Investigating the Influence of Satellite Sensor Characteristics and Atmospheric Correction on Arctic Lake Mapping

by

James Maze

A thesis accepted and approved in partial fulfillment of the

Requirements for the degree of

Master of Science

in Geography

Thesis Committee:

Sarah Cooley, Chair

Johnathan C. Ryan, Co-Chair

Mark Fonstad, Co-Chair

University of Oregon

Spring 2025

THESIS ABSTRACT

James Maze

Master of Science in Geography

Title: Investigating the influence of satellite sensor characteristics and atmospheric correction on Arctic lake mapping

Shifting hydrologic balances and thawing permafrost are transforming lake area dynamics in the Arctic. Many studies have used satellite imagery to investigate seasonal to long-term trends in lake area; however, persistent uncertainty remains regarding both the timing and magnitude of these fluctuations. A potential explanation for some of this disagreement may be differences between satellite sensors and processing levels. Here we investigate the influence of sensor properties and data-processing specifications – specifically spatial resolution, band characteristics, and atmospheric correction – on surface water detection in the western Canada and Alaska. Using 226 sets of contemporaneous Landsat 8 and Sentinel-2 images, we compare reflectance values and water classification outcomes at the surface reflectance (SR) and top-of-atmosphere (TOA) processing levels. We explicitly quantify reflectance biases relative to lake positions to better understand how these biases may impact water classification. In TOA imagery, we find Sentinel-2 reported higher NDWI values at lake margins, which resulted in wider lake extents. Though using SR data mitigated these cross-sensor reflectance biases, reduced water fractions by lowering Green reflectance more than NIR (i.e., lower NDWI values) along lake margins and thus underestimated lake areas compared to TOA images. SR data also led to increased image-to-image variability and lower consistency in lake detection. Both cross-sensor and cross-atmospheric correction differences were amplified at lake edges and over small lakes (<0.5 km^2). Additionally, regions with high seasonal lake variability exhibited greater disagreement in both cross-sensor and atmospheric correction level comparisons. These results demonstrate the surprising sensitivity of water classification to both sensor characteristics and atmospheric correction when tracking small lakes and dynamic shorelines. Overall, we conclude that while fusing multiple satellites can offer increased observation frequency, careful attention must be paid to differences between sensors to ensure consistency in detection of lake dynamics.

ACKNOWLEDGMENTS

I am grateful for the support of my advisor Professor Sarah Cooley and counsel from my committee members Professor Johnathan Ryan and Professor Mark Fonstad. In addition to her insights, Professor Cooley gave me autonomy to explore an ambitious research topic of my choosing. She trusted my judgement while steering me whenever I needed guidance. Like any compelling research, this thesis has raised new questions, and I only wish there was more time to explore them.

My analysis was possible because Professor Ryan showed me the tools of computationally intensive science; these hard-won analytical skills will foster a stable and fulfilling career. For that, I am deeply grateful. I also thank other members of the COOL lab group, particularly Dr. Elizabeth Webb and Eric Levenson. Dr. Webb's ideas about Arctic lake trend uncertainties inspired me to explore novel sources of inconsistency.

Finally, and most importantly, I would like to thank my parents, Lori and Chester Maze. Their constant encouragement throughout my education gave me the freedom to explore a wide range of opportunities, including this one.

TABLE OF CONTENTS

THESIS ABSTRACT ... 3
ACKNOWLEDGMENTS ... 4
1. INTRODUCTION ... 7
2. METHODS ... 11
2.1 Study Regions ... 11
2.2 Input datasets ... 12
2.2.1 Image acquisition and pre-processing ... 12
2.2.2 The Prior Lake Database (PLD) ... 15
2.2.3 GSWO and GLAD seasonal fractions ... 16
2.3 Image analysis ... 17
2.3.1 Reflectance comparisons ... 17
2.3.2 Surface water classification ... 18
2.4 Statistical tests ... 20
3 RESULTS ... 22
3.1 Comparing Sentinel-2 and Landsat 8 reflectance ... 22
3.1.1 Agreement between Sentinel-2 and Landsat 8 within AC processing ... 22
3.1.2 Agreement between TOA and SR within sensors ... 24
3.2 Sentinel-2 and Landsat 8 water fractions ... 26
3.2.1 Water fraction differences in TOA images ... 26
3.2.2 AC's impact on water fraction discrepancies ... 30
3.3 Lake size and seasonal fraction explain regional impacts ... 34
3.3.1 Size influences AC and sensor discrepancies ... 35
3.3.2 Seasonal associations with sensor discrepancies ... 37
4 DISCUSSION ... 39
4.1 Implications for dynamic lake mapping ... 39
4.2 Future work and limitations ... 41
5 CONCLUSION ... 44
BIBLIOGRAPHY ... 46

LIST OF FIGURES AND TABLES

Figure 1: Hydrogeographic study regions and image target areas 12

Table 1: Coincident image observation counts by hydrogeographic region and month 12

Figure 2: RGB and NIR relative spectral response functions for Sentinel-2 and Landsat 8........ 14

Figure 3: Visualizing the pre-processing workflow and PLD lake positions. 16

Figure 4: Example NDWI histogram for Otsu and Adaptive Classification............................... 20

Figure 5: Correlation coefficients and percent of biased pixels between Sentinel-2 and Landsat 8 reflectance comparisons.. 24

Figure 6: Correlation coefficients and percent of biased pixels between TOA and SR reflectance comparisons 26

Figure 7: Sentinel-2 and Landsat 8 water fractions in lake position masks by atmospheric correction level and resampling method.. 28

Figure 8: Relative and absolute differences between Sentinel-2 and Landsat 8 water fractions across lake positions, atmospheric correction, and resampling methods.................................... 29

Figure 9: Relative and absolute differences between TOA and SR water fractions across lake positions, sensors, and resampling methods 31

Figure 10: Visualizing RGB images, NDWI values, and lake extents 33

Figure 11: Regional impact on cross-sensor and AC discrepancies 35

Figure 12: Lake size influence on cross-sensor and AC differences.. 37

Figure 13: Associations between image target area's GSWO & GLAD seasonal water fractions and magnitude of cross-sensor discrepancies in coincidence scenes... 38

1. INTRODUCTION

Northern lakes and ponds are currently experiencing substantial change due to climate warming and thawing permafrost (Bring et al., 2016). The transformation of northern waters threatens vulnerable ecologies, cultural practices such as subsistence hunting (Knoll et al., 2019; Roach & Griffith, 2015), and carries global significance for permafrost carbon emissions (Hastie et al., 2018; Walter Anthony et al., 2016). Because in-situ data is limited, quantifying changes to northern lakes at regional to global scales typically requires the use of satellite imagery. Instability of northern lakes caused by thawing permafrost and shifting meteorological water balances is a decadal-scale process (Lara et al., 2021; Ulrich et al., 2017; Webb & Liljedahl, 2023) requiring continuous years of observation. The timescale of an individual lake's area change can be rapid (e.g., weeks) or span multiple decades. Catastrophic lake drainage can occur, where lakes lose more than 50% of their area in a few months (Jones & Arp, 2015; Ingmar Nitze et al., 2020); however, this represents only one of many geomorphological and climatological mechanisms for lake growth or shrinkage. Even in regions with catastrophic drainage, total lake areas may still be increasing due to other factors (Grosse et al., 2013). Oftentimes, lake change in the Arctic occurs gradually at lake margins (Grosse et al., 2013; Lara et al., 2021), requiring finer-scale observations for accurate detection.

Despite a large and long-standing record of research into Arctic lake change (Jones & Arp, 2015; Lara et al., 2021; Smith et al., 2005), studies often yield opposite lake area trends even in the same geographical regions (Webb & Liljedahl, 2023). For example, in the Alaska Coastal Plain three studies (Jones et al., 2017; Jorgenson et al., 2018; Pastick et al., 2019) found positive lake area trends, while two studies (Andresen & Lougheed, 2015; Ingmar Nitze et al., 2017) found negative lake area trends another study (Riordan et al., 2006) found no trend. Other places such as the Western Siberian Lowlands have more consistent reports of negative lake area trends (Kirpotin et al., 2009; I. Nitze et al., 2018; Smith et al., 2005). Geographic disagreement in ongoing lake area trends makes forecasts about future trends highly uncertain with climate change. Accurate projections about northern lake extents are important to carbon mineralization and release in thawing permafrost (Hastie et al., 2018; Walter Anthony et al., 2016).

The reasons for this disagreement remain an area of active research and debate, but one key cause may be differences between sensors used to track lake area change. For example, studies use sensors with a range of image resolutions. While most decadal-scale studies use various Landsat satellites (M. Chen et al., 2014; Lara et al., 2021; Lindgren et al., 2021; Olthof et al., 2015; Webb & Liljedahl, 2023), coarser (e.g. MODIS) sensors are sometimes used for global and Pan-Arctic analyses (Carroll et al., 2009; Sharma et al., 2015; Webb et al., 2022). More recently, higher resolution aerial and satellite imagery are increasingly used to enable tracking of finer-scale processes (Cooley et al., 2019; Finger Higgens et al., 2019; Jones et al., 2020; Mullen et al., 2023). The increasing use of global surface water products (Y. Chen et al., 2023; Pastick et al., 2019; Pi et al., 2022) which fuse image timeseries from multiple Landsat satellites to generate monthly occurrence, may also influence lake trend detection. Two notable global products are the Global Surface Water Occurrence (Pekel et al., 2016) and the Global Land Analysis & Discovery (Pickens et al., 2020) datasets. Despite their utility, these global datasets contain errors in Arctic regions (Olthof & Rainville, 2022) with elevated disagreement between products at high latitudes (Rajib et al., 2024).

The differences between satellites can be broadly summarized into their temporal resolution, spatial resolution, spectral characteristics, as well as processing levels (i.e., atmospheric correction). A sensor's spatial resolution has clearly documented impacts on lake mapping (Cooley et al., 2019), with finer spatial resolution sensors capturing more small water features and higher lake fractions (Mullen et al., 2023; Muster et al., 2013). Another contributing challenge in the Arctic is that the long durations of ice cover and frequent summer clouds limit observation frequencies, making surface water mappings derived from coarse temporal resolution (i.e. Landsat) imagery reliant on relatively few ice-free scenes per year.

Along with differences in spatial and temporal resolution, spectral differences between sensors may also influence lake mapping. The range and central wavelengths of each sensor's various bands can influence reflectance values, yet there is little research on this effect specifically for surface water. In analyses of vegetation (NDVI) trends, studies have found that sensors' unique spectral response functions can shift reflectance values as sensors change (Guay et al., 2014; Holden & Woodcock, 2016). Empirical corrections can harmonize sensors with different spectral response functions (Chastain et al., 2019; Claverie et al., 2018; Roy et al., 2016), but their performance can vary by region (J. Chen & Zhu, 2022; Flood, 2017) and across landscape features, where reflectance traits can be unique. Interestingly, one analysis comparing Sentinel-2 and Landsat 8 reflectance in the Arctic removed all surface water and shoreline pixels due to high disagreement, particularly in near-infrared (NIR) (Runge & Grosse, 2019). A likely explanation is the difference in the spectral response functions of Sentienl-2 and Landsat 8's NIR bands (Figure 2) (Chastain et al., 2019). Crucially, NIR is an important feature for lake mapping due to water's low reflectance in NIR relative to visible wavelengths (McFeeters, 1996; Pickens et al., 2020). Because lakes generally exhibit low reflectance but heterogenous optical properties (Pahlevan et al., 2021), cross-sensor reflectance could be inconsistent along lakes and shorelines. Whether these reflectance differences significantly impact lake mapping outcomes remains an open question. Nevertheless, the expansion of satellite constellations and increasing use of sensor fusion studies necessitate understanding of continuity and consistency between sensors.

Lastly, differences in image processing levels, specifically the use of atmospheric correction (AC), may also contribute to disagreement between trends. The majority of surface water area trend studies use top-of-atmosphere (TOA) data (uncorrected for atmospheric effects), for two reasons. First, applying AC to decadal-scale studies can be complicated by algorithmic differences as data available for surface reflectance conversion change over time. Second, many studies argue that water features are easily separable from land in TOA data such that conversion to surface reflectance is not required for classification (Pickens et al., 2020; Yamazaki et al., 2015). However, there are several examples of analyses that use SR data (Janiec et al., 2023; Lara et al., 2021; Qin et al., 2023), which some papers (e.g., Yamazaki et al. 2015) have postulated that consistent use and availability of SR data would improve water classification accuracy and consistency. Still, no studies have yet formally quantified the benefits and trade-offs of modern AC algorithms specifically for water classification. AC significantly lowers reflectance values over surface water, but the degree of reduction differs across spectral bands (Kuhn et al., 2019). Because lake mapping relies on band indices like NDWI, AC-induced shifts in band reflectance could either sharpen or muddle classification accuracy. Despite rigorous evaluation of AC in water quality remote sensing (Lima et al., 2025; Maciel et al., 2023), the its impact on lake mapping remains unexplored.

Here we investigate sources of uncertainty and disagreement in lake area detection caused by differences in sensor specifications and atmospheric correction levels. Specifically, we analyze 226 pairs of same-day Landsat 8 and Sentinel-2 images covering Alaska and Western Canada during the ice-free seasons over 2018-2024. We first quantify the correlations and biases between Green, NIR, and NDWI reflectance values, examining how they vary with atmospheric correction and the distinct spectral responses of Sentinel-2 and Landsat 8 sensors. While other studies have previously compared Sentinel-2 and Landsat 8-pixel reflectance values, our analysis provides new insight relevant for surface water classification by specifically determining how pixel's reflectance correlations between sensors and biases vary depending on their lake position (i.e. is the pixel lake, shoreline, or land). Next, we compare water fraction classifications between sensors at these distinct lake positions. By establishing relationships between reflectance values and binary water classification in lakes and around lake margins, we determine whether lake fractions detected using unharmonized sensors (without empirical corrections between sensors) are congruent. Additionally, by analyzing both TOA and SR images, we determine if AC can influence the cross-satellite differences. Lastly, we relate the magnitude of satellite and AC water fraction discrepancies to both lake size and seasonal dynamism in surface water. We conclude with a discussion of implications of our findings and recommendations for accurately tracking surface water variability in Arctic lakes.

