Precision	Recall	F1	OA
OUR	0.728	0.883	0.806	0.843	0.990
NDWI-L-opt	0.702	0.922	0.747	0.825	0.990
NDWI-G-opt	0.603	0.814	0.700	0.753	0.985

Figure 4. Performance analysis on the Denver dataset. (a) Proposed method's water map vs. CIR image; (b) detailed views comparing CIR, NDWI, and three water maps against USGS NHDPlus HR reference; (c) statistical comparisons of three methods presented in box plots and tables, each reflecting different subsets of the entire tile.

approach excelled at detecting intricate and small water bodies, such as stream branches, with greater precision than the alternatives. As shown in the first row of Figure 4b, the proposed method exhibited fewer false positives in urban areas compared to the NDWI-based methods. NDWI-L-opt, even with its local optimization, produced many false positives, while NDWI-G-opt identified more water bodies but at the cost of an increased false positive rate, resulting in reduced OA and IoU. As depicted in subsequent rows of Figure 4b, our method consistently surpassed the NDWI methods in OA and IoU metrics and closely emulated the USGS reference, particularly in delineating intricate stream branches. Additionally, as shown in the Boxplot of Figure 4c, our method exhibits the smallest deviation in OA and higher accuracy compared to other methods, underscoring its robustness and scalability. Tables C2-C4 further reinforce this robustness, with consistently high accuracy across various tile subsets.

The Wind River dataset results (Figure 5) demonstrate the effectiveness of our method in mountainous terrain, characterized by steep topography and extensive shadows. Our method adeptly differentiates water bodies from terrain shadows, whereas NDWI-based methods often incur significant errors. In particular, these methods tend to misinterpret shadows as water, introducing noise in regions influenced by rugged terrain and vegetation. In contrast, the proposed method closely aligns with the reference data overall. Boxplot C1 (Figure 5c) highlights our method's superior OA with minimal variance compared to NDWI-based methods, demonstrating its robustness. This is particularly notable as no parameter adjustments were made for the entire mapping coverage. Across metrics C2–C4, our method consistently shows high accuracy.

The Dallas dataset results (Figure 6) demonstrate the robustness of our method across various water bodies, including a large river, branching creeks,

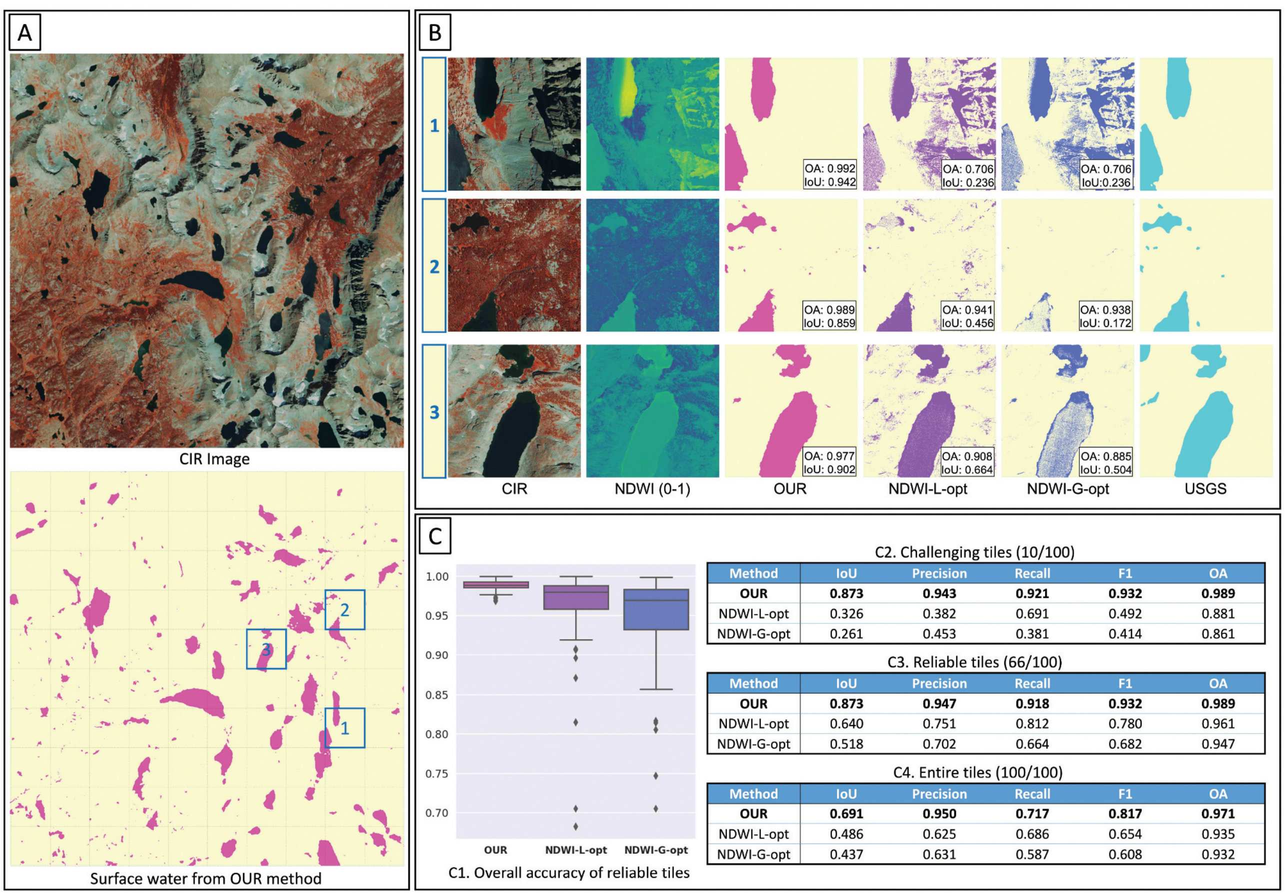
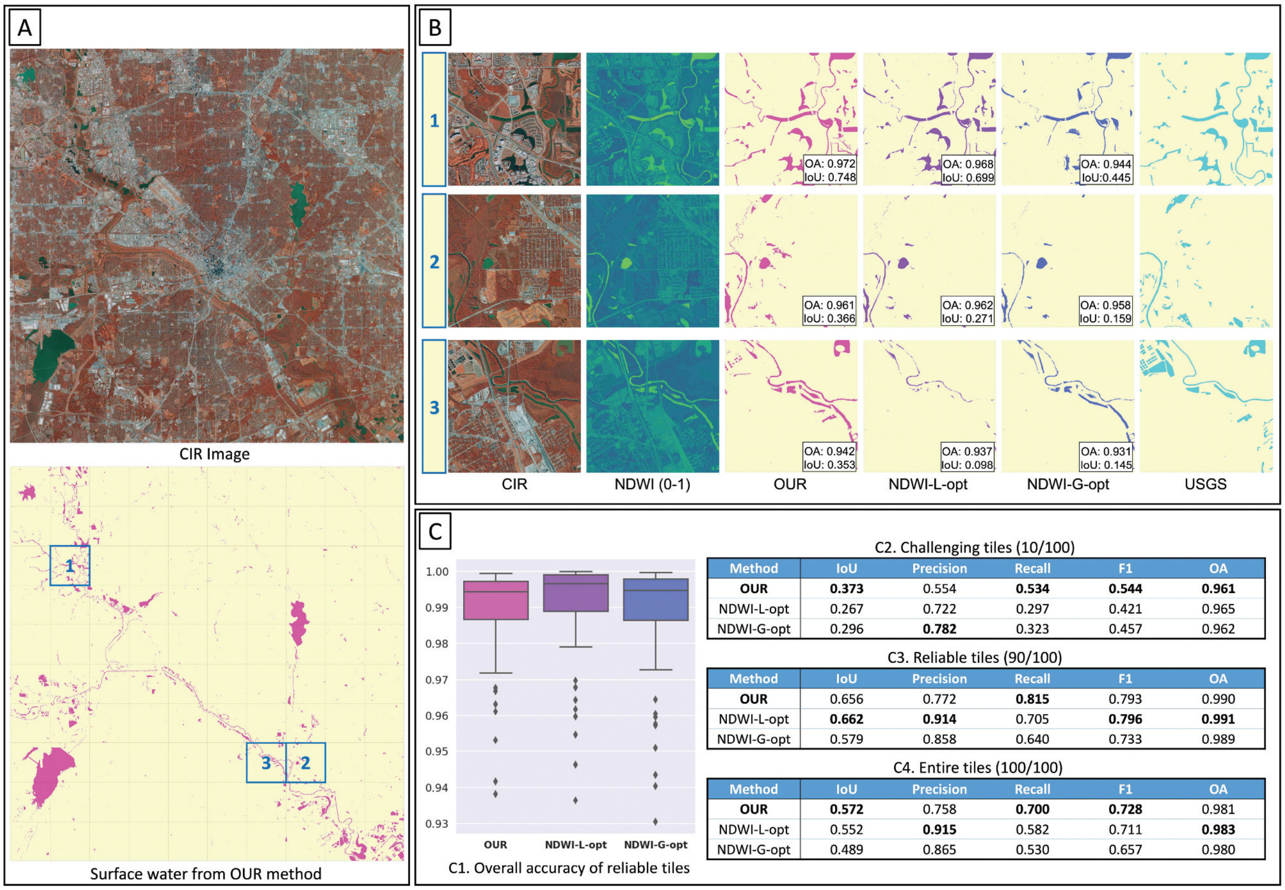


Figure 5. Performance analysis on the wind River dataset. (a) Proposed method's water map vs. CIR image; (b) detailed views comparing CIR, NDWI, and three water maps against USGS NHDPlus HR reference; (c) statistical comparisons of three methods presented in box plots and tables, each reflecting different subsets of the entire tile.