2. METHODS

2.1 Study Regions

We selected six hydrogeographic study regions in western Canada and Alaska with ranging, climates, riverine influences, vegetation types, permafrost extent, lake-ice dynamics and geomorphological characteristics. The common attribute for all of these regions was high lake density. While landscape diversity was not the focus of our study, the hydrogeographic regions presented a range of lake sizes and seasonal fractions that we associated with lake area discrepancies. To facilitate our ability to identify same-day, cloud-free Landsat 8 and Sentinel-2 image tiles in Google Earth Engine, we produced 24 smaller image target areas within our hydrogeographic regions. By isolating the analysis to smaller areas, we can more granularly track the reflectance differences and water fractions of individual scenes instead of composites with multiple images. Figure 1 depicts the main study regions, and the image target areas contained within them. For exact counts of coincident images within hydrogeographic regions and image target areas, see Table 1.

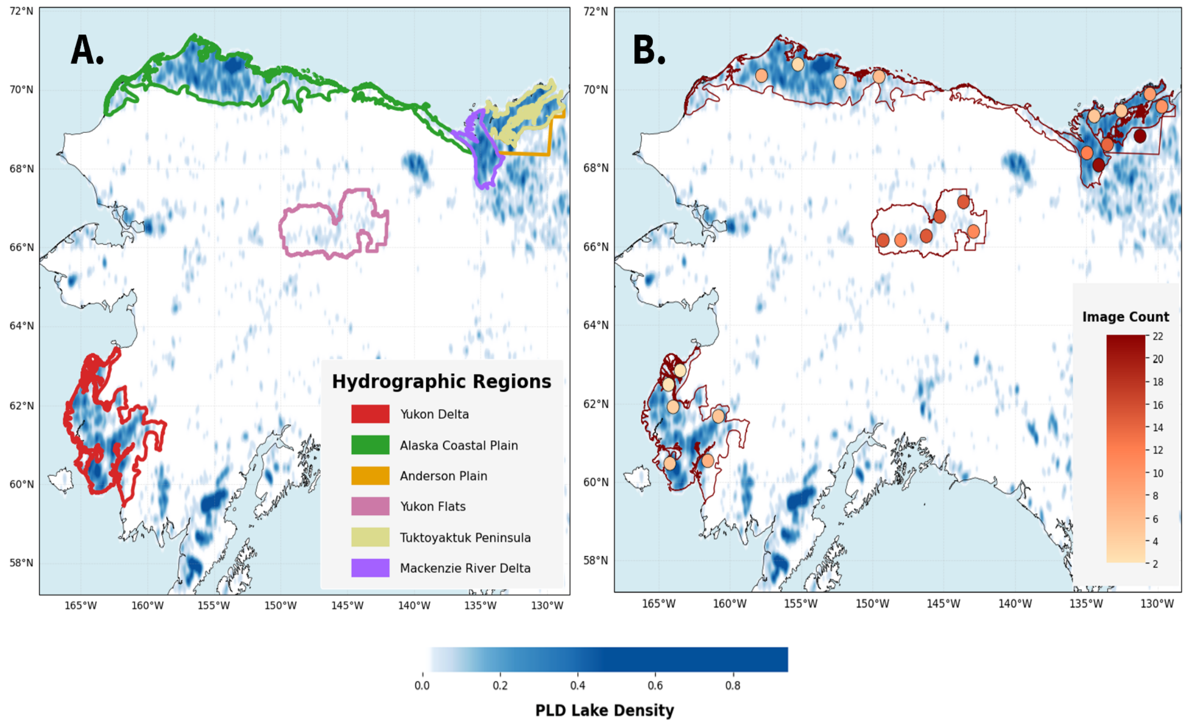

Figure 1: Hydrogeographic study regions and image target areas. Panel A shows the main hydrographic study regions in the Western Arctic. Meanwhile, Panel B shows the image target areas within the main study regions, the color-scale corresponds to the number of quality same-day Sentinel-2 and Landsat 8 image sets for the target area.

Hydrogeographic Region	Observation Count	Image Target Areas		Month	Observation Count
Alaska Coastal Plain	21	4		May	24
Anderson Plain	36	2		June	66
Mackenzie River Delta	33	2		July	62
Tuktoyaktuk Peninsula	29	4		August	41
Yukon Delta	25	6		September	33
Yukon Flats	82	6			
Total	226	24			

Table 1: Coincident image observation counts by hydrogeographic region and month

2.2 Input datasets

2.2.1 Image acquisition and pre-processing

Using Google Earth Engine, we downloaded all same-day Sentinel-2 and Landsat 8 image pairs with total cloud cover below 40% between May and September 2018-2024. We downloaded Sentinel-2 images processed to two levels TOA (Level-1C orthorectified top-of-atmosphere reflectance) and SR (Level-2A orthorectified atmospherically corrected surface reflectance). The Sentinel-2 A/B satellites carry the Multispectral Imager (MSI) sensor, which has 13 multispectral bands with 10-meter resolution for Blue, Green, Red, and NIR, but we specifically evaluate Band 3 (Green, 543-578nm) and Band 8 (NIR, 785-907nm) for NDWI mapping. The Sentinel-2 Surface Reflectance images we use are processed using the Sen2Cor algorithm, which corrects TOA data for atmospheric, terrain and cirrus cloud effects (Main-Knorn et al., 2017). To complete our image sets, we downloaded Landsat 8 (Level 2, Collection 2, Tier 1) images processed to both the TOA and SR levels. The Operational Land Imager (OLI) sensor aboard Landsat 8 contains 9 multispectral bands with 30-meter resolution. Similar to Sentinel-2, we limit our evaluation to Band 3 (Green, 530-590nm) and Band 5 (NIR, 850-880nm). The Landsat Surface Reflectance Code (LaSRC) algorithm adjusts Landsat 8 TOA images for aerosols and thin clouds (Vermote et al., 2018), but unlike Sen2Cor, LaSRC does not perform any terrain correction. None of the datasets – Landsat 8 SR/TOA or Sentinel-2 SR/TOA – perform correction for bidirectional reflectance effects. While bidirectional reflectance impacts can be meaningful in individual scene comparisons with disparate view angles (Zhang et al., 2018), analyzing a large number of scenes balances this effect (Roy et al., 2016). While the Sentinel-2 and Landsat 8 Green bands have similar spectral response functions, the NIR bands (MSI band 8, OLI band 5) vary more widely in wavelengths (Figure 2). The sensors' incongruent NIR bands is a focal point for our analysis.

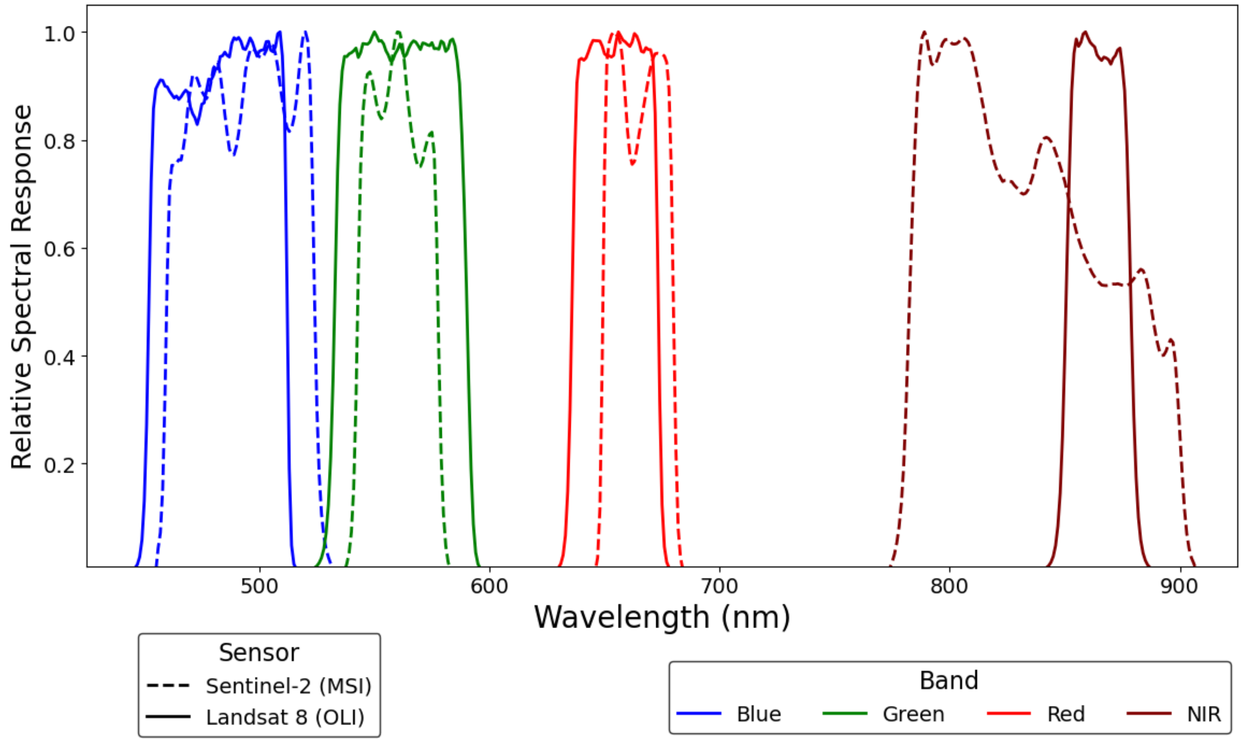

Figure 2: RGB and NIR relative spectral response functions for Sentinel-2 and Landsat 8. Line types indicate spectral response functions for Sentinel-2's MSI (dashed line) and Landsat 8's OLI (solid line) sensors. For simplicity we only plot Sentinel-2 A's relative spectral response, because Sentinel-2 B's sensor is very similar.

To account for slightly different imaging times between Sentinel-2 and Landsat 8, we produced a common set of masked values constrained by both satellites. We omitted pixels affected by opaque clouds, snow/ice, cirrus clouds, as identified by quality algorithms from NASA (Landsat 8) and ESA (Sentinel-2). The Sentinel-2 data mask was generated using the Cloud Probability dataset (excluding pixels with cloud probability greater than 70%), combined with pixels flagged in the Scene Classification Layer (SCL) for snow/ice, cirrus clouds, and cloud shadows. The Landsat 8 data mask was derived from the bitwise QA_PIXEL layer, where we eliminated the same four attributes. We then combined these Sentinel-2 and Landsat 8 masks into a unified common mask, capturing all invalid pixels from both datasets. To reduce speckle, we sieved out small mask features (fewer than 50 Landsat pixels or 0.045 km^2) and then dilated the remaining features by 0.5 km for a conservative mask extent. This processing ensured that only the highest-quality pixels from both sensors remained available for reflectance and water fraction comparisons.

Because our analysis included shoulder season images (May & September), lake ice remained even after applying Earth Engine quality filters. Due to their thermal inertia, lakes often maintain ice cover after snow melts on land, meaning in some early season images, lake surfaces disproportionately contained un-measured ice pixels. To address this, we eliminated image sets where more than 60% of the total buffered lake pixels were masked. From our initial catalog of 302 images sets, this additional filtering step removed 76 sets, leaving 226 sets meeting our criteria. Omitting images with comparatively few lake pixels improved the consistency between water fractions and did not qualitatively alter our results. Due to geographic variability in cloud-cover, latitude-dependent overlaps between Sentinel-2 and Landsat 8 swaths, and variations in lake ice duration, the number of quality image sets varied across hydrogeographic regions (Figure 1; Table 1) and months. May notably lacked quality images due to extensive ice cover, accounting for only 9% of observations. The lack of shoulder season observations underscores a source of uncertainty measuring seasonal lake fluctuations in the Arctic. Table 1 provides an account of spatial and temporal image coverage.

2.2.2 The Prior Lake Database (PLD)

We used the NASA Prior Lake Database (PLD) (Wang et al., 2025) to demarcate lake position within each image. PLD is a globally available database derived from over a dozen input data sources, primarily from the Circa-2015 UCLA Global Lake Database (Sheng et al., 2016). PLD's minimum lake size (0.01 km^2) is well within Landsat 8's size detection limits (Olthof & Fraser, 2024). Using PLD, we define four categories of lake position categories (Figure 3):

1. Lakes – within original PLD boundaries
2. Buffered Lakes – pixels within 60 meters of the PLD boundary
3. Land – pixels outside PLD buffered by 60 meters
4. Shoreline – pixels between PLD eroded by 60 meters and PLD buffered by 60 meters.

We also grouped PLD lake areas by size into four classes: smallest [0.01-0.05 km^2], small (0.05-0.5 km^2], medium (0.5-1 km^2], and large (> 1km^2). This categorization enabled us to assess how lake size influenced observed discrepancies in water fraction measurements between Sentinel-2 and Landsat 8 sensors and their AC processing levels.

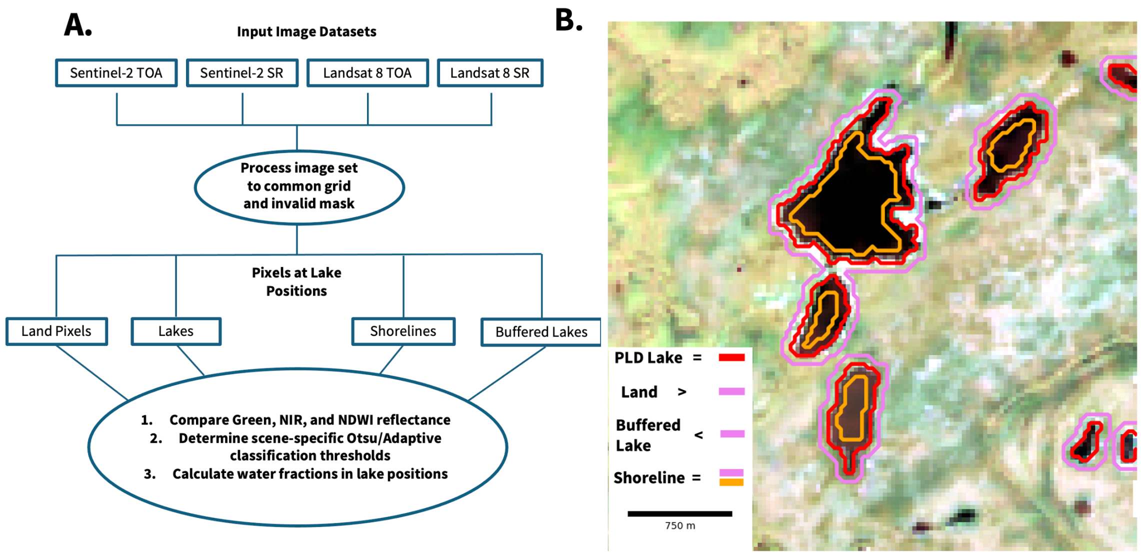

Figure 3: Visualizing the pre-processing workflow and PLD lake positions. Panel A shows a high-level overview of our image QAQC and processing workflow, where we apply a common mask to the image set and designate lake position according to PLD. Then, we conduct reflectance and water fraction comparisons for each lake position. Panel B shows demarcated lake positions with PLD boundaries over a Landsat 8 SR image from the Yukon Flats on July 1st in 2021. The colored contours illustrate how we demarcated land (outside the violet contour), lakes (inside the red contour), buffered lakes (inside the violet contour), and shorelines (between orange and violet contours).

2.2.3 GSWO and GLAD seasonal fractions

We used both the Global Surface Water Occurrence (GSWO) dataset (Pekel et al., 2016) and Global Land Analysis and Discovery (GLAD) dataset (Pickens et al., 2020) to investigate how differences between coincident Sentinel-2 and Landsat 8 images relate to seasonal variability in inundation extent. Specifically, we downloaded monthly binary surface water occurrence grids covering the years 1999-2020, the period common to both datasets. For June and August, we aggregated these monthly binary grids across all years, producing a monthly percent occurrence dataset (0-100%). We then applied an 80% occurrence threshold to generate June and August surface water grids, which we then constrained to our PLD buffered mask to restrict analysis to lake area. For each image target area, we calculated the net seasonal lake fraction as the difference between June and August lake areas relative to June lake areas.

2.3 Image analysis

2.3.1 Reflectance comparisons

Because Sentinel-2 and Landsat 8 have different spatial resolutions (10-meter vs. 30-meter), and the overlapping tiles are sometimes referenced in different UTM zones, we resampled images to align the pixels precisely for reflectance comparisons. There are established methods for matching Sentinel-2 and Landsat 8 to a common grid involving resampling one or both images (Chastain et al., 2019; Claverie et al., 2018; Flood, 2017; Runge & Grosse, 2019). To mitigate image and water fraction degradations that can occur with successive resampling (Lyons et al., 2013), we ensured both the Landsat 8 and Sentinel-2 images were both resampled only once into a predefined transformation matching our image target areas' extent and UTM zones.

We assessed the sensitivity of our results using two different grid resolutions: 30 meters and 60 meters. 30 meters is the most common resample grid resolution (Chastain et al., 2019; J. Chen & Zhu, 2022; Flood, 2017; Lima et al., 2025), but we also assessed 60 meters following Runge & Grosse, (2019) to ensure reflectance and water fraction differences were consistent at a coarse scale. Additionally, we tested three resampling algorithms – bilinear, cubic, and lanczos – since various resampling techniques can introduce artifacts, particularly around feature edges and no-data values (Claverie et al., 2018; Lima et al., 2025). We also calculated water fractions for the un-resampled images in their native resolution. The choice of resampling method did not qualitatively affect reflectance or water fraction outcomes. Unless otherwise reported, our results are bilinearly resampled to a 30-meter grid.

Using co-registered pixels on resampled grids, we tested agreement between both sensors (Sentinel-2's MSI versus Landsat 8's OLI) and AC (TOA versus SR) reflectance values in the Green, NIR, and NDWI bands. From each image, we randomly selected 10,000 pixels from distinct lake positions (land, lakes, shorelines, and buffered lakes) and plotted the collocated reflectance values. For each comparison, we calculated the Pearson correlation coefficient, the proportion of pixels biased above a 1:1 line, and performed linear regression using Reduced Major Axis (RMA) regression. RMA regression is particularly suitable for cross-sensor comparisons, because it assumes both the independent and dependent variables have errors with similar magnitude (Flood, 2017; Roy et al., 2016).

2.3.2 Surface water classification

On each image, we employed two similar NDWI histogram-based classification methods – Otsu thresholding (Steps 1-3) (Otsu, 1979) and Adaptive Thresholding (Steps 1-5) – over PLD's predefined buffered lake mask, an approach similar to other studies (Cooley et al., 2017, 2019; Donchyts et al., 2016; Levenson et al., 2025; Olthof et al., 2015). The classification processes involved the following steps:

1. **Generate an NDWI image** for the entire scene using equation 1, where ρ is a given band's reflectance.

$$NDWI = \frac{\rho_{Green} - \rho_{NIR}}{\rho_{Green} + \rho_{NIR}} \quad (eq. 1)$$

2. **Mask the NDWI image** to buffered lake areas, capturing both water and adjacent land pixels as distinct classes for the Otsu algorithm.

3. **Generate a bimodal histogram** with distinct land and water NDWI classes (see figure 3). Use this histogram to calculate the Otsu threshold (t) in equation 2 that minimizes weighted within-class variance $\sigma_\omega^2(t)$. This threshold is chosen such that it optimally separates the two classes based on their probabilities (weights ω):

$$\sigma_\omega^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \quad (eq. 2)$$

4. **For adaptive thresholding, determine the land limits (LL) and water limits (WL)** as NDWI values, associated with histogram counts (h), see figure 3 for a visualization. Heights for the LL and WL NDWI values are calculated using equations 3 and 4. The respective peaks for the land (h_l) and water (h_w) classes, and histogram's height at the Otsu threshold (h_t).

$$h_{ll} = h_l - 0.9(h_l - h_t) \quad (eq. 3)$$

$$h_{wl} = h_w - 0.9(h_w - h_t) \quad (eq. 4)$$

5. **Linearly interpolate intermediate NDWI** pixels between the Land and Water bounds using equation 5. For these intermediate values, we chose a 75% water probability threshold.

$$Water\ Probability = \frac{NDWI - LL}{WL - LL} \quad (eq. 5)$$

Because these algorithms dynamically adjust the NDWI threshold for each image, they are well-suited to TOA images and comparisons across atmospheric correction methods, where the overall scene brightness varies substantially (Donchyts et al., 2016). Furthermore, calculating the thresholds over the buffered lakes resolves the issue where histogram-based methods perform poorly in low water fractions (Cordeiro et al., 2021; Donchyts et al., 2016). Adaptive Thresholding, in particular, is effective at handling ambiguous or mixed pixels along shorelines by linearly interpolating between defined land and water limits (Olthof et al., 2015).

While simpler than machine learning and deep learning classification models, these histogram methods offer greater interpretability, which aligns well with our goal of comparing reflectance values across sensors. A key advantage of histogram-based classification is its transparent link between pixel values and classification outcomes—an interpretability often lost in more complex models. Since the choice between Otsu and Adaptive Thresholding did not meaningfully affect our results, we report only the Adaptive Thresholding water fractions for brevity.

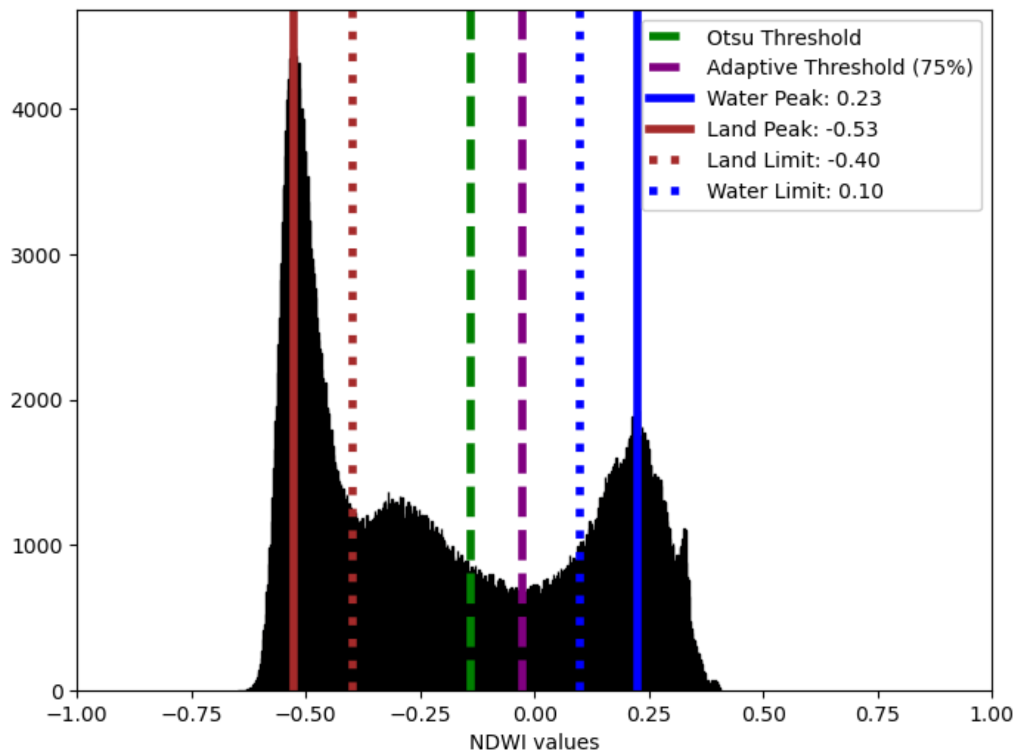

Figure 4: Example NDWI histogram for Otsu and Adaptive Classification. Illustrates how Otsu threshold (green dashed line) and Adaptive threshold (purple dashed line) differentiate the land and water classes. The intermediate NDWI values are between the land limit (dashed brown line) and water limit (dashed blue line), which are relative to the land and water peaks (solid lines). The 75% water probability threshold for intermediate NDWI values is linearly interpolated using (eq. 5).

2.4 Statistical tests

We established relationships between water fraction inconsistencies and lake position, lake size categories, and seasonal lake fraction using descriptive statistics. One important framing is water fractions differences in absolute versus relative terms. Because water fractions varied considerably between image target areas (between 4% and 52% lake area as measured by PLD) and lake positions (i.e., unbuffered lakes will have higher water fractions than shorelines), we also contextualized image discrepancies relative to the baseline water. Understanding lake area discrepancies relative to a landscape's lake fractions is especially important in a dynamic lake mapping context, because the percentage change is relative to baseline lake area. For a given constraint (e.g. shoreline, total landscape, or particular lake size category), we calculated a scene's water fraction (0-100%). Calculations for our area difference results are as follows:

$$Absolute\ Sensor\ Difference = LS8\% - S2\% \quad \text{(eq. 6)}$$

$$Absolute\ AC\ Difference = TOA\% - SR\% \quad \text{(eq. 7)}$$

$$Relative\ Sensor\ Difference = \frac{LS8\% - S2\%}{0.5(LS8\% + S2\%)} \times 100 \quad \text{(eq. 8)}$$

$$Relative\ AC\ Difference = \frac{TOA\% - SR\%}{0.5(TOA\% + SR\%)} \times 100 \quad \text{(eq. 9)}$$

Across all lake positions, we used one sample t-tests to determine whether the relative and absolute differences between satellites and AC methods were biased from zero. To determine if the lake position (e.g., shoreline), and lake size category (e.g., below 0.5 km^2) had outsized water fraction discrepancies, we used a one-way ANOVA with a post-hoc Tukey's Honest Significant Differences (HSD). When evaluating the relationship between an image target area's sensitivity to relative satellite differences and lake seasonal fraction, as measured by GLAD and GSWO, we used Ordinary Least Squares regression.

3. RESULTS

3.1 Comparing Sentinel-2 and Landsat 8 reflectance

3.1.1 Agreement between Sentinel-2 and Landsat 8 within AC processing

We observed strong overall correlations between Landsat 8 and Sentinel-2 reflectance values, especially for the NIR band (mean Pearson's R = 0.94 for TOA and R = 0.92 for SR). However, the Green band exhibited marginally weaker correlations (R = 0.87 for TOA and R = 0.83 for SR) (Figure 5 panels A and B). This overall strong correlation between the sensors aligns well with prior landscape-wide studies comparing Sentinel-2 and Landsat 8 reflectance (Chastain et al., 2019; Claverie et al., 2018; Runge & Grosse, 2019). Our analysis builds upon these findings by explicitly examining how lake positions (lake, shoreline, buffered lakes, and land) influence reflectance agreement. Notably, our results indicate that lake position does not substantially affect the strength of correlation between sensors. For instance, the mean Pearson's R values for the TOA image's NIR band were R = 0.92 over lakes and R = 0.90 over land. Image sets with unusually low correlations (R < 0.75) likely resulted from undetected cirrus clouds, not flagged by ESA or NASA's quality algorithms, or significant sun angle differences between paired satellite acquisitions (J. Chen & Zhu, 2022). For shorelines, 10% of Green, < 1% of NIR, and 0% of NDWI images in TOA data had R values below 0.75. The number of low correlation images was comparable in SR images, where 11% of Green, < 1% of NIR, and 6% of NDWI images had R values below 0.75.