C2. Challenging tiles (10/100)

Method	IoU	Precision	Recall	F1	OA
OUR	**0.373**	0.554	**0.534**	**0.544**	**0.961**
NDWI-L-opt	0.267	0.722	0.297	0.421	0.965
NDWI-G-opt	0.296	**0.782**	0.323	0.457	0.962

C3. Reliable tiles (90/100)

Method	IoU	Precision	Recall	F1	OA
OUR	0.656	0.772	**0.815**	0.793	0.990
NDWI-L-opt	**0.662**	**0.914**	0.705	**0.796**	**0.991**
NDWI-G-opt	0.579	0.858	0.640	0.733	0.989

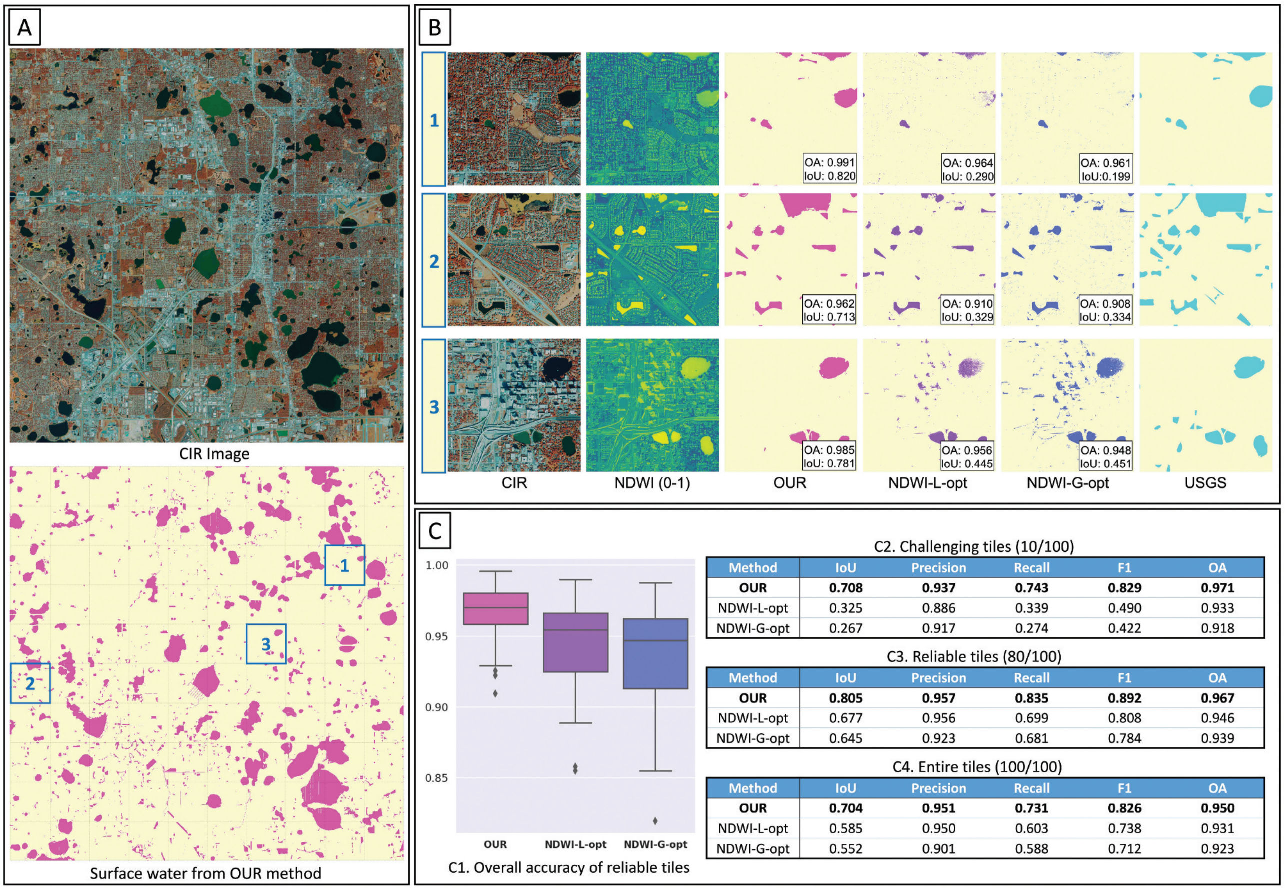
C4. Entire tiles (100/100)

Method	IoU	Precision	Recall	F1	OA
OUR	**0.572**	0.758	**0.700**	**0.728**	0.981
NDWI-L-opt	0.552	**0.915**	0.582	0.711	**0.983**
NDWI-G-opt	0.489	0.865	0.530	0.657	0.980

Figure 6. Performance analysis on the Dallas dataset. (a) Proposed method's water map vs. CIR image; (b) detailed views comparing CIR, NDWI, and three water maps against USGS NHDPlus HR reference; (c) statistical comparisons of three methods presented in box plots and tables, each reflecting different subsets of the entire tile.

lakes, and scattered wetlands, streams, and ponds. As shown in Figure 6b, which depicts reservoirs, streams, and lakes, our method excels in terms of OA and IoU. Notably, it captures even minor creeks not outlined in the reference. In contrast, the NDWI-L-opt and NDWI-G-opt methods struggle to precisely map water bodies. While NDWI-L-opt achieves the highest OA in some cases, our method effectively identifies minor creeks when juxtaposed with the CIR image, highlighting areas where NDWI-based methods falter. This underscores the fact that quantitative metrics can sometimes fail to reflect the true efficacy of a method due to imperfections in the reference. In Figure 6c, our method's overall performance aligns with NDWI-L-opt in terms of IoU, F1-score, and OA; however, the higher recall and lower precision indicate that our method detects more water. This is further supported by the observation that our method identifies smaller water bodies not delineated in the reference map, as evident in Figure 6b. The Dallas dataset predominantly consists of small water bodies, many of which are inaccurately represented. Consequently, the quantitative results may deviate more than those from other datasets due to misalignment with the actual performance of the different methods.

Our method accurately mapped diverse lakes and ponds across the flat metropolitan terrain, as demonstrated in the Orlando dataset results (Figure 7a). The abundance and variability of water bodies in this dataset made surface water mapping particularly challenging for elevation-based approaches. Figure 7b provides detailed insights from three selected regions. The first row displays urban lakes and ponds, where our method shows superior accuracy in both OA and IoU. In contrast, the two NDWI-based methods omit numerous water bodies due to some non-aquatic areas having higher NDWI values than actual water bodies. While adjusting the



C2. Challenging tiles (10/100)

Method	IoU	Precision	Recall	F1	OA
OUR	**0.708**	**0.937**	**0.743**	**0.829**	**0.971**
NDWI-L-opt	0.325	0.886	0.339	0.490	0.933
NDWI-G-opt	0.267	0.917	0.274	0.422	0.918

C3. Reliable tiles (80/100)

Method	IoU	Precision	Recall	F1	OA
OUR	**0.805**	**0.957**	**0.835**	**0.892**	**0.967**
NDWI-L-opt	0.677	0.956	0.699	0.808	0.946
NDWI-G-opt	0.645	0.923	0.681	0.784	0.939

C4. Entire tiles (100/100)

Method	IoU	Precision	Recall	F1	OA
OUR	**0.704**	**0.951**	**0.731**	**0.826**	**0.950**
NDWI-L-opt	0.585	0.950	0.603	0.738	0.931
NDWI-G-opt	0.552	0.901	0.588	0.712	0.923

Figure 7. Performance analysis on the Orlando dataset. (a) Proposed method's water map vs. CIR image; (b) detailed views comparing CIR, NDWI, and three water maps against USGS NHDPlus HR reference; (c) statistical comparisons of three methods presented in box plots and tables, each reflecting different subsets of the entire tile.

threshold might include more water areas, it would also increase the likelihood of false positives. NDWI-L-opt detected more water than NDWI-G-opt but introduced additional noise. The second row highlights our method's ability to discern algae-covered water bodies, a scenario where NDWI methods struggle. Our technique effectively detects these due to the similar elevations of algae-covered and clear lakes. The third row presents areas where high-rise buildings and lakes are juxtaposed. NDWI imagery shows that shadows from buildings often have higher NDWI values than real water areas, leading to inaccuracies in NDWI-based techniques. Statistical results in Figure 7c further demonstrate the efficacy and reliability of the proposed method across diverse scenarios.

The proposed method effectively extracted water bodies in a coastal urban area adjacent to the ocean, as shown in the Hollywood dataset (Figure 8a), which features various creeks, bays, and artificial waterways. Figure 8b displays three of the ten tiles from the Hollywood dataset. Overall, both the proposed method and the NDWI-based method produced satisfactory results in water body detection. In every case, the OA and IoU values exceeded 0.9 and 0.8, respectively. Compared to the USGS's water body layer, the most common errors arose from occlusions and shadows cast by tall coastal buildings. Other discrepancies came from minor inland water bodies or false negatives caused by watercraft docked along the waterways. Among all the methods tested, the proposed method achieved the highest accuracy, particularly in preventing errors from shadows or occlusions. Figure 8c furnishes quantitative results. The three omitted tiles in C3 correspond to the bottom three tiles in Figure 8a, as NHDPlus HR classified mangroves as water bodies. All methods, except for the three exclusions, achieved metrics above 0.85, with our method demonstrating the highest accuracy.