While correlations between sensors were stable across lake positions, the reflectance biases between sensors varied substantially in the TOA data (Figure 5, panel C). Sentinel-2 TOA consistently measured higher NIR reflectance over lake surfaces in 87% of images, whereas it measured lower NIR reflectance over land in nearly all cases. This opposing NIR reflectance bias notably impacted the NDWI index across lake positions. Sentinel-2 TOA systematically recorded higher NDWI values along shorelines compared to Landsat 8 in 98% of images, highlighting its relatively lower NIR sensitivity in these transitional environments. Across all TOA images, the mean proportion of NDWI shoreline pixels biased higher for Sentinel-2 was 75%. Consequently, Sentinel-2's tendency to measure higher NDWI values along shorelines resulted in significantly larger lake areas and higher water fractions than Landsat 8.

Applying atmospheric correction substantially reduced systematic biases between sensors in both NIR reflectance and NDWI values. In SR images, there was not a consistent pattern where Sentinel-2 measured higher reflectance than Landsat 8 based on lake position (Figure 5, Panel D); however, independent of lake position, Sentinel-2's SR data measured lower NIR reflectance and higher NDWI than Landsat 8's SR data. The NDWI index remained higher for Sentinel-2 versus Landsat 8 along shorelines in 68% of SR images, a decrease compared to the 98% observed in TOA data. Still, this NDWI bias did not vary substantially across lake positions. Since the mean proportion of biased NDWI pixels was comparable between lakes, land, and shorelines, it did not systematically impact the separation between land and water classes. Consequently, SR data did not exhibit the directional bias in water fractions between Sentinel-2 and Landsat 8 that was apparent in TOA.

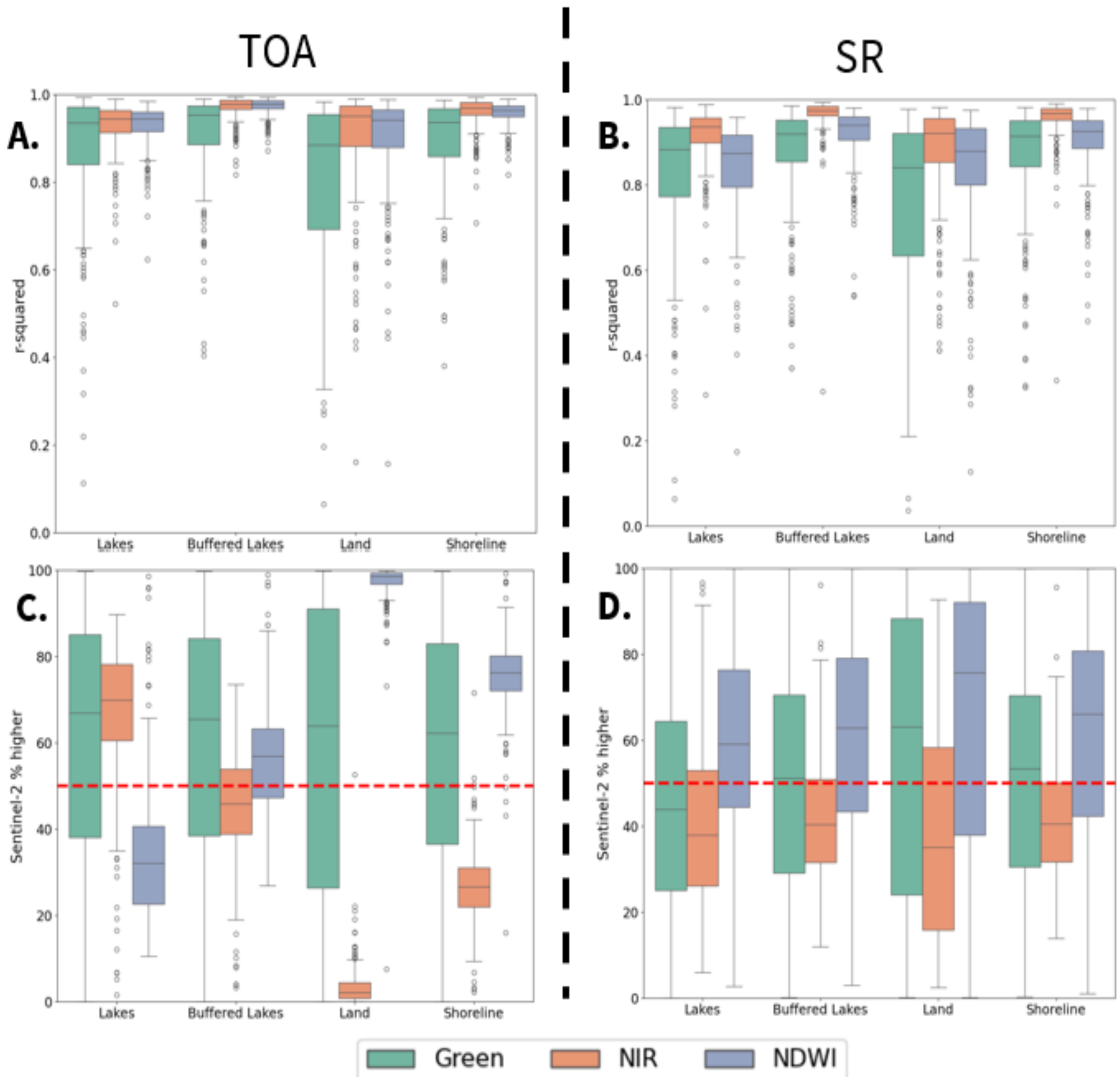

Figure 5: Correlation coefficients and percent of biased pixels between Sentinel-2 and Landsat 8 reflectance comparisons. The panels are split between SR and TOA data. The top row (panels A and B) shows the correlation coefficients between Sentinel-2 and Landsat 8 scene reflectance across lake positions colored by bands. The bottom row (panels C and D) shows the percentage of each scene's pixels where Sentinel-2 measured higher reflectance than Landsat 8 across bands and lake positions.

3.1.2 Agreement between TOA and SR within sensors

Along with lessening bias between sensors, AC also significantly reduced the reflectance in the Green and NIR bands (Figure 6). Correlations between TOA and SR data remained strong and consistent across lake positions for both sensors (Figure 6 panels A and B). Mean Pearson's

R values between TOA and SR data were higher in Green (Landsat 8 = 0.92, Sentinel-2 = 0.98) and NIR bands (Landsat 8 = 0.99, Sentinel-2= 0.99) compared to NDWI (Landsat 8 = 0.85, Sentinel-2 = 0.81). The lower NDWI agreement arises because AC disproportionately reduces Green reflectance compared to NIR reflectance over surface water (Kuhn et al., 2019), a pattern confirmed by our analysis.

For both Landsat 8 and Sentinel-2, 0% of SR images exhibited higher average Green reflectance over buffered lakes compared to their TOA counterparts. In the Green band, the average proportion of buffered lake pixels biased higher was < 0.1% for Landsat 8 and 0.3% for Sentinel-2. In contrast, the proportion of pixels biased lower in the NIR band was 6% for Landsat 8 and 39% for Sentinel-2 (Figure 6). This differential reduction in Green and NIR reflectance resulted in systematically lower NDWI values at lake shorelines for all SR images. Across the entire dataset, the mean proportion of shoreline pixels with higher NDWI values in SR data was 3% for Landsat8 and 6% for Sentinel-2. AC's impact in lowering NDWI values at shorelines is crucial as it contributes to underestimating lake classifications.

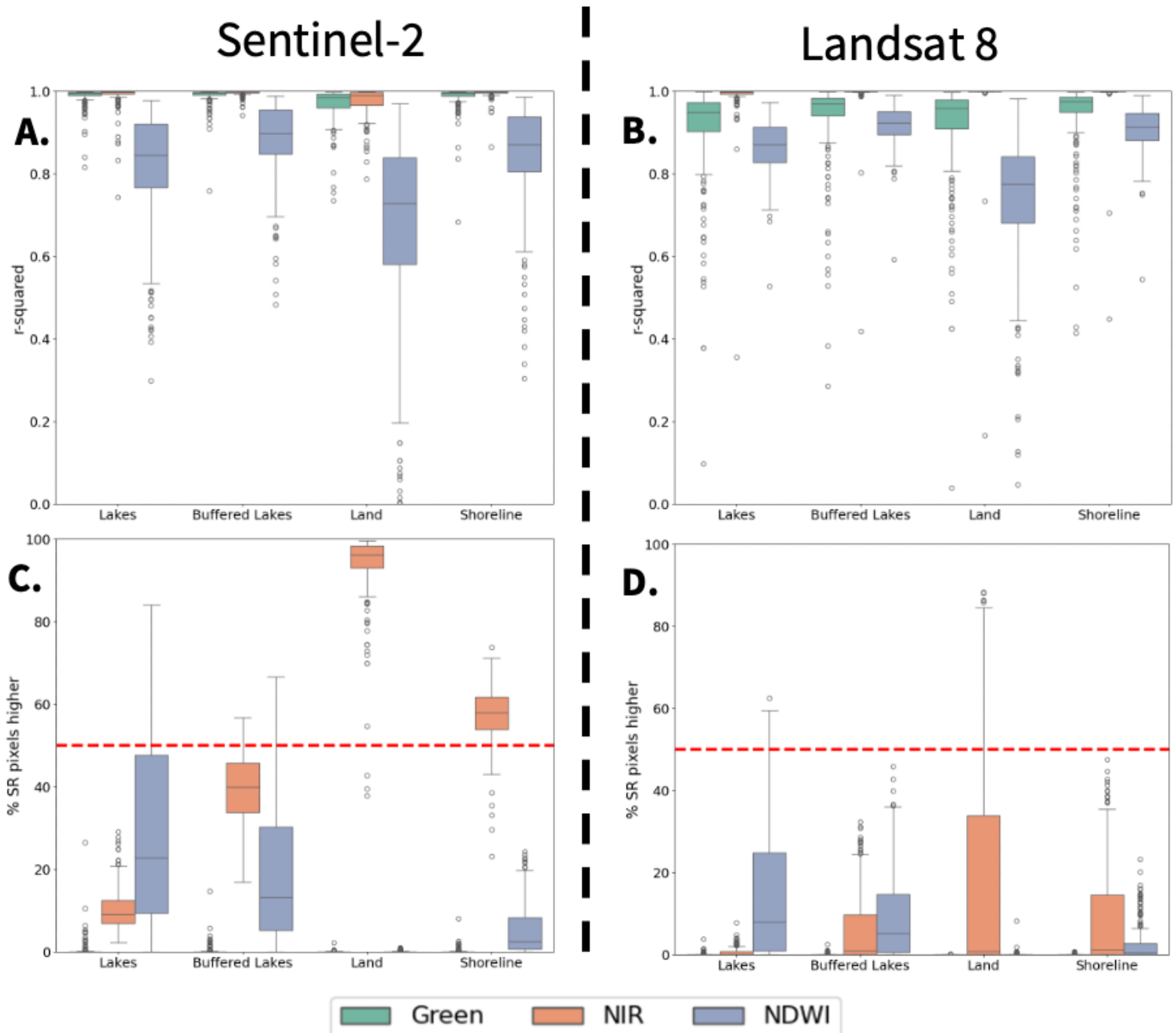

Figure 6: Correlation coefficients and percent of biased pixels between TOA and SR reflectance comparisons. The panels are split between Sentinel-2 and Landsat 8 data. The top row (panels A and B) shows the correlation coefficients between TOA and SR data for each sensor across lake positions colored by bands. The bottom row (panels C and D) shows the percentage of each scene's pixels where SR data measured higher reflectance than TOA data.

3.2 Sentinel-2 and Landsat 8 water fractions

3.2.1 Water fraction differences *in TOA images*

For TOA images, Sentinel-2 consistently measured higher water fractions using both Otsu and Adaptive water classifications compared to Landsat 8. For un-resampled images, mean landscape water fractions were 16.4% for Sentinel-2 and 14.3% for Landsat 8, while mean lake (PLD) water fractions were 82.5% and 74.8% respectively. The lake water fractions were considerably lower than 100%, because PLD's lake geometries appear to broadly err on the side of overestimating lake extent rather than underestimating. Resampling the images to a common, coarser grid reduced—but did not eliminate—these cross-sensor differences. For instance, when both sensors were bilinearly resampled to a 30-meter grid, Sentinel-2's landscape water fraction was 14.5% (75.5% over PLD lakes), compared to Landsat 8's 13.4% (70.5% over PLD lakes). Even at a 30-meter resolution and when restricting analyses exclusively to unbuffered PLD lake areas—lake sizes clearly within Landsat 8's detection capabilities (Olthof & Fraser, 2024; Wang et al., 2025)—differences remained statistically significant.

Although the nominal percentage differences in landscape water fractions appear relatively small, their statistical significance underscores a notable inter-sensor bias. In dynamic lake mapping contexts, such differences should also be evaluated relative to existing lake and shoreline areas, as lake area changes are inherently proportional to baseline lake extents. We find that Landsat 8 measured 5% lower mean water fractions within the PLD mask for images resampled to a 30-meter grid (bilinear interpolation). Evaluating lake fraction discrepancies relative to the baseline areas highlights a profound impact, particularly along shorelines. Shoreline positions disproportionately influence TOA water fraction differences (Figure 8, panel B) due to their high proportion of mixed pixels. Specifically, Landsat 8 underestimated water fractions by an average relative difference of 7.7% for buffered lakes (IQR=5.6%), and 19.3% for shorelines (IQR=11.6%). In relative terms, Landsat 8's TOA underestimation along shorelines was disproportionate to other lake positions, when compared to the total landscape, lakes, and buffered lakes (HSD p-values < 0.01). These results emphasize that shoreline uncertainty is critical for accurately interpreting lake dynamics in multi-sensor studies.

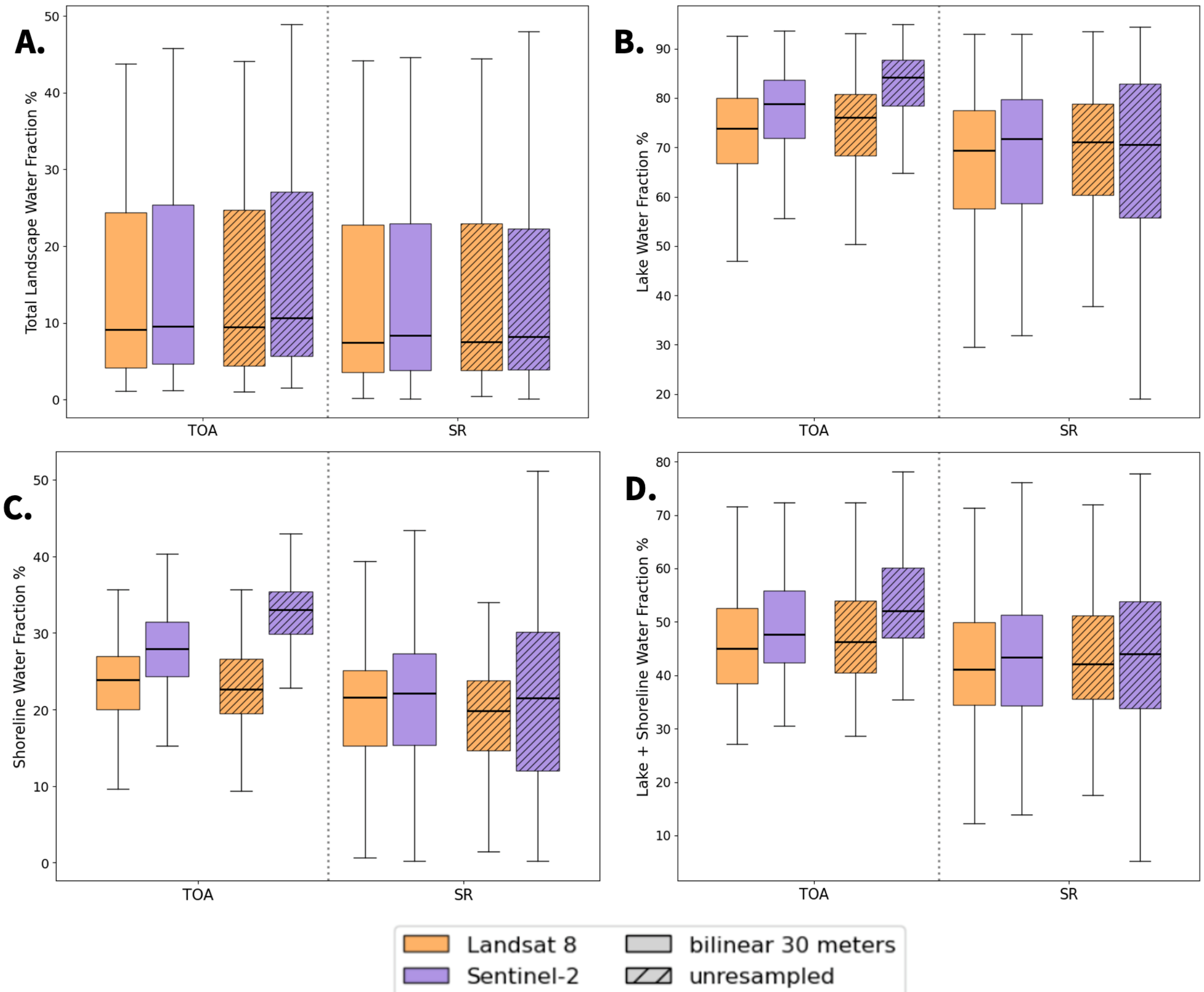

Figure 7: Sentinel-2 and Landsat 8 water fractions in lake position masks by atmospheric correction level and resampling method. Water fractions across all 226 scenes are reported as the percentage of water pixels within a given lake position's mask. Panel A corresponds to total scene water fractions (i.e., all valid pixels), while Panel B, C, and D correspond to PLD lakes, shorelines, and buffered lakes respectively. The grey dashed line separates TOA data from SR data within the panels. The y-axes have different scales, because the water fractions vary substantially with different lake positions.

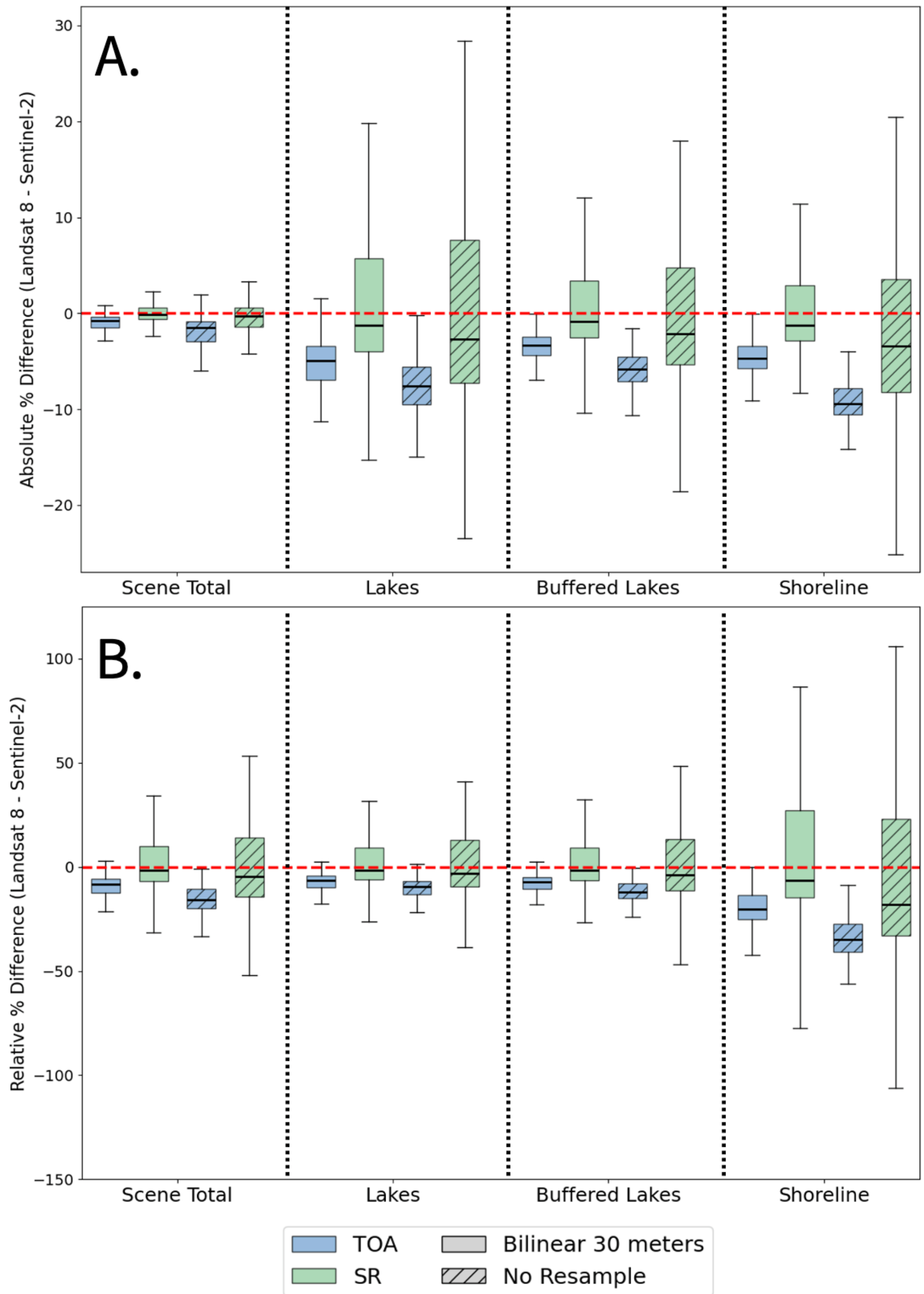

Figure 8: Relative and absolute differences between Sentinel-2 and Landsat 8 water fractions across lake positions, atmospheric correction, and resampling methods. Panel A shows absolute sensor differences between Landsat 8 and Sentinel-2 water fractions with Panel B showing relative differences. The red horizontal line marks zero. Outliers beyond 1.5 X IQR are not shown. Note, the relative and absolute plots have different y-axis scales.

3.2.2 AC's impact on water fraction discrepancies

Processing images to surface reflectance mitigated the mean bias where Landsat 8 measured lower water fractions relative to Sentinel-2. However, this process simultaneously introduced greater variability among individual observations and resulted in poorer overall agreement between sensors (Figure 8). While SR's mean bias between sensors was lower, relative difference of -7.9% on the total landscape, the inter-quartile range was much wider at +-16.5%, compared to TOA at +-7.3%, and the IQR was amplified in other lake positions, particularly on shorelines (Figure 9). Both sensors systematically underestimated water fractions with SR data compared to TOA, but AC's impact was greater with Sentinel-2. The mean relative difference between Sentinel-2 TOA and SR was 20.8% for the total landscape and 18.8% over lakes. Meanwhile, Landsat 8's mean relative differences were 11.4% on the total landscape and 10.7% on lakes. Interestingly, the relative differences for AC were greater over the total landscape than for lakes. A likely explanation for the wider total landscape differences is AC's tendency to omit many small water features on the landscape, which would be outside of PLD. The SR data's systematic underestimation of water fractions was statistically significant (t-test $p < 0.01$) for all resampling methods and landscape positions in both sensors. SR images were particularly prone to omitting smaller water bodies and marginal lake edges, leading to systematic underestimation of total lake fractions. Similar to cross sensor discrepancies, the impact of AC was greatest along lake shorelines. For both Landsat 8 and Sentinel-2, SR data's relative low water bias along shorelines was disproportionate to other lake positions, (HSD p-values < 0.01) when compared to the total landscape, lakes, and buffered lakes. Thus, while atmospheric correction effectively reduces systematic reflectance and area biases between Sentinel-2 and Landsat 8, it introduces its own methodological sensitivities, namely increased variability and lake-edge detection uncertainties.

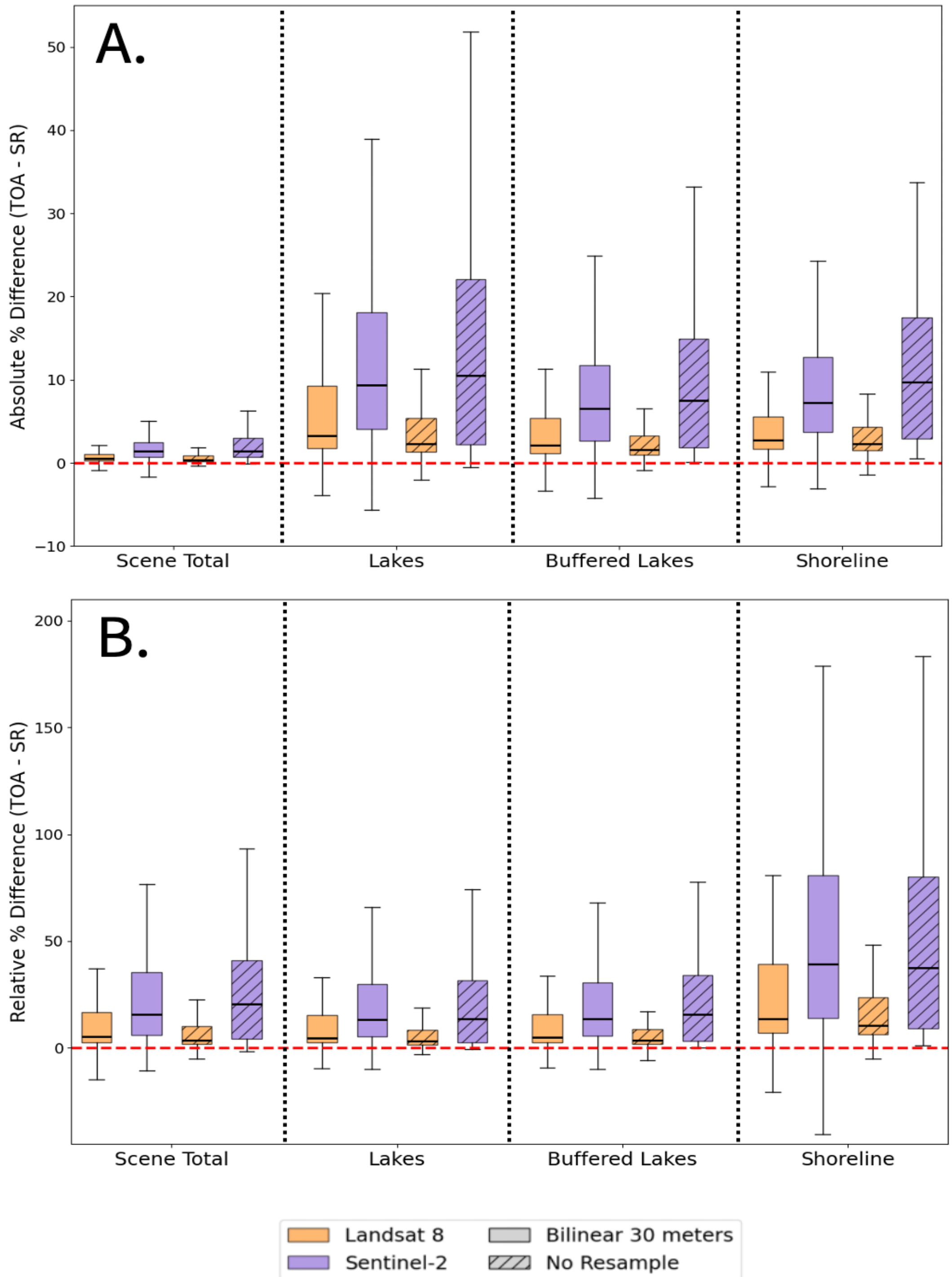

Figure 9: Relative and absolute differences between TOA and SR water fractions across lake positions, sensors, and resampling methods. Panel A shows the absolute differences between TOA and SR water fractions with Panel B showing the relative differences. The horizontal red line marks zero (i.e., no difference). Outliers beyond 1.5 X IQR are not shown. Note, the relative and absolute plots have different y-axes.