Table 2 summarizes the performance of different methods across five datasets. The mean values of IoU,

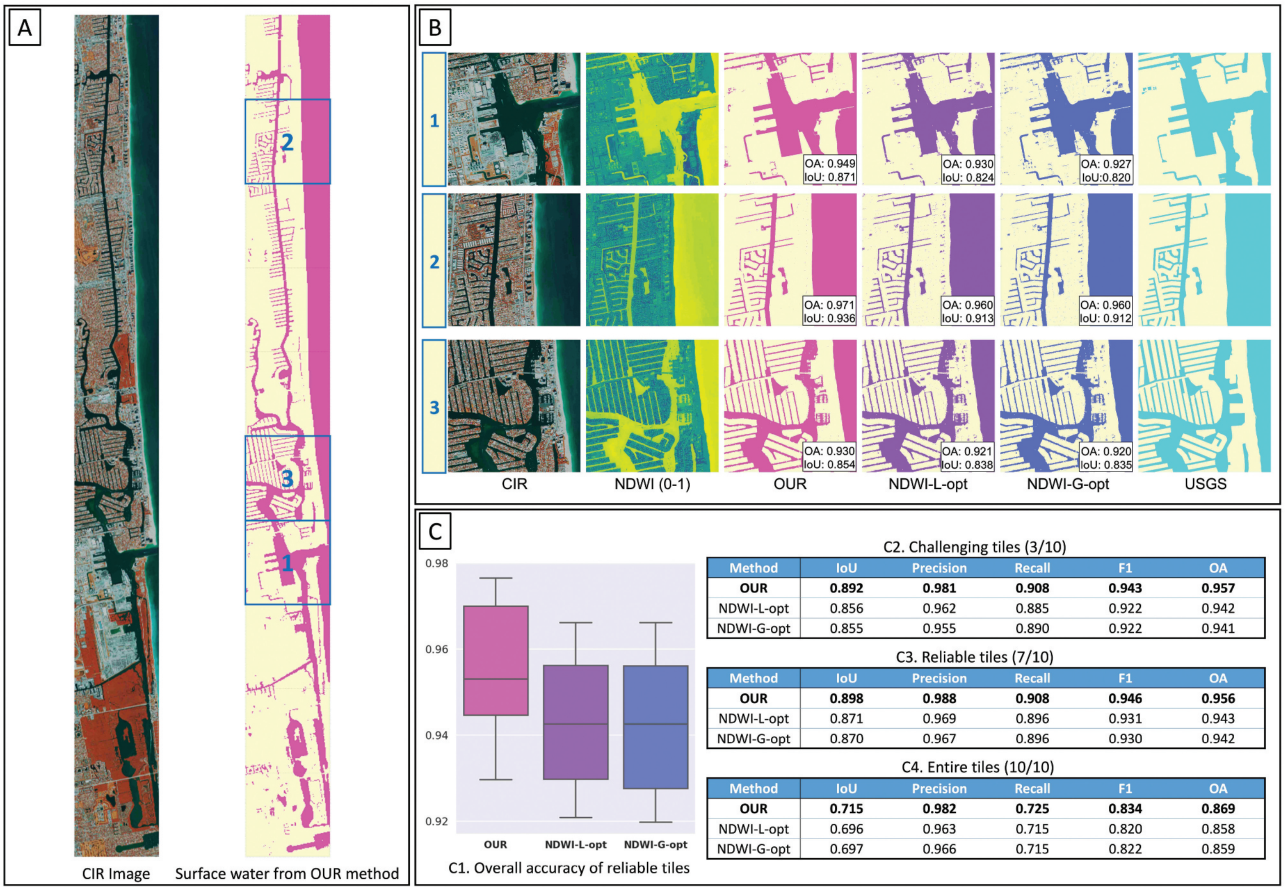


Figure 8. Performance analysis on the Hollywood dataset. (a) Proposed method's water map vs. CIR image; (b) detailed views comparing CIR, NDWI, and three water maps against USGS NHDPlus HR reference; (c) statistical comparisons of three methods presented in box plots and tables, each reflecting different subsets of the entire tile.

Table 2. Summary of the proposed method's performance against ndwi-based methods across five datasets: this table presents mean intersection over union (IoU), precision, recall, F1-score, and overall accuracy (OA) evaluated using reliable tiles. The "NDWI-G-opt"method selects the global optimal threshold for each test dataset, while "NDWI-L-opt"chooses the local optimal threshold for each tile based on the reference label.

Method	Metric	Denver	Wind River	Dallas	Orlando	Hollywood	Average
NDWI-G-opt	IoU	0.669	0.518	0.579	0.645	0.870	0.656
	Precision	0.836	0.702	0.858	0.923	0.967	0.857
	Recall	0.769	0.664	0.640	0.681	0.896	0.730
	F1-score	0.802	0.682	0.733	0.784	0.930	0.786
	OA	0.987	0.947	0.989	0.939	0.942	0.961
NDWI-L-opt	IoU	0.773	0.640	0.662	0.677	0.871	0.725
	Precision	0.937	0.751	0.914	0.956	0.969	0.905
	Recall	0.815	0.812	0.705	0.699	0.896	0.785
	F1-score	0.872	0.780	0.796	0.808	0.931	0.837
	OA	0.992	0.961	0.991	0.946	0.943	0.967
Proposed Method	IoU	0.795	0.873	0.656	0.805	0.898	0.805
	Precision	0.908	0.947	0.772	0.957	0.988	0.914
	Recall	0.865	0.918	0.815	0.835	0.908	0.868
	F1-score	0.886	0.932	0.793	0.892	0.946	0.890
	OA	0.993	0.989	0.990	0.967	0.956	0.979

precision, recall, F1-score, and OA, assessed using Reliable Tiles, are tabulated for each method, along with their average over the five datasets. Our method showed the highest accuracy in all metrics, significantly outperforming the NDWI-based methods. Notably, our method achieved a 22.7% higher mean

IoU compared to NDWI-G-opt and an 11.0% improvement over NDWI-L-opt on average. Additionally, it achieved a 13.2% higher mean F1-score compared to NDWI-G-opt and a 6.3% improvement over NDWI-L-opt on average, underscoring its superior efficacy in scalable and accurate surface water mapping.

5.2. Comparative analysis against supervised methods

For machine learning methods, we employed pixel-based classification using random forest (RF) and support vector machine (SVM) classifiers. These methods were chosen for their wide adaptability and high performance in surface water mapping tasks (Bijeesh and Narasimhamurthy 2020; Ko, Kim, and Nam 2015; Sun et al. 2015). While we acknowledge that the ideal benchmark would be an automated unsupervised airborne LiDAR-based method, we chose machine learning-based methods due to the limited availability of such unsupervised methods in this field. It is important to note that machine learning methods using airborne LiDAR data are currently the state-of-the-art practice for surface water mapping, as existing methods typically employ machine learning classifiers after collecting training samples manually (Malinowski et al. 2016; Yu et al. 2022) or automatically (Yan 2023).

For the experiment, we created a subset of the Dallas dataset, covering an area of 14 km by 8 km, consisting of diverse surface water bodies (Figure 9). To imitate practical conditions, we created 122 training polygons using QGIS software, totaling over 100,000 pixels of water and non-water classes. From this set, we randomly selected 5,000 pixels for training. We then employed a five-fold cross-validation approach combined with parameter grid searching. For the SVM classifier, using the radial basis function kernel, we tested different values for the "C" parameter (0.1, 1, 10, 100) and the "Gamma" parameter (0.01, 0.001, 0.0001). For the RF classifier, we experimented with various settings for the number of estimators (50, 100, 200) and the maximum depth of the trees (0, 10, 20). We replicated five experiments, each with a different random sampling of training samples. The implementation was carried out using Python's scikit-learn library, version 1.3.2.

Our method generally obtained higher accuracy compared to pixel-based machine learning approaches for both optical and LiDAR datasets at their highest accuracies (Table 3). Specifically, our method obtained an IoU of 0.706 and an F1-score of 0.827, compared to the highest IoU of 0.696 and F1-score of 0.821 using RGB-NIR with SVM, and the highest IoU of 0.562 and F1-score of 0.719 using LiDAR-based features with RF. The accuracy trends with varying training sample sizes indicate that both machine learning methods have reached close to their best obtainable accuracies.

Figure 9 illustrates the subset area within the Dallas dataset with corresponding (a) RGB imagery, (b) Digital Elevation Model, (c) LiDAR point clouds projected into a gridded array, (d) LiDAR intensity map, and (e) training sample polygons for water and ground. Figures 9 (f-h) display the generated surface water maps with omission and commission errors for the proposed method, SVM with RGB-NIR, and RF with LiDAR features, respectively. Correctly identified surface water pixels are shown in blue, omission errors in orange, and commission errors in red.

Compared to our method, RF with LiDAR features showed lower accuracy, with more salt-and-pepper noise and commission errors. Our method's superior performance can be attributed to its ability to leverage the spatial context and physical properties of surface water, ensuring that connected water bodies remain flat. Advanced feature development (Malinowski et al. 2016), higher-quality training samples, and incorporating more spatial-contextual information beyond the pixel level – such as using patches (Song, Kim, and Kim 2019) or larger receptive fields, as seen in deep learning (Wieland et al. 2023) – would also improve performance. Nonetheless, the unique advantages offered by our proposed method – operating without a training procedure and ensuring clear elevation demarcation at the interface of water bodies and terrain – remain undiminished.

For deep learning, we used the U-Net model, renowned for its effectiveness in high-resolution surface water mapping (Wieland et al. 2023).While specialized and pre-trained models (Chen et al. 2020b; Isikdogan, Bovik, and Passalacqua 2019; Luo, Tong, and Hu 2021) could improve results, we opted for standard U-Net to establish a clear benchmark and minimize complexity. We used the Denver and Wind River datasets (Table 1) and employed three different training sampling methods: "Random Mix,""North/South,"and "Cross-state."Each method experienced different levels of data distribution shift. Random Mix involved completely randomly selected training

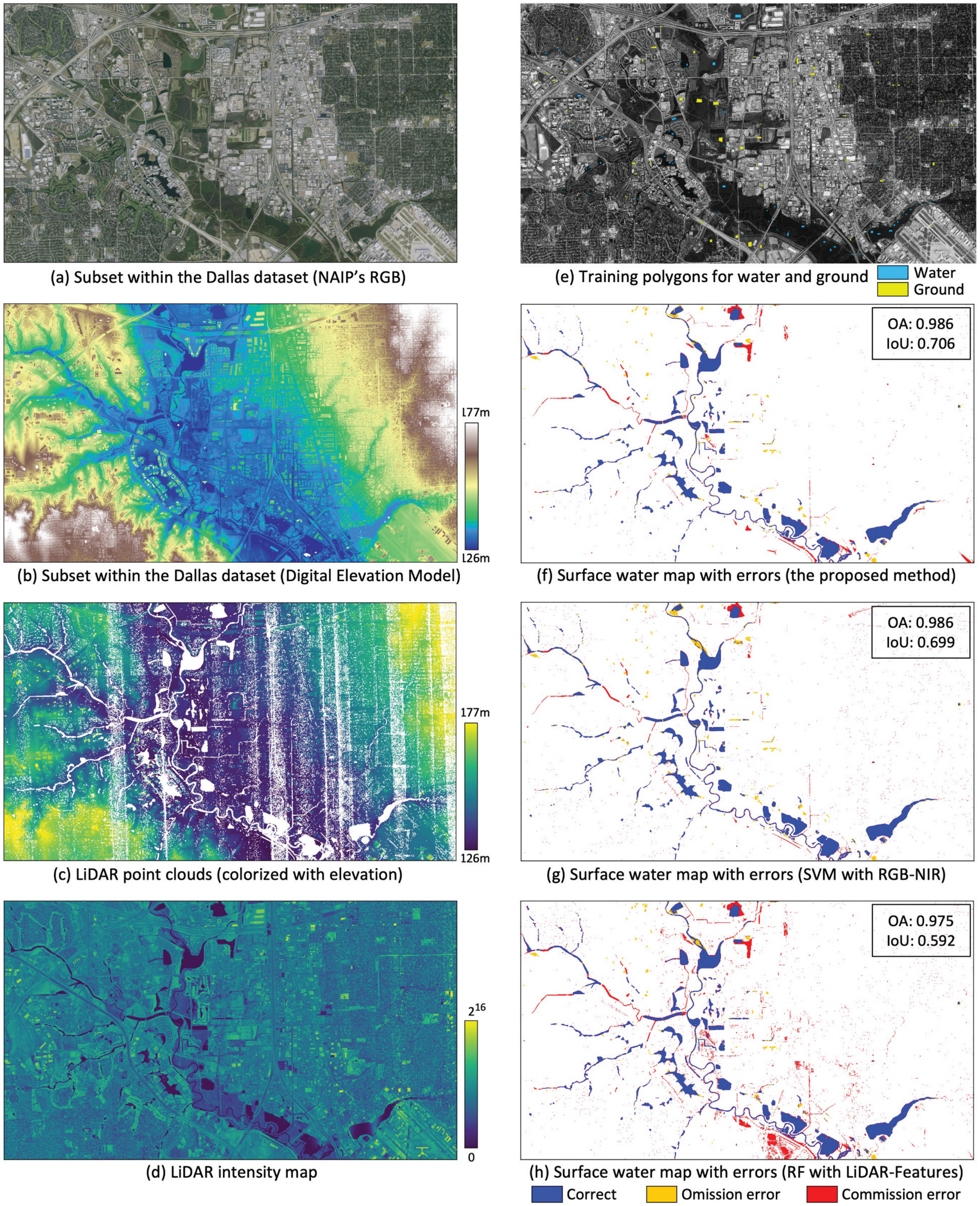


Figure 9. Visualization of the subset area within the Dallas dataset and surface water mapping results. (a) RGB imagery. (b) Digital elevation model. (c) LiDAR point clouds colorized by elevation values. (d) LiDAR intensity map after interpolation. (e) Training sample polygons for water and ground over the grayscale image. (f-h) generated surface water maps and errors for (f) the proposed method, (g) support vector machine with RGB-NIR, and (h) random forest with LiDAR features, respectively. Correctly identified surface water pixels are shown in blue, omission errors in orange, and commission errors in red. Overall accuracy (OA) and intersection over union (IoU) values are provided for each method.

Table 3. Summary of the proposed method's performance against pixel-based machine learning methods for the subset of the Dallas dataset: this table presents the mean and standard deviation of intersection over union (IoU), precision, recall, F1-score, and overall accuracy (OA) evaluated using the NHD plus HR dataset with different numbers of training samples. "RF" indicates the random forest classifier. "SVM"indicates the support vector machine classifier.

Method	Metric	Number of Training Samples 5,000	10,000	15,000	20,000	25,000
RF (RGB-NIR)	IoU	0.667 ± 0.01	0.679 ± 0.01	0.685 ± 0.01	0.687 ± 0.01	0.688 ± 0.01
	Precision	0.807 ± 0.01	0.822 ± 0.01	0.829 ± 0.01	0.831 ± 0.01	0.830 ± 0.01
	Recall	0.794 ± 0.01	0.797 ± 0.00	0.798 ± 0.00	0.798 ± 0.00	0.800 ± 0.00
	F1-score	0.800 ± 0.01	0.809 ± 0.01	0.813 ± 0.00	0.814 ± 0.00	0.815 ± 0.00
	OA	0.984 ± 0.00	0.985 ± 0.00	0.986 ± 0.00	0.986 ± 0.00	0.985 ± 0.00
SVM (RGB-NIR)	IoU	0.690 ± 0.01	0.696 ± 0.01	0.695 ± 0.01	0.694 ± 0.00	0.696 ± 0.00
	Precision	0.808 ± 0.02	0.815 ± 0.01	0.814 ± 0.01	0.813 ± 0.01	0.815 ± 0.00
	Recall	0.826 ± 0.00	0.826 ± 0.00	0.826 ± 0.00	0.826 ± 0.00	0.826 ± 0.00
	F1-score	0.816 ± 0.01	0.821 ± 0.00	0.820 ± 0.00	0.820 ± 0.00	0.821 ± 0.00
	OA	0.985 ± 0.00	0.986 ± 0.00	0.986 ± 0.00	0.986 ± 0.00	0.986 ± 0.00
RF (LiDAR-Features[a])	IoU	0.546 ± 0.04	0.558 ± 0.03	0.557 ± 0.03	0.554 ± 0.03	0.562 ± 0.04
	Precision	0.580 ± 0.04	0.592 ± 0.04	0.593 ± 0.03	0.588 ± 0.03	0.599 ± 0.05
	Recall	0.906 ± 0.01	0.905 ± 0.01	0.903 ± 0.01	0.905 ± 0.01	0.903 ± 0.01
	F1-score	0.706 ± 0.03	0.715 ± 0.03	0.715 ± 0.02	0.712 ± 0.02	0.719 ± 0.03
	OA	0.970 ± 0.01	0.971 ± 0.00	0.971 ± 0.00	0.971 ± 0.00	0.972 ± 0.01
SVM (LiDAR-Features[a])	IoU	0.468 ± 0.03	0.461 ± 0.02	0.460 ± 0.02	0.462 ± 0.02	0.468 ± 0.02
	Precision	0.504 ± 0.03	0.495 ± 0.02	0.493 ± 0.02	0.495 ± 0.03	0.503 ± 0.02
	Recall	0.868 ± 0.01	0.873 ± 0.01	0.874 ± 0.01	0.872 ± 0.01	0.874 ± 0.01
	F1-score	0.637 ± 0.03	0.631 ± 0.02	0.630 ± 0.01	0.632 ± 0.02	0.638 ± 0.02
	OA	0.960 ± 0.00	0.960 ± 0.00	0.959 ± 0.00	0.959 ± 0.00	0.960 ± 0.00
Proposed Method		No Training Needed				
	IoU	0.706				
	Precision	0.800				
	Recall	0.857				
	F1-score	0.827				
	OA	0.986				

[a]LiDAR-Features:Four stacked feature maps derived from LiDAR, including the DSM, intensity, standard deviation of elevation, and point density.

samples, providing a well-mixed dataset, the most favorable for training. North/South involved collecting training data from the northern half, while the southern half was reserved for testing. Cross-state entailed training on one state's dataset (e.g. Denver) and testing on another's (e.g. Wind River), and vice versa. We randomly sampled non-overlapping training and test datasets from Reliable Tiles, which excluded tiles with significant noise, to ensure fair and accurate benchmarking of performance. The same test datasets were used for Random Mix and Cross-state, while only the southern part was used for North/South. Our method's evaluation also used the same corresponding datasets. For U-Net training, we used 256 by 256 pixels size with 1-m resolution input data using the Adam optimizer with a batch size of 16. We explored various learning rates (1×10^{-3}, 1×10^{-4}, 1×10^{-5}) over five iterations for each scenario and reported the highest mean IoU among the three learning rates for each scenario to provide a strong baseline. We applied early stopping with a patience of 5. The implementation was conducted using PyTorch version 2.1.1.