Visually inspecting the true-color and NDWI images corroborates our quantitative findings that atmospheric correction reduces NDWI values for both Sentinel-2 and Landsat 8 muddling water classifications. The impacts of atmospheric correction are most pronounced over mixed shoreline pixels and when measuring small lakes. SR classifications frequently missed small lakes, even those large enough to be include in PLD (Figure 10, panel A).

Compared to reflectance differences introduced by atmospheric correction, the reflectance differences between Sentinel-2 and Landsat 8 TOA data were more subtle, but still influential. Cross-sensor reflectance variations notably influenced the intensity of NDWI values at lake margins. Higher shoreline NDWI intensity led to larger estimated lake areas in Sentinel-2 TOA data, despite Landsat 8 exhibiting higher NDWI values at the lake center (Figure 10, panel B).

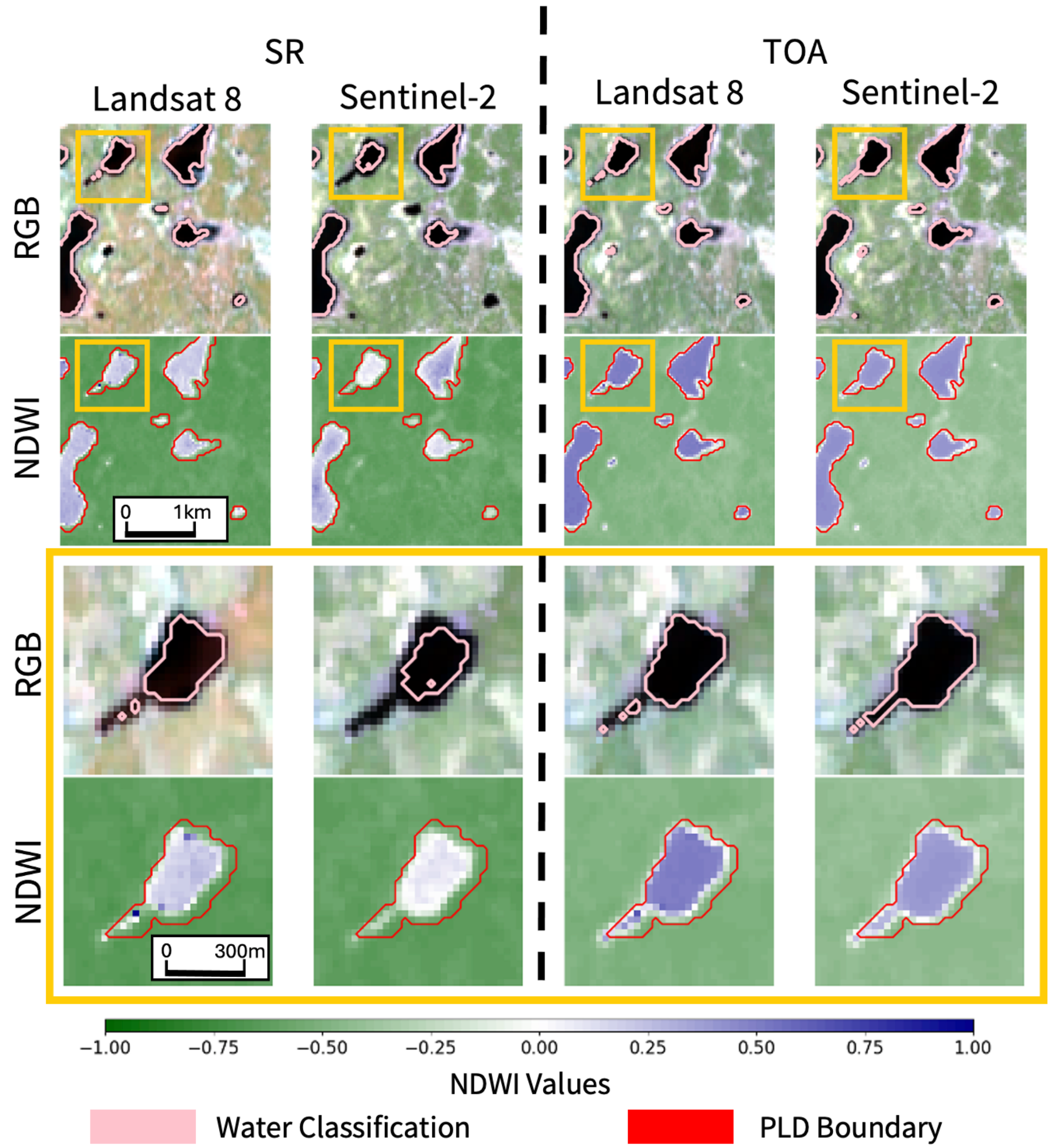

Figure 10: Visualizing RGB images, NDWI values, and lake extents. Example imagery from the Anderson Plain hydrogeographic region on June 16th, 2021, each image in the set is bilinearly resampled to 30 meters. TOA and SR images for each sensor are split by the dashed line with Landsat 8 in the right sub-columns and Sentinel-2 in the left sub-columns. Rows 1 and 3 show the RGB images with the Adaptive lake classification (pink contours), while rows 2 and 4 show the NDWI images with PLD boundaries (red contours). Panel A illustrates how SR imagery omits small lakes across the landscape, even those large enough to be included in PLD. Panel B zooms to the 1km-by-1km box in Panel A to highlight the pattern of shoreline omission by SR data and by Landsat 8 compared to Sentinel-2.

3.3 Lake size and seasonal fraction explain regional impacts

The relative magnitude of both sensor and AC water fraction discrepancies varied substantially throughout our hydrogeographic study regions. For example, Landsat 8's mean relative bias below Sentinel-2 for TOA data was only -4.0% (IQR=2.7%) in the Tuktoyaktuk Peninsula, but it was -9.4% (IQR=5.9%) in the Yukon Flats. This highlights that cross-sensor uncertainty is more impactful in some regions compared to others (Figure 10). Similar to cross-sensor differences, AC's impact on relative water fractions varied widely across regions. For example, the mean relative difference between TOA and SR data in the Mackenzie River Delta was 11.2% for Landsat 8 and 13.0% for Sentinel-2. On the other hand, the Yukon Delta had 22.6% (Landsat 8) and 18.2% (Sentinel-2) relative AC differences. Regions with high AC water fraction discrepancies also exhibited high cross-sensor water fraction discrepancies (Figure 10). This fact suggests that while dynamic surface water mapping uncertainty is not uniform across the Arctic, regions susceptible to one aspect of methodological uncertainty will be susceptible to others. While multiple factors could explain behavior across regions, outsized shoreline impacts (Figures 8 and 9) indicate that morphology plays an important role. Regional lake size, shoreline complexities, and adjacent wetlands with shallow water and emergent vegetation thus could all contribute to regional differences in methodological sensitivity. Comprehensive analysis with each of these morphological causes is outside the scope of our study, but we explore lake size and seasonal fraction's influence.

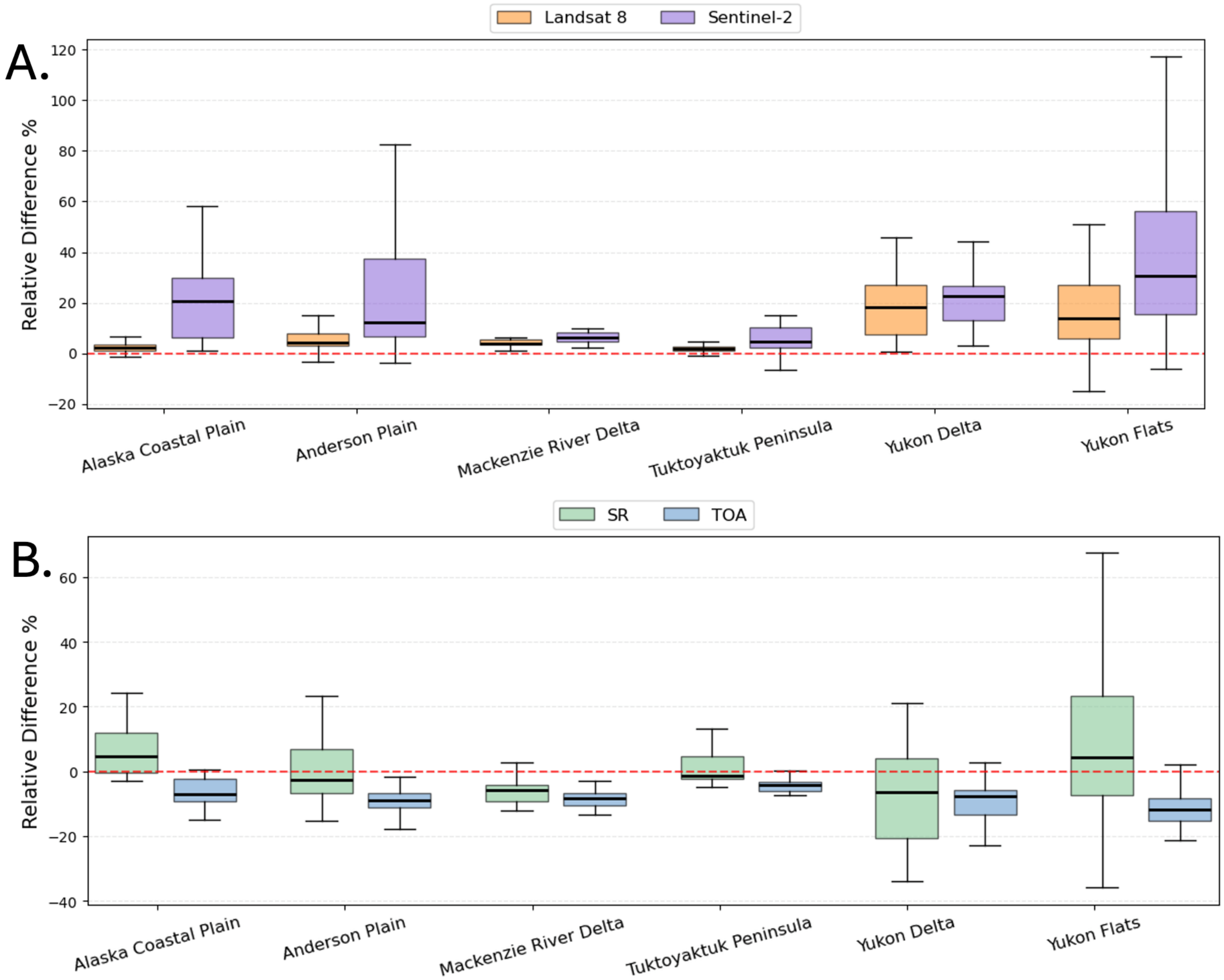

Figure 11: Regional impact on cross-sensor and AC discrepancies. Shows the cross-sensor and AC water fraction biases in relative terms grouped across hydrogeographic regions. Panel A shows relative bias between TOA and SR data colored by sensor, while Panel B shows relative bias between Landsat 8 and Sentinel-2 colored by AC processing level.

3.3.1 Size influences AC and sensor discrepancies

We found that the relative magnitude of both cross-sensor and AC water fraction differences increased as lake size decreased (Figure 11). For each coincident scene, we calculated aggregate water fraction discrepancies in specific lake size categories within PLD lakes (un-buffered). For resampled (bilinear 30 m) TOA images, the mean relative sensor difference was 19.2% (IQR=14.5%) over the smallest PLD lakes, but only 3.6% (IQR=2.9%) over lakes above 1 km^2. When comparing relative sensor differences in TOA data, the smallest lakes (0.01-0.05 km^2) had significantly outsized biases (HSD p-values < 0.01). As with lake positions, the cross-sensor differences were not directionally biased for SR data in all lake size categories (t-test p values > 0.01); however, the scene-to-scene variability – standard deviation in relative differences – was extreme for small lakes (σ= 77.2%) and high for large lakes (σ= 31.6%). This underscores the elevated image-to-image disagreement that AC can introduce between different sensors.

When comparing Landsat 8 and Sentinel-2's atmospheric correction impacts, we found that SR water fraction's bias lower relative to TOA was outsized in the smallest lakes (Figure 13). Though all lake size categories had significant bias (t-test p-values < 0.01), the bias for the smallest lakes (0.01-0.05 km^2) was outsized compared to other size categories (HSD p-values < 0.01). For Landsat 8, a scene's relative differences between TOA and SR data had a mean of 45.0% (IQR=38.8%) for the smallest lakes and 6.0% (IQR=3.9%) for the large lakes. The effect was similar for Sentinel-2, where the mean relative AC difference was 74.3% (IQR=104.9%) for the smallest lakes and 11.7% (IQR=12.1%) for the largest lakes. These large differences highlight that AC is a critical methodological consideration when evaluating relative change in small lakes. In addition to higher mean AC differences, smaller lakes had more scene-to-scene variability, where TOA and SR images produced wide ranging relative differences for the same sensor. Landsat 8's standard deviation in relative AC differences was 52% over the smallest lakes while Sentinel-2's was 57%.