As shown in Table 4, our method achieved comparable accuracy to the supervised method in cases where deep models were trained and tested in the same experimental area (Random Mix and North/South) and obtained significantly higher accuracy in

Table 4. Comparative analysis of our method against deep learning-based semantic segmentation methods (i.e. U-Net) in various scenarios: this table presents the mean IoU and its standard deviation values, evaluated using reliable tiles. "Random Mix"with fully random samples, "North/South"distinguishes training in the northern region and testing in the southern, and "cross-state"involves separate regions (Denver and Wind River) for training and testing. The standard U-Net models are tested using two types of inputs: RGB-NIR imagery and LiDAR-derived feature maps.

Test Dataset	Denver			Wind River		
Training Sampling Method	Random Mix	North/South	Cross-state	Random Mix	North/South	Cross-state
U-Net (RGB-NIR)	0.782 ± 0.01	0.742 ± 0.02	0.574 ± 0.06	0.870 ± 0.01	0.810 ± 0.05	0.711 ± 0.01
U-Net (LiDAR-Features[a])	0.780 ± 0.01	0.771 ± 0.01	0.735 ± 0.04	0.889 ± 0.00	0.887 ± 0.00	0.836 ± 0.06
Proposed Method	0.763	0.753	0.763	0.877	0.877	0.877

[a]LiDAR-Features:Four stacked feature maps derived from LiDAR, including the DSM, intensity, standard deviation of elevation, and point density.

scenarios where models were trained and tested in different areas (Cross-state). Unlike pixel-based machine learning methods, the U-Net with LiDAR-based features demonstrated a higher IoU compared to the one with RGB-NIR, indicating that the deep learning method effectively leveraged the spatial context. However, it was still affected by data distribution shifts, highlighting the need for high-quality training datasets. In contrast, our method automatically ensures precise water boundary demarcation based on elevation, reducing the need for post-processing and streamlining topographic mapping.

6. Discussion

6.1. The impact of parameters

With more extensive experiments, we found that adjustments to our method's parameters within reasonable bounds exhibited a minimal effect on the overall statistics. However, this limited impact is also largely because the statistics are predominantly influenced by substantial water bodies, for which our method rarely generated errors. Specifically, in our tests, the total accuracy seldom varied by more than 1% within an acceptable parameter range (detailed below). Considering the inherent variability in water bodies due to elements like precipitation and evaporation, minor discrepancies made such variations even less significant.

To elucidate the dynamics behind key parameters – z-score (Z), elevational range (ER), and minimum size (MS)—,we selected a region where the impact of parameters is most pronounced. Z sets the confidence interval for initial water segment identification. For more detail, please refer to Section 3.2. A Z value of 1.96 corresponds to a 95% confidence level from a standard normal distribution. In our method, the default Z value is 2. For the parameter analysis, we experimented with Z values of 1.5 (for a more aggressive water extraction) and 2.5 (for a more conservative water extraction). ER specifies an elevation range to identify areas with similar elevations by slicing the DSM in the WERM algorithm. The default for ER is set at 0.1, representing a range of $\pm$ 0.1 m, in line with the typical vertical precision of ALS. We adjusted this default to either 0.05 or 0.15, aiming for conservative and aggressive region merging, respectively. The MS parameter indicates the minimum size of the initial water segment to be expanded by WERM. Its default setting is 500m^2, which excludes minor water entities like puddles. We experimented with values of 100m^2 (aggressive) and 1,000m^2 (conservative).

Figure 10a displays a CIR image and water map derived from a Dallas dataset. With its flat topography and diverse water features, such as rivers, wetlands, and ponds, parameter adjustments noticeably influenced the results. Figure 10b depicts conservative case on the left and an aggressive case on the right. Variations in Z to 2.5 or 1.5 led to subtle changes, underscoring the stability of water pixel identification by Z's role. In contrast, ER adjustments to 0.05 or 0.15 yielded marked differences. Specifically, an ER of 0.05 mitigated "flooding"errors, where minor puddles mistakenly expanded their boundaries to neighboring terrains, while 0.15 amplified these errors. Similarly, an MS of 1,000m^2 reduced such errors, whereas 100m^2 exacerbated them.

We have found that ER and MS influence the outcomes. Yet, their effects, rooted in the physical implications associated with water level and area, are predictable, rendering its error relatively manageable.

6.2. Computational complexity analysis and implementation details

This section presents the programmatic implementation details and computational complexity analysis of the key operations in the proposed surface water mapping.

(1) **DSM Generation**: This operation iterates through all LiDAR data points to create a DSM grid. It starts by creating a grid based on the spatial extent and iterates through each LiDAR point, registering them to the grid based on their x and y coordinates. Both DSMs with and without interpolation are saved. This results in a time complexity of $O(n)$, where n is the total number of points in the LiDAR file. The interpolation step has a time complexity of $O(m \log m)$, where m represents the number of cells in the DSM grid.
(2) **Initial Water Segments Extraction**: This module calculates a cell density map using convolution by counting the number of registered points using the DSM without interpolation.

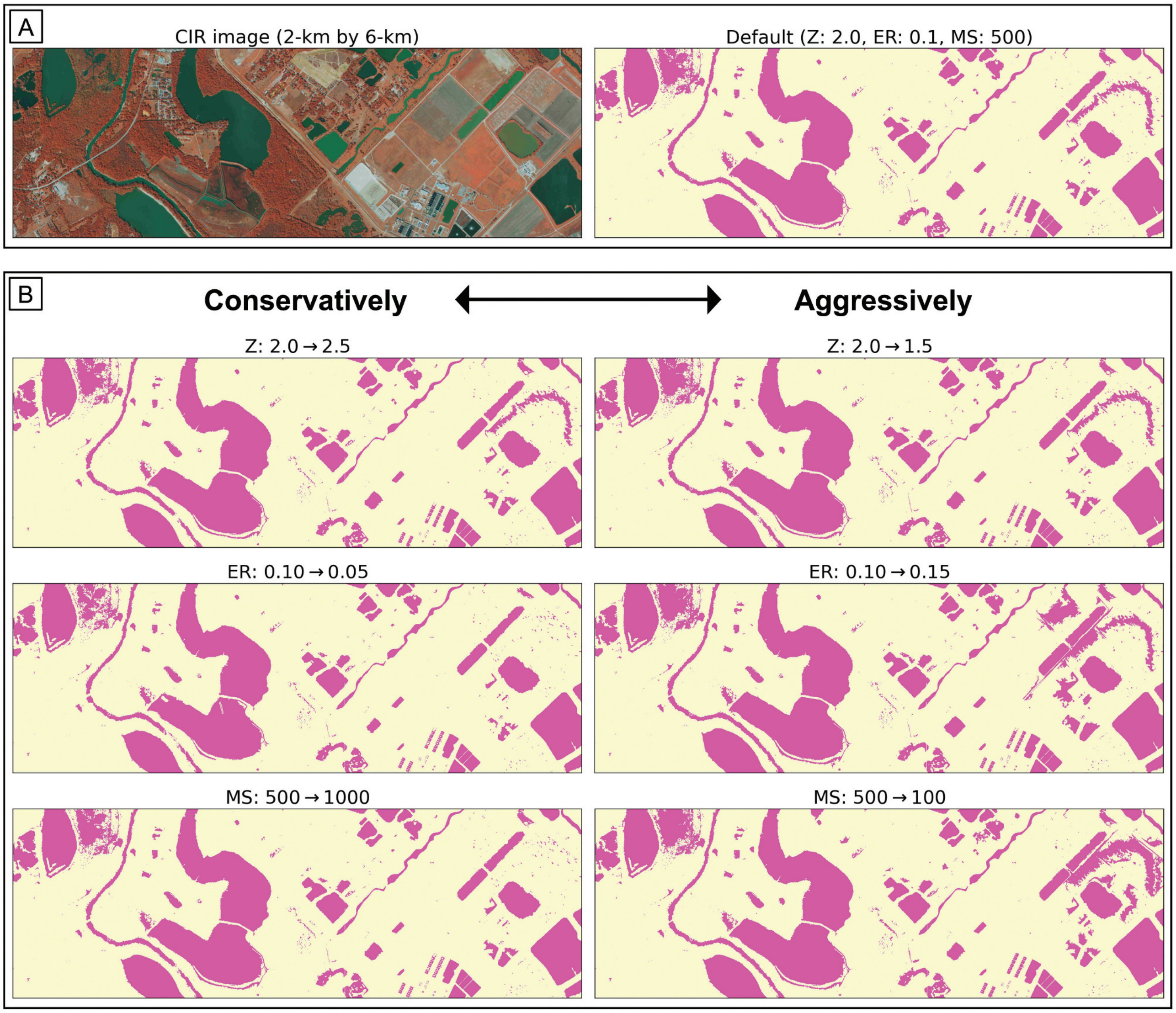


Figure 10. The impact of parameters: z-score (Z), elevation range (ER), and minimum size (MS). (Aa) CIR image contrasted with the water map using default parameters; (b) side-by-side comparison of water maps with conservative (left) versus aggressive (right) parameter adjustments.

Several array operations follow to classify potential water bodies based on density thresholds. Small water bodies are then removed based on a size threshold, resulting in a map of initial water segments. Each of these steps, including convolution and array operations, has a time complexity of $O(m)$, where m is the number of cells in the DSM grid.

(3) **Water Elevation-based Region Merging**: This module uses the map of initial water segments and handles each segment individually. For each segment k, it performs DSM slicing to identify regions within a specific elevation interval and then applies connected component labeling to find contiguous regions. Segments are merged based on elevation similarity, iterating through labeled segments and updating the merged result and elevation map. Both DSM slicing and connected component labeling are $O(m)$ operations. Consequently, the overall complexity is $O(k \times m)$, where k is the number of initial water segments and m is the number of cells in the DSM grid.

The overall system complexity is $O(n + m \log m + k \times m)$, where n is the total number of LiDAR points, m is the number of cells in the DSM grid, and k is the number of initial water segments. In a typical scenario, the total time complexity is primarily influenced by the DSM generation (including interpolation) and the water elevation-based region merging processes. The space complexity is largely determined by the size of the DSM.

The source code for our workflow, written in Python, is available on GitHub: https://github.com/hunsoosong/airborne-lidar-water-mapping. It is important to note that the actual runtime will also be affected by factors such as the characteristics of the input data, the implementation of algorithms for interpolation, the connected components algorithm, and other operations, as well as the computational resources available. Particularly, if the number of initial water segments (k) is large, the computation will be primarily determined by the water elevation-based region merging process. Adjusting the threshold to refine initial water segments by their sizes, thereby eliminating insignificant water bodies before the region merging process, is advised based on the purpose of the study and available resources.

One key aspect of our method is its consideration of spatial connectivity (i.e. connected surface water has a similar elevation). Therefore, providing sufficient spatial context, especially in study areas with large water bodies, would enhance performance. This is also why operating with a tile-based approach (handling a tile that consists of multiple strips of LiDAR scans) is recommended rather than a LiDAR-strip-based operation, which potentially limits spatial context since each strip may not cover enough area. In practice, if computational resources are limited, it is advisable to segment the entire project area into smaller tiles, such as several kilometers by several kilometers, based on the complexity and characteristics of the study area. Under these conditions, the creation of the DSM file often consumes the most time.

6.3. Significance and limitations of the study

Traditional surface water mapping algorithms often require repetitive tuning of site-specific parameters or extensive training processes, complicating scalability. This challenge is especially pronounced in high-resolution water body mapping due to increased noise and difficulty in collecting high-quality training samples for small water bodies (Wieland et al. 2023). Our method addresses these issues by focusing on the geometric properties of water, utilizing only 3D coordinate observations from airborne LiDAR data. This approach has demonstrated robustness compared to other methods, eliminating the need for repetitive parameter tuning and training procedures.

A notable strength of our method is its ability to automatically generate a 3D surface water map with elevation details for each water body. Traditionally, 3D hydrographic data is often produced by overlaying a 2D water layer from optical imagery onto a digital elevation model from LiDAR or SAR data, a process called "hydro-flattening"(Heidemann 2012). This approach can introduce temporal discrepancies and registration errors. In contrast, our method simplifies the process by relying solely on airborne LiDAR data, ensuring both accuracy and seamless integration. The water elevations generated are reliable, with water surfaces remaining flat and boundaries integrating perfectly with adjacent terrains. These features, combined with the elimination of repetitive parameter tuning and training, contribute to scalable and accurate 3D topographic mapping.

Our approach leverages knowledge of the physical properties of water instead of data-driven solutions. This not only improves mapping efficiency but also ensures that errors are relatively predictable and explainable, making them more manageable compared to data-driven approaches. In extensive surface water mapping projects, especially those including small water bodies, it is often infeasible to track and validate all outputs and errors. Data-driven solutions depend heavily on the quality of training samples, and their processes lack transparency, often resulting in biased and less trustworthy outputs (Tuia, Persello, and Bruzzone 2016). Existing literature has found that current surface water maps frequently underestimate the extent of small water bodies, likely due to insufficient training samples for them and a bias toward having samples for larger water bodies (Mao et al. 2022; Ogilvie et al. 2018). In this context, our method, which leverages the physical knowledge explicitly represented in remote sensing data, demonstrates promising potential for knowledge-based mapping systems in this field (Kadambi et al. 2023).

Despite these strengths, there are limitations. Our method identifies initial water segments based on point density, as surface water exhibits laser dropouts due to significant specular reflections (Höfle et al. 2009).

However, small water bodies near the nadir angle may exhibit high point densities over their entire extent, similar to non-water bodies, leading to false negatives. Conversely, large high-rise structures can cause a substantial extent of laser dropouts beyond the size threshold, resulting in false positives. Additionally, while our method assumes that connected surface water has similar elevation, this may not hold in areas with rapid flow (Milan et al. 2010), leading to potential artifacts. In our experiments, we calculated cell density for every LiDAR input tile, typically ranging from 1–10 km^2, and did not identify significant artifacts. However, significant noise in LiDAR data may cause erroneous cell density calculations, potentially leading to artifacts, requiring preprocessing depending on the quality of the LiDAR dataset.

Regarding the experiment, it is essential to acknowledge that the reference data is not without errors. Temporal gaps and the exclusion of small water bodies in the reference contributed to inaccuracies. Evaluating performance against a perfect reference was infeasible due to the expansive evaluation of our study and the dynamic nature of water bodies. While we enhanced the robustness of our comparison by evaluating performance against widely adopted benchmarks and employing both large-area and tile-based evaluations, slight differences in accuracy may not definitively indicate the superiority of one method over another, given the imperfect nature of the reference. The development of extensive benchmark datasets that include diverse small water bodies and multimodal remote sensing data would significantly contribute to this field.

Our method requires high-quality topographic airborne LiDAR data, which, although increasingly available (Moudr et al. 2023; Stoker and Miller 2022), is still limited compared to optical imagery and may not suffice for applications needing frequent water dynamics observations. Additionally, while we confirmed our method's robustness with various topographic ALS campaigns, it may require modifications for other high-altitude laser scanning systems (Yi et al. 2014), where specular reflection in water bodies may not be as apparent. Nevertheless, advancements in LiDAR sensors, particularly in their spatial and spectral resolution, allow LiDAR technology to play a more diverse role in Earth observations, including more detailed classification of water types (Milan et al. 2010; Ricker et al. 2023) and other land cover classes (Morsy et al. 2016; Yu et al. 2021). Future research that integrates the geometric properties of water in LiDAR acquisition with additional spectral information and other modalities would be promising for achieving complementary benefits.

7. Conclusion

This paper presents a novel method for high-resolution surface water mapping, capturing both the extent and elevation of water bodies. Unlike existing approaches, our method relies exclusively on 3D coordinate information from ALS. This unique dependence on coordinates not only enables scalable surface water mapping but also facilitates the generation of comprehensive 3D topographic maps using only airborne LiDAR data. A key advantage is that our method eliminates the need for training while demonstrating robust performance, independent of repetitive site-specific parameters. This approach ensures scalability for large-area mapping projects, offers resilience against potential training sample biases, and holds great potential to address the challenges of increasing noise in high-resolution surface water mapping.

We thoroughly validated our method over diverse landscapes spanning roughly 2,500 km^2 of urban, coastal, and mountainous terrains. The results affirm our method's reliability in generating surface water maps. Our method notably outperformed a highly competitive version of the widely adopted NDWI baseline, even when using only a default set of parameters. Particularly, our technique excels in intricate urban settings with small water bodies, and in mountainous areas, showcasing robustness against factors like shadows, snow, and variable water reflectance. Additionally, it demonstrated effectiveness in terms of both accuracy and scalability, even when contrasted with supervised methods on optical and LiDAR datasets. Although parameter choices in our method can influence the outcomes, their direct correlation to water extent and elevation ensures that any resultant errors are relatively predictable and manageable, especially when compared to optical image-based and data-driven methods.

Acknowledgments

The authors would like to thank the teams at the U.S. Geological Survey's 3D Elevation Program (3DEP).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the Federal Award/Contract [No. HM157522D0009/HM157523F0135] (Subaward No. SPC-1000012108/GR132562).

ORCID

Hunsoo Song http://orcid.org/0000-0001-6899-6770
Jinha Jung http://orcid.org/0000-0003-1176-3540

Author contributions

H.S. developed the algorithm, conducted the analysis, and wrote the manuscript. J.J. offered guidance throughout the process. Both authors designed the experiment and reviewed the final manuscript.

Data availability statement

All the airborne LiDAR data employed in this study are available through the U.S. Geological Survey's 3D Elevation Program (3DEP). The code will be made public upon publication. The source code for our workflow, written in Python, along with sample data, can be accessed on GitHub: https://github.com/hunsoosong/airborne-lidar-water-mappinghttps://github.com/hunsoosong/airborne-lidar-water-mapping.

References

Akwensi, P. H., Z. Kang, and R. Wang. 2023. "Hyperspectral Image-Aided Lidar Point Cloud Labeling via Spatio-Spectral Feature Representation Learning." *International Journal of Applied Earth Observation and Geoinformation* 120:103302. https://doi.org/10.1016/j.jag.2023.103302.

Arrighi, C., and L. Campo. 2019. "Effects of Digital Terrain Model Uncertainties on High-Resolution Urban Flood Damage Assessment." *Journal of Flood Risk Management* 12 (S2): e12530. https://doi.org/10.1111/jfr3.12530.

Awadallah, M. O. M., A. Juárez, and K. Alfredsen. 2022. "Comparison Between Topographic and Bathymetric Lidar Terrain Models in Flood Inundation Estimations." *Remote Sensing* 14 (1): 227. https://doi.org/10.3390/rs14010227.

Becker, R. H., M. Sayers, D. Dehm, R. Shuchman, K. Quintero, K. Bosse, and R. Sawtell. 2019. "Unmanned Aerial System Based Spectroradiometer for Monitoring Harmful Algal Blooms: A New Paradigm in Water Quality Monitoring." *Journal of Great Lakes Research* 45 (3): 444–453. https://doi.org/10.1016/j.jglr.2019.03.006.