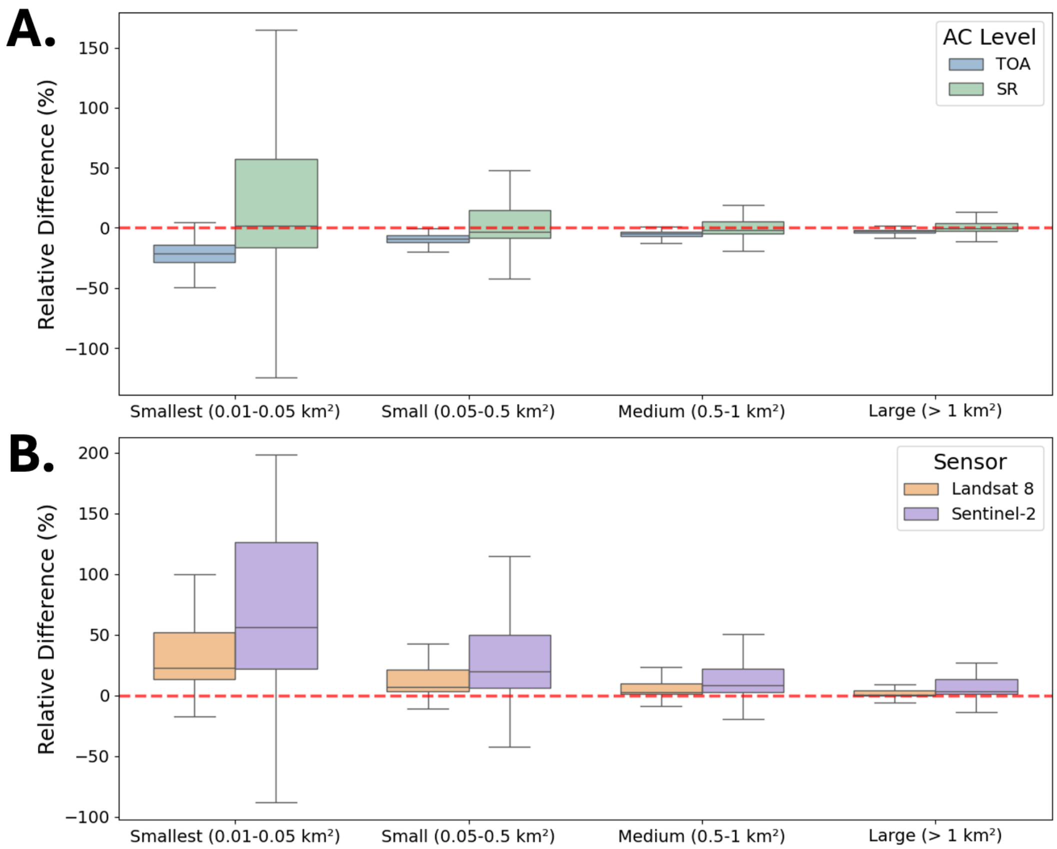

Figure 12: Lake size influence on cross-sensor and AC differences. Panel A shows how the relative cross-sensor differences varied by lake size categories and AC processing. Panel B shows how the relative cross-AC differences varied by lake size and sensor.

3.3.2 Seasonal associations with sensor discrepancies

We found a strong positive correlation between seasonal lake fractions as reported by GSWO and GLAD, and our observations of disagreement between Sentinel-2 and Landsat 8's TOA data (Figure 15, panel A) ($p < 0.01$, R=0.46 for GLAD, and R=0.57 for GSWO). Both GSWO and GLAD use aggregated Landsat 5, 7, and 8 TOA imagery (Pekel et al., 2016; Pickens et al., 2020). Interestingly, the seasonal lake fractions (i.e., June to August changes) were surprisingly high in both datasets. For GLAD, 9 of the 24 image target areas had seasonal lake fractions in excess of 20%, and 12 of 24 for GSWO; however, the high fraction of seasonal water could be a data artifact. In places like the Yukon Flats with small lakes and complex morphologies, GSWO and GLAD seasonality estimates were especially dynamic. Given the outsized sensitivity of small lakes to differences between sensors, there are questions about whether this extreme seasonality reflects hydrology, or if it may result from artifacts from combining multiple sensors.

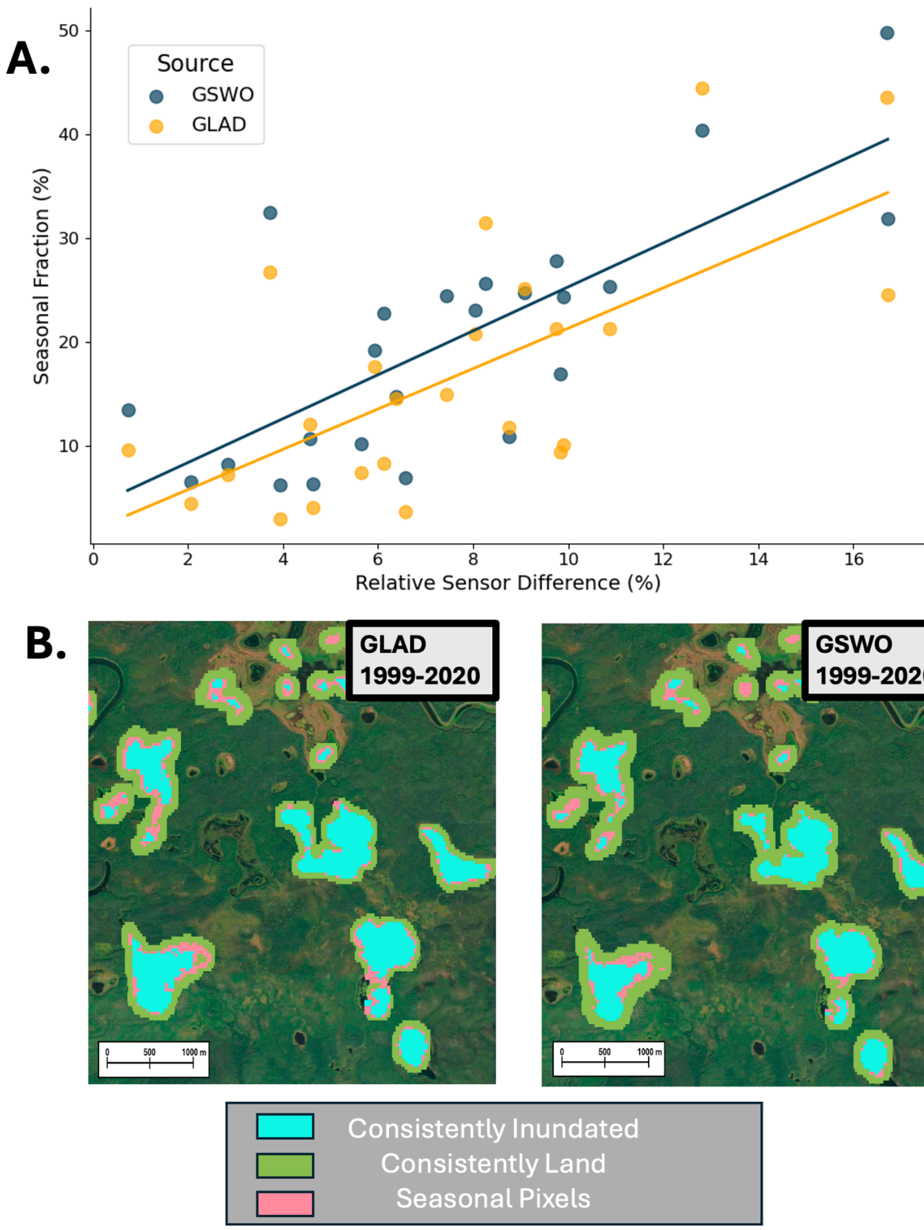

Figure 13: Associations between image target area's GSWO & GLAD seasonal water fractions and magnitude of cross-sensor discrepancies in coincidence scenes. Panel A shows the linear regression where the mean relative cross-sensor differences for PLD buffered 60 meters are plotted against the aggregate GLAD and GSWO seasonal lake fractions PLD buffered 60 meters. Panel B shows an example of PLD buffered 60-meter seasonal pixels as measured by GSWO and GLAD in the Yukon Flats.

4. DISCUSSION

We investigate two under-researched methodological sensitivities in Arctic lake mapping, and found they could potentially influence lake trend studies, if left unaddressed. Our research uniquely quantified how lake position dictates cross-sensor and AC reflectance biases, which thus translate to lake classification differences. In TOA images, we found that Sentinel-2's NIR bias relative to Landsat 8 depended on lake position, which produced higher NDWI values at lake margins. This produced larger lake extents for Sentinel-2 TOA relative to Landsat 8. In accordance with previous findings, atmospheric correction substantially lowered both sensor's reflectance (Kuhn et al., 2019), but the Green band was lowered more than NIR. This caused lower NDWI values at lake margins, which produced substantially smaller lake extents for SR data compared to TOA.

4.1 Implications for dynamic lake mapping

Our findings from this analysis have some notable implications for dynamic lake mapping. First, because atmospheric correction reduces the Green band's reflectance more than NIR over surface water, it leads to lower NDWI values and often omits lake margins. While atmospheric correction on average lessened biases between Sentinel-2 and Landsat 8's water fractions, image to image variability was substantially higher. This result thus suggests that studies should consider using TOA data whenever analyses' goal is to avoid omission errors, particularly over small waterbodies. Many northern regions contain an outsized proportion of these small water bodies (Muster et al., 2019). Beyond the Arctic, omission of small water bodies can be a persistent problem irrespective of the classification algorithm (Cordeiro et al., 2021). The strong impact of atmospheric correction on surface water fractions also means that switching AC methods could hide or inflate lake area trends. We thus suggest lake area trend analyses should use a consistent atmospheric correction algorithm throughout the period of record.

Our results also emphasize the importance of sensitivity analysis between satellites whenever studies fuse multiple sensors. We found that Landsat 8 consistently measured lower water fractions than Sentinel-2, particularly along mixed shoreline pixels. Fitting with previous research on the impact of spatial resolution (Mullen et al., 2023; Muster et al., 2013), Sentinel-2's finer 10-meter pixels explain a portion of – but not all – the water fraction discrepancies.

Significant water fraction differences between coarse resampled images over PLD's lake mask, water features with areas well within Landsat 8's detection limits, suggests that a given band's (particularly NIR's) spectral responses likely also have a meaningful influence. The discrepancies between Landsat 8 and Sentinel-2's relative water fractions were disproportionately large along shorelines and small lakes. Because lake change happens along theses uncertain margins, studies fusing multiple satellites should place more emphasis on explicit shoreline accuracy comparisons between sensors.

Failing to properly account for different sensor behavior in mixed pixels could artificially inflate or hide lake area trends, as the composition of satellites changes across years (e.g., adding Landsat 8). This concept has already been demonstrated in NDVI trends (Guay et al., 2014; Holden & Woodcock, 2016); however, we could not find any studies applying spectral corrections for lake area change. Not applying spectral corrections between sensors with stark differences could explain some trend disagreements in Arctic regions. Older studies (pre-2015) could be especially problematic because the image timeseries are extremely sparse, typically less than five observations over decades (Webb & Liljedahl, 2023; Supplemental Information). This means many Arctic lake area trend studies are reliant a small number of images from first-generation sensors, and a few subsequent observations from modern systems. Given our mean relative PLD lake difference of 7% between coincident Sentinel-2 and Landsat 8 TOA imagery, trends could be materially amplified or completely hidden depending on sensor choice over decades.

While the sparse temporal sampling of our matched image sets meant we did not conduct any explicit timeseries analysis, we did find strong associations between regions with high seasonal lake fractions (as measured by GSWO/GLAD) and cross-sensor lake discrepancies. We thus interpret a large portion of this seasonality as noise caused by differences in satellite (Landsat 5, 7, or 8) behavior at lake margins. The NIR band for Landsat 7 has different spectral response function from Landsat 8 with closer similarity and stronger reflectance agreement to Sentinel-2 (band 8) (Chastain et al., 2019). In the Arctic, this noise between satellites could be amplified by the paucity of annual lake observations, because a given monthly composite (from GSWO/GLAD) may only have one satellite. A subsequent monthly composite with a different satellite, could misinterpret sensor change as lake area change.

This result is important for dynamic water mapping due to the large number of studies that rely on GSWO, GLAD, and or other similar sensor fusion approaches for analyses of small glacial lakes (Song et al., 2025), seasonal water at high latitudes (Pickens et al., 2022), and Arctic lake change (Y. Chen et al., 2023). Notably, both GSWO and GLAD have especially low accuracy in seasonally inundated pixels; Users:44%, Producers:73% for GLAD, and GSWO Users:17.4%, Producers:36.3%. (Pickens et al., 2020) For GSWO, Pekel et al. (2016) noted the substantially lower omission accuracy in seasonally inundated pixels across satellites (Landsat 5 = 74.9%, Landsat 7 = 73.8%, and Landsat 8 = 77.4%). Because small lakes have a higher proportion of shoreline pixels, noise between sensors is elevated in these water bodies. A recent study utilizing GSWO highlighted the outsized variability of small lakes (Pi et al., 2022), yet our results suggest differences in shoreline behavior between sensors (even different Landsat sensors) could inflate seasonality detection, especially when frequent ice or cloud cover makes monthly composites reliant on a single sensor's observation. GLAD built separate bagged classification trees for each sensor, while GSWO used the same expert systems classifier on all sensors. In theory, separate models for each sensor could mitigate these differences; however, we still found anomalously high seasonality in GLAD data. Our findings thus indicate that applying established cross-sensor empirical corrections (Chastain et al., 2019; Claverie et al., 2018; Holden & Woodcock, 2016; Roy et al., 2016) to account for band's spectral responses could improve consistency between lake area trends and seasonality estimates.

4.2 Future work and limitations

As this study provides helpful insights for two potentially important but relatively understudied methodological sensitivities for dynamic surface water mapping, there are many opportunities for follow-on research. While our findings are unlikely to be unique to the Arctic, testing coincident shoreline discrepancies in imagery globally is worthwhile. In addition to expanding the geographic reach, there are three questions ripe for exploration: 1. Does using an aquatic-specific atmospheric correction algorithms improve lake classification?-2. Do differences in spectral response functions influence lake classification from other Landsat constellation sensors? 3. Can a more sophisticated classification algorithm incorporating multiple features such as SWIR bands, reduce discrepancies satellite-derived lake areas?

While we found that AC increased sensor disagreement and omitted water fractions, particularly for small lakes, Sen2Cor and LaSCR are not designed for aquatic contexts. We used them, because they are analysis ready in Earth Engine and therefore are a likely first choice for remote sensing hydrologists, who generally do not consider the nuances of AC methods. Additionally, our choice was determined by the fact that aquatic AC algorithms can struggle in small waterbodies with bright shorelines due to adjacency effects (Kuhn et al., 2019), where brightly reflective shorelines cause AC over-correction on water. Nevertheless, there are a variety of atmospheric correction algorithms designed for satellite-derived water quality observations with their own trade-offs. (Pahlevan et al., 2021) Whether modern aquatic AC algorithms improve surface water classification accuracy is an unanswered question.

We compared two sensors as a proof-of-concept to illustrate how differences in band spectral response functions translate into differences in surface water mapping. Given the exponential growth of Earth observation satellites, we believe our findings warrant similar analyses (i.e., coincident water fraction comparisons) with other sensors. Within the context of Landsat 8 and Sentinel-2, the Harmonized Landsat Sentinel-2 (HLS) dataset is an exciting prospect. Even though HLS's empirical corrections are robust for general landscape applications (Claverie et al., 2018), determining if these empirical relationships hold true for lakes and shorelines would enhance the validation HLS for dynamic surface water mapping. Additionally, future work should examine coincident image agreement within the Landsat constellation (Landsat 5, Landsat 7, Landsat 8, and Landsat 9), explicitly around lakes and shorelines. This analysis would build more confidence in the ability fused datasets like GSWO and GLAD to accurately measure dynamic surface water change at the margins.

The advantage of our Adaptive-Otsu classification is the interpretable relationships between reflectance, NDWI, and water fraction outcomes, but we did not examine the effect of machine learning algorithms with more input features. Still, NDWI is often a dominant feature in machine learning algorithms (Pickens et al., 2020), and Otsu thresholding over a prior buffered water mask performs reasonably compared to more sophisticated algorithms (Cordeiro et al., 2021). Many other studies combine NDWI with SWIR indices, jointly improving lake mapping accuracy (Janiec et al., 2023; Sui et al., 2022). While Sentinel-2 and Landsat 8's SWIR band are similar resulting in stronger correlations than their NIR bands (Chastain et al., 2019), other sensors can exhibit significant differences in SWIR bands. For example, Band 5 in Landsat 7

(1550-1750 nm) versus its counterpart Band 6 in Landsat 8 (1570-1650 nm) present notable biases (Roy et al., 2016). Explicitly evaluating lake and shoreline behavior among sensors with contrasting SWIR bands warrants further investigation. It remains to be determined whether band discrepancies' impact diminishes when surface water classifications incorporate multiple spectral features.

5. CONCLUSION

This study analyzed the impact of sensor resolution, spectral band characteristics and atmospheric correction on dynamic surface water mapping in the Western Arctic. Building on previous reflectance comparison approaches (Chastain et al., 2019; Claverie et al., 2018; Holden & Woodcock, 2016), we explicitly investigated the how reflectance agreement shifts relative to lake position. We find these reflectance changes can significantly influence water classifications on lake margins, where most dynamic surface water change occurs. Even accounting for the spatial resolution differences between Sentinel-2 and Landsat 8 through resampling and restricting analysis to lakes defined by the Prior Lake Database (PLD), Sentinel-2 consistently reported significantly higher lake water fractions. Additionally, SR data measured much lower water fractions than TOA data for both Sentinel-2 and Landsat 8.

Due to the systematic tendency of SR data to underestimate surface water, particularly along transitional regions like shorelines and smaller lakes, we recommend utilizing TOA data, apart from rare cases where the analysis goal requires more conservative (i.e., high omission) estimates of surface water. These biases— both Sentinel-2 versus Landsat 8 and SR versus TOA—were particularly pronounced in mixed pixels where uncertainty is greatest. This issue is especially pressing since meaningful surface water change occurs within these uncertain margins. While understanding discrepancies between sensors are less critical for static maps (i.e., not assessing relative change over time), the magnitude of observed discrepancies becomes highly consequential when monitoring seasonal or long-term changes in surface water extent. Consequently, studies integrating data from multiple satellite sensors could be more robust by performing more cross-sensor sensitivity analyses, specifically targeting mixed pixels in coincident images.

As the constellation of Earth observation satellites expands, there will be increasing opportunities—and an increasing necessity—to cross-validate and harmonize data across sensors. Initial efforts, such as the Harmonized Landsat-Sentinel (HLS) initiative, have already made outstanding strides in aligning reflectance across diverse landscapes and spectral ranges. Nonetheless, scientific studies targeting specific environmental features must carefully assess how variations in sensor characteristics and processing influence their unique classification problems, both aquatically (Lima et al., 2025; Maciel et al., 2023; Pahlevan et al., 2021), and terrestrially (Bayle et al., 2024; Ghasempour et al., 2023). Because surface water is dynamic at sub-seasonal timescales and change occurs at uncertain margins, involving more exhaustive cross-sensor calibration when fusing datasets may help lessen disagreements between sensors and more accurately quantify surface water variability.

BIBLIOGRAPHY

Andresen, C. G., & Lougheed, V. L. (2015). Disappearing Arctic tundra ponds: Fine-scale analysis of surface hydrology in drained thaw lake basins over a 65 year period (1948–2013). *Journal of Geophysical Research: Biogeosciences*, *120*(3), 466–479. https://doi.org/10.1002/2014JG002778

Bayle, A., Gascoin, S., Berner, L. T., & Choler, P. (2024). Landsat-based greening trends in alpine ecosystems are inflated by multidecadal increases in summer observations. *Ecography*, *2024*(12), e07394. https://doi.org/10.1111/ecog.07394

Bring, A., Fedorova, I., Dibike, Y., Hinzman, L., Mård, J., Mernild, S. H., et al. (2016). Arctic terrestrial hydrology: A synthesis of processes, regional effects, and research challenges. *Journal of Geophysical Research: Biogeosciences*, *121*(3), 621–649. https://doi.org/10.1002/2015JG003131

Carroll, M. L., Townshend, J.R., DiMiceli ,C.M., Noojipady ,P., & Sohlberg, R. A. (2009). A new global raster water mask at 250 m resolution. *International Journal of Digital Earth*, *2*(4), 291–308. https://doi.org/10.1080/17538940902951401

Chastain, R., Housman, I., Goldstein, J., Finco, M., & Tenneson, K. (2019). Empirical cross sensor comparison of Sentinel-2A and 2B MSI, Landsat-8 OLI, and Landsat-7 ETM+ top of atmosphere spectral characteristics over the conterminous United States. *Remote Sensing of Environment*, *221*, 274–285. https://doi.org/10.1016/j.rse.2018.11.012

Chen, J., & Zhu, W. (2022). Comparing Landsat-8 and Sentinel-2 top of atmosphere and surface reflectance in high latitude regions: case study in Alaska. *Geocarto International*, *37*(20), 6052–6071. https://doi.org/10.1080/10106049.2021.1924295

Chen, M., Rowland, J. C., Wilson, C. J., Altmann, G. L., & Brumby, S. P. (2014). Temporal and spatial pattern of thermokarst lake area changes at Yukon Flats, Alaska. *Hydrological Processes*, *28*(3), 837–852. https://doi.org/10.1002/hyp.9642

Chen, Y., Cheng, X., Liu, A., Chen, Q., & Wang, C. (2023). Tracking lake drainage events and drained lake basin vegetation dynamics across the Arctic. *Nature Communications*, *14*(1), 7359. https://doi.org/10.1038/s41467-023-43207-0

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., et al. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, *219*, 145–161. https://doi.org/10.1016/j.rse.2018.09.002

Cooley, S. W., Smith, L. C., Stepan, L., & Mascaro, J. (2017). Tracking Dynamic Northern Surface Water Changes with High-Frequency Planet CubeSat Imagery. *Remote Sensing*, *9*(12), 1306. https://doi.org/10.3390/rs9121306

Cooley, S. W., Smith, L. C., Ryan, J. C., Pitcher, L. H., & Pavelsky, T. M. (2019). Arctic-Boreal Lake Dynamics Revealed Using CubeSat Imagery. *Geophysical Research Letters*, *46*(4), 2111–2120. https://doi.org/10.1029/2018GL081584

Cordeiro, M. C. R., Martinez, J.-M., & Peña-Luque, S. (2021). Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with

Level 2A processors. *Remote Sensing of Environment*, *253*, 112209. https://doi.org/10.1016/j.rse.2020.112209

Donchyts, G., Schellekens, J., Winsemius, H., Eisemann, E., & Van de Giesen, N. (2016). A 30 m Resolution Surface Water Mask Including Estimation of Positional and Thematic Differences Using Landsat 8, SRTM and OpenStreetMap: A Case Study in the Murray-Darling Basin, Australia. *Remote Sensing*, *8*(5), 386. https://doi.org/10.3390/rs8050386

Finger Higgens, R. A., Chipman, J. W., Lutz, D. A., Culler, L. E., Virginia, R. A., & Ogden, L. A. (2019). Changing Lake Dynamics Indicate a Drier Arctic in Western Greenland. *Journal of Geophysical Research: Biogeosciences*, *124*(4), 870–883. https://doi.org/10.1029/2018JG004879

Flood, N. (2017). Comparing Sentinel-2A and Landsat 7 and 8 Using Surface Reflectance over Australia. *Remote Sensing*, *9*(7), 659. https://doi.org/10.3390/rs9070659

Ghasempour, F., Sekertekin, A., & Kutoglu, S. H. (2023). How Landsat 9 is superior to Landsat 8: comparative assessment of land use land cover classification and land surface temperature. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *X-4-W1-2022*, 221–227. https://doi.org/10.5194/isprs-annals-X-4-W1-2022-221-2023

Grosse, G., Jones, B., & Arp, C. (2013). 8.21 Thermokarst Lakes, Drainage, and Drained Basins. In J. F. Shroder (Ed.), *Treatise on Geomorphology* (pp. 325–353). San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-374739-6.00216-5

Guay, K. C., Beck, P. S. A., Berner, L. T., Goetz, S. J., Baccini, A., & Buermann, W. (2014). Vegetation productivity patterns at high northern latitudes: a multi-sensor satellite data assessment. *Global Change Biology*, *20*(10), 3147–3158. https://doi.org/10.1111/gcb.12647

Hastie, A., Lauerwald, R., Weyhenmeyer, G., Sobek, S., Verpoorter, C., & Regnier, P. (2018). CO2 evasion from boreal lakes: Revised estimate, drivers of spatial variability, and future projections. *Global Change Biology*, *24*(2), 711–728. https://doi.org/10.1111/gcb.13902

Holden, C. E., & Woodcock, C. E. (2016). An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations. *Remote Sensing of Environment*, *185*, 16–36. https://doi.org/10.1016/j.rse.2016.02.052

Janiec, P., Nowosad ,J., & Zwoliński, Z. (2023). A machine learning method for Arctic lakes detection in the permafrost areas of Siberia. *European Journal of Remote Sensing*, *56*(1), 2163923. https://doi.org/10.1080/22797254.2022.2163923

Jones, B. M., & Arp, C. D. (2015). Observing a Catastrophic Thermokarst Lake Drainage in Northern Alaska. *Permafrost and Periglacial Processes*, *26*(2), 119–128. https://doi.org/10.1002/ppp.1842

Jones, B. M., Arp, C. D., Whitman, M. S., Nigro, D., Nitze, I., Beaver, J., et al. (2017). A lake-centric geospatial database to guide research and inform management decisions in an Arctic watershed in northern Alaska experiencing climate and land-use changes. *Ambio*, *46*(7), 769–786. https://doi.org/10.1007/s13280-017-0915-9

Jones, B. M., Tape, K. D., Clark, J. A., Nitze, I., Grosse, G., & Disbrow, J. (2020). Increase in beaver dams controls surface water and thermokarst dynamics in an Arctic tundra region, Baldwin Peninsula, northwestern Alaska. *Environmental Research Letters*, *15*(7), 075005. https://doi.org/10.1088/1748-9326/ab80f1

Jorgenson, J. C., Jorgenson, M. T., Boldenow, M. L., & Orndahl, K. M. (2018). Landscape Change Detected over a Half Century in the Arctic National Wildlife Refuge Using High-Resolution Aerial Imagery. *Remote Sensing*, *10*(8), 1305. https://doi.org/10.3390/rs10081305

Kirpotin, S. N., Berezin, A., Bazanov, V., Polishchuk, Y., Vorobiov, S., Mironycheva-Tokoreva, N., et al. (2009). Western Siberia wetlands as indicator and regulator of climate change on the global scale. *International Journal of Environmental Studies*, *66*(4), 409–421. https://doi.org/10.1080/00207230902753056

Knoll, L. B., Sharma, S., Denfeld, B. A., Flaim, G., Hori, Y., Magnuson, J. J., et al. (2019). Consequences of lake and river ice loss on cultural ecosystem services. *Limnology and Oceanography Letters*, *4*(5), 119–131. https://doi.org/10.1002/lol2.10116

Kuhn, C., de Matos Valerio, A., Ward, N., Loken, L., Sawakuchi, H. O., Kampel, M., et al. (2019). Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-*a* and turbidity. *Remote Sensing of Environment*, *224*, 104–118. https://doi.org/10.1016/j.rse.2019.01.023

Lara, M. J., Chen, Y., & Jones, B. M. (2021). Recent warming reverses forty-year decline in catastrophic lake drainage and hastens gradual lake drainage across northern Alaska. *Environmental Research Letters*, *16*(12), 124019. https://doi.org/10.1088/1748-9326/ac3602

Levenson, E. S., Cooley, S., Mullen, A., Webb, E. E., & Watts, J. (2025). Glacial History Modifies