Bijeesh, T., and K. Narasimhamurthy. 2020. "Surface Water Detection and Delineation Using Remote Sensing Images: A Review of Methods and Algorithms." *Sustainable Water Resources Management* 6 (4): 1–23. https://doi.org/10.1007/s40899-020-00425-4.

Brzank, A., C. Heipke, J. Goepfert, and U. Soergel. 2008. "Aspects of Generating Precise Digital Terrain Models in the Wadden Sea from Lidar–Water Classification and Structure Line Extraction." *ISPRS Journal of Photogrammetry & Remote Sensing* 63 (5): 510–528. https://doi.org/10.1016/j.isprsjprs.2008.02.002.

Chen, F., X. Chen, T. Van de Voorde, D. Roberts, H. Jiang, and W. Xu. 2020 a. "Open Water Detection in Urban Environments Using High Spatial Resolution Remote Sensing Imagery." *Remote Sensing of Environment* 242:111706. https://doi.org/10.1016/j.rse.2020.111706.

Chen, Y., L. Tang, Z. Kan, M. Bilal, and Q. Li. 2020 b. "A Novel Water Body Extraction Neural Network (Wbe-Nn) for Optical High-Resolution Multispectral Imagery." *Journal of Hydrology* 588:125092. https://doi.org/10.1016/j.jhydrol.2020.125092.

Crasto, N., C. Hopkinson, D. Forbes, L. Lesack, P. Marsh, I. Spooner, and J. Van Der Sanden. 2015. "A Lidar-Based Decision-Tree Classification of Open Water Surfaces in an Arctic Delta." *Remote Sensing of Environment* 164:90–102. https://doi.org/10.1016/j.rse.2015.04.011.

Dong, Y., L. Fan, J. Zhao, S. Huang, C. Geiß, L. Wang, and H. Taubenböck. 2022. "Mapping of Small Water Bodies with Integrated Spatial Information for Time Series Images of Optical Remote Sensing." *Journal of Hydrology* 614:128580. https://doi.org/10.1016/j.jhydrol.2022.128580.

Guo, F., Z. Li, Q. Meng, G. Ren, L. Wang, J. Wang, H. Qin, and J. Zhang. 2023. "Semi-Supervised Cross-Domain Feature Fusion Classification Network for Coastal Wetland Classification with Hyperspectral and Lidar Data." *International Journal of Applied Earth Observation and Geoinformation* 120:103354. https://doi.org/10.1016/j.jag.2023.103354.

Heidemann, H. K. 2012. Lidar Base Specification. *Technical Report*. US Geological Survey.

Hoekstra, A. Y., J. Buurman, and K. C. Van Ginkel. 2018. "Urban Water Security: A Review." *Environmental Research Letters* 13 (5): 053002. https://doi.org/10.1088/1748-9326/aaba52.

Höfle, B., M. Vetter, N. Pfeifer, G. Mandlburger, and J. Stötter. 2009. "Water Surface Mapping from Airborne Laser Scanning Using Signal Intensity and Elevation Data." *Earth Surface Processes and Landforms* 34 (12): 1635–1649. https://doi.org/10.1002/esp.1853.

Irwin, K., D. Beaulne, A. Braun, and G. Fotopoulos. 2017. "Fusion of Sar, Optical Imagery and Airborne Lidar for Surface Water Detection." *Remote Sensing* 9 (9): 890. https://doi.org/10.3390/rs9090890.

Isikdogan, L. F., A. Bovik, and P. Passalacqua. 2019. "Seeing Through the Clouds with Deepwatermap." *IEEE Geoscience*

& Remote Sensing Letters 17 (10): 1662–1666. https://doi.org/10.1109/LGRS.2019.2953261.

Janowski, Ł., D. Skarlatos, P. Agrafiotis, P. Tysiac, A. Pydyn, M. Popek, A. M. Kotarba-Morley, et al. 2024. "High Resolution Optical and Acoustic Remote Sensing Datasets of the Puck Lagoon." *Scientific Data* 11 (1): 360. https://doi.org/10.1038/s41597-024-03199-y.

Ji, L., L. Zhang, and B. Wylie. 2009. "Analysis of Dynamic Thresholds for the Normalized Difference Water Index." *Photogrammetric Engineering & Remote Sensing* 75 (11): 1307–1317. https://doi.org/10.14358/PERS.75.11.1307.

Kadambi, A., C. de Melo, C.-J. Hsieh, M. Srivastava, and S. Soatto. 2023. "Incorporating Physics into Data-Driven Computer Vision." *Nature Machine Intelligence* 5 (6): 572–580. https://doi.org/10.1038/s42256-023-00662-0.

Kastdalen, L., M. Stickler, C. Malmquist, and J. Heggenes. 2024. "Evaluating Methods for Measuring In-River Bathymetry: Remote Sensing Green Lidar Provides High-Resolution Channel Bed Topography Limited by Water Penetration Capability." *River Research and Applications* 40 (4): 467–482. https://doi.org/10.1002/rra.4245.

Kelly-Quinn, M., J. Biggs, S. Brooks, P. Fortuño, S. Hegarty, J. Jones, and F. Regan. 2022. "Opportunities, Approaches and Challenges to the Engagement of Citizens in Filling Small Water Body Data Gaps." *Hydrobiologia* 850 (15): 3419–3439. https://doi.org/10.1007/s10750-022-04973-y.

Khandelwal, A., A. Karpatne, M. E. Marlier, J. Kim, D. P. Lettenmaier, and V. Kumar. 2017. "An Approach for Global Monitoring of Surface Water Extent Variations in Reservoirs Using Modis Data." *Remote Sensing of Environment* 202:113–128. https://doi.org/10.1016/j.rse.2017.05.039.

Ko, B. C., H. H. Kim, and J. Y. Nam. 2015. "Classification of Potential Water Bodies Using Landsat 8 Oli and a Combination of Two Boosted Random Forest Classifiers." *Sensors (Switzerland)* 15 (6): 13763–13777. https://doi.org/10.3390/s150613763.

Kyzivat, E. D., and L. C. Smith. 2023. "Contemporary and Historical Detection of Small Lakes Using Super Resolution Landsat Imagery: Promise and Peril." *GIScience & Remote Sensing* 60 (1): 2207288. https://doi.org/10.1080/15481603.2023.2207288.

Liu, Q., S. Zhang, N. Wang, Y. Ming, and C. Huang. 2022. "Fusing Landsat-8, Sentinel-1, and Sentinel-2 Data for River Water Mapping Using Multidimensional Weighted Fusion Method." *IEEE Transactions on Geoscience & Remote Sensing* 60:1–12. https://doi.org/10.1109/TGRS.2022.3187154.

Luo, X., X. Tong, and Z. Hu. 2021. "An Applicable and Automatic Method for Earth Surface Water Mapping Based on Multispectral Images." *International Journal of Applied Earth Observation and Geoinformation* 103:102472. https://doi.org/10.1016/j.jag.2021.102472.

Ma, D., L. Jiang, J. Li, and Y. Shi. 2023. "Water Index and Swin Transformer Ensemble (Wiste) for Water Body Extraction from Multispectral Remote Sensing Images." *GIScience & Remote Sensing* 60 (1): 2251704. https://doi.org/10.1080/15481603.2023.2251704.

Malinowski, R., B. Höfle, K. Koenig, G. Groom, W. Schwanghart, and G. Heckrath. 2016. "Local-Scale Flood Mapping on Vegetated Floodplains from Radiometrically Calibrated Airborne Lidar Data." *ISPRS Journal of Photogrammetry & Remote Sensing* 119:267–279. https://doi.org/10.1016/j.isprsjprs.2016.06.009.

Mandlburger, G., M. Pfennigbauer, R. Schwarz, S. Flöry, and L. Nussbaumer. 2020. "Concept and Performance Evaluation of a Novel Uav-Borne Topo-Bathymetric Lidar Sensor." *Remote Sensing* 12 (6): 986. https://doi.org/10.3390/rs12060986.

Mao, W., K. Yang, W. Zhang, Y. Wang, and M. Li. 2022. "High-Resolution Global Water Body Datasets Underestimate the Extent of Small Rivers." *International Journal of Remote Sensing* 43 (11): 4315–4330. https://doi.org/10.1080/01431161.2022.2111531.

Martins, V. S., C. C. F. Barbosa, L. A. S. De Carvalho, D. S. F. Jorge, F. D. L. Lobo, and E. M. L. D. M. Novo. 2017. "Assessment of Atmospheric Correction Methods for Sentinel-2 Msi Images Applied to Amazon Floodplain Lakes." *Remote Sensing* 9 (4): 322. https://doi.org/10.3390/rs9040322.

Milan, D., G. Heritage, A. Large, and N. Entwistle. 2010. "Mapping Hydraulic Biotopes Using Terrestrial Laser Scan Data of Water Surface Properties." *Earth Surface Processes and Landforms* 35 (8): 918–931. https://doi.org/10.1002/esp.1948.

Moore, R. B., L. D. McKay, A. H. Rea, T. R. Bondelid, C. V. Price, T. G. Dewald, and C. M. Johnston. 2019. "User's Guide for the National Hydrography Dataset Plus (Nhdplus) High Resolution." *Open-File Report-US Geological Survey* 0 (2019–1096). https://doi.org/10.3133/ofr20191096.

Morsy, S., A. Shaker, A. El-Rabbany, and P. E. LaRocque. 2016. "Airborne Multispectral Lidar Data for Land-Cover Classification and Land/Water Mapping Using Different Spectral Indexes." *ISPRS Annals of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 3:217–224. https://doi.org/10.5194/isprs-annals-III-3-217-2016.

Moudr, V., A. F. Cord, L. Gábor, G. V. Laurin, V. Barták, K. Gdulová, M. Malavasi, et al. 2023. "Vegetation Structure Derived from Airborne Laser Scanning to Assess Species Distribution and Habitat Suitability: The Way Forward." *Diversity & Distributions* 29 (1): 39–50. https://doi.org/10.1111/ddi.13644.

Musa, Z., I. Popescu, and A. Mynett. 2015. "A Review of Applications of Satellite Sar, Optical, Altimetry and Dem Data for Surface Water Modelling, Mapping and Parameter Estimation." *Hydrology and Earth System Sciences* 19 (9): 3755–3769. https://doi.org/10.5194/hess-19-3755-2015.

Ogilvie, A., G. Belaud, S. Massuel, M. Mulligan, P. Le Goulven, and R. Calvez. 2018. "Surface Water Monitoring in Small Water Bodies: Potential and Limits of Multi-Sensor Landsat Time Series." *Hydrology and Earth System Sciences* 22 (8): 4349–4380. https://doi.org/10.5194/hess-22-4349-2018.

Pan, S., H. Guan, Y. Yu, J. Li, and D. Peng. 2019. "A Comparative Land-Cover Classification Feature Study of Learning Algorithms: Dbm, Pca, and Rf Using Multispectral Lidar Data." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 12 (4): 1314–1326. https://doi.org/10.1109/JSTARS.2019.2899033.

Pekel, J.-F., A. Cottam, N. Gorelick, and A. S. Belward. 2016. "High-Resolution Mapping of Global Surface Water and Its

Long-Term Changes." *Nature* 540 (7633): 418–422. https://doi.org/10.1038/nature20584.

Pickens, A. H., M. C. Hansen, M. Hancher, S. V. Stehman, A. Tyukavina, P. Potapov, B. Marroquin, and Z. Sherani. 2020. "Mapping and Sampling to Characterize Global Inland Water Dynamics from 1999 to 2018 with Full Landsat Time-Series." *Remote Sensing of Environment* 243:111792. https://doi.org/10.1016/j.rse.2020.111792.

Prošek, J., K. Gdulová, V. Barták, J. Vojar, M. Solsk, D. Rocchini, and V. Moudr. 2020. "Integration of Hyperspectral and Lidar Data for Mapping Small Water Bodies." *International Journal of Applied Earth Observation and Geoinformation* 92:102181. https://doi.org/10.1016/j.jag.2020.102181.

Ricker, R., S. Fons, A. Jutila, N. Hutter, K. Duncan, S. L. Farrell, N. T. Kurtz, and R. M. Fredensborg Hansen. 2023. "Linking Scales of Sea Ice Surface Topography: Evaluation of Icesat-2 Measurements with Coincident Helicopter Laser Scanning During Mosaic." *The Cryosphere* 17 (3): 1411–1429. https://doi.org/10.5194/tc-17-1411-2023.

Sanders, B. F., J. E. Schubert, D. T. Kahl, K. J. Mach, D. Brady, A. AghaKouchak, F. Forman, R. A. Matthew, N. Ulibarri, and S. J. Davis. 2022. "Large and Inequitable Flood Risks in Los Angeles, California." *Nature Sustainability* 6 (1): 47–57. https://doi.org/10.1038/s41893-022-00977-7.

Shaker, A., W. Y. Yan, and P. E. LaRocque. 2019. "Automatic Land-Water Classification Using Multispectral Airborne Lidar Data for Near-Shore and River Environments." *Isprs Journal of Photogrammetry & Remote Sensing* 152:94–108. https://doi.org/10.1016/j.isprsjprs.2019.04.005.

Smeeckaert, J., C. Mallet, N. David, N. Chehata, and A. Ferraz. 2013. "Large-Scale Classification of Water Areas Using Airborne Topographic Lidar Data." *Remote Sensing of Environment* 138:134–148. https://doi.org/10.1016/j.rse.2013.07.004.

Song, H., and J. Jung. 2022. "An Unsupervised, Open-Source Workflow for 2d and 3d Building Mapping from Airborne Lidar Data." *arXiv Preprint arXiv: 2205.14585*.

Song, H., Y. Kim, and Y. Kim. 2019. "A Patch-Based Light Convolutional Neural Network for Land-Cover Mapping Using Landsat-8 Images." *Remote Sensing* 11 (2): 114. https://doi.org/10.3390/rs11020114.

Stoker, J., and B. Miller. 2022. "The Accuracy and Consistency of 3d Elevation Program Data: A Systematic Analysis." *Remote Sensing* 14 (4): 940. https://doi.org/10.3390/rs14040940.

Sun, X., L. Li, B. Zhang, D. Chen, and L. Gao. 2015. "Soft Urban Water Cover Extraction Using Mixed Training Samples and Support Vector Machines." *International Journal of Remote Sensing* 36 (13): 3331–3344. https://doi.org/10.1080/01431161.2015.1042594.

Szafarczyk, A., and C. To. 2023. "The Use of Green Laser in Lidar Bathymetry: State of the Art and Recent Advancements." *Sensors (Switzerland)* 23 (1): 292. https://doi.org/10.3390/s23010292.

Tayer, T. C., M. M. Douglas, M. C. Cordeiro, A. D. Tayer, J. N. Callow, L. Beesley, and D. McFarlane. 2023. "Improving the Accuracy of the Water Detect Algorithm Using Sentinel-2, Planetscope and Sharpened Imagery: A Case Study in an Intermittent River." *GIScience & Remote Sensing* 60 (1): 2168676. https://doi.org/10.1080/15481603.2023.2168676.

Tuia, D., C. Persello, and L. Bruzzone. 2016. "Domain Adaptation for the Classification of Remote Sensing Data: An Overview of Recent Advances." *IEEE Geoscience and Remote Sensing Magazine* 4 (2): 41–57. https://doi.org/10.1109/MGRS.2016.2548504.

Wang, Y., G. Foody, X. Li, Y. Zhang, P. Zhou, and Y. Du. 2023. "Regression-Based Surface Water Fraction Mapping Using a Synthetic Spectral Library for Monitoring Small Water Bodies." *GIScience & Remote Sensing* 60 (1): 2217573. https://doi.org/10.1080/15481603.2023.2217573.

Wieland, M., S. Martinis, R. Kiefl, and V. Gstaiger. 2023. "Semantic Segmentation of Water Bodies in Very High-Resolution Satellite and Aerial Images." *Remote Sensing of Environment* 287:113452. https://doi.org/10.1016/j.rse.2023.113452.

Xu, P., M. Herold, N.-E. Tsendbazar, and J. G. Clevers. 2020. "Towards a Comprehensive and Consistent Global Aquatic Land Cover Characterization Framework Addressing Multiple User Needs." *Remote Sensing of Environment* 250:112034. https://doi.org/10.1016/j.rse.2020.112034.

Yan, W. Y. 2023. "Airborne Lidar Data Artifacts: What We Know Thus Far." *IEEE Geoscience and Remote Sensing Magazine* 11 (3): 21–45. https://doi.org/10.1109/MGRS.2023.3285261.

Yan, W. Y., A. Shaker, and P. E. LaRocque. 2019. "Scan Line Intensity-Elevation Ratio (Slier): An Airborne Lidar Ratio Index for Automatic Water Surface Mapping." *Remote Sensing* 11 (7): 814. https://doi.org/10.3390/rs11070814.

Yang, X., Q. Qin, P. Grussenmeyer, and M. Koehl. 2018. "Urban Surface Water Body Detection with Suppressed Built-Up Noise Based on Water Indices from Sentinel-2 Msi Imagery." *Remote Sensing of Environment* 219 (0): 259–270. https://doi.org/10.1016/j.rse.2018.09.016.

Yi, D., J. P. Harbeck, S. S. Manizade, N. T. Kurtz, M. Studinger, and M. Hofton. 2014. "Arctic Sea Ice Freeboard Retrieval with Waveform Characteristics for nasa's Airborne Topographic Mapper (Atm) and Land, Vegetation, and Ice Sensor (Lvis)." *IEEE Transactions on Geoscience & Remote Sensing* 530 (3): 1403–1410. https://doi.org/10.1109/TGRS.2014.2339737.

Yu, Y., T. Jiang, J. Gao, H. Guan, D. Li, S. Gao, E. Tang, W. Wang, P. Tang, and J. Li. 2022. "Capvit: Cross-Context Capsule Vision Transformers for Land Cover Classification with Airborne Multispectral Lidar Data." *International Journal of Applied Earth Observation and Geoinformation* 111:102837. https://doi.org/10.1016/j.jag.2022.102837.

Yu, Y., C. Liu, H. Guan, L. Wang, S. Gao, H. Zhang, Y. Zhang, and J. Li. 2021. "Land Cover Classification of Multispectral Lidar Data with an Efficient Self-Attention Capsule Network." *IEEE Geoscience & Remote Sensing Letters* 19:1–5. https://doi.org/10.1109/LGRS.2021.3071252.

Zhao, P., H. Guan, D. Li, Y. Yu, H. Wang, K. Gao, J. M. Junior, and J. Li. 2021. "Airborne Multispectral Lidar Point Cloud Classification with a Feature Reasoning-Based Graph Convolution Network." *International Journal of Applied Earth Observation and Geoinformation* 105:102634. https://doi.org/10.1016/j.jag.2021.102634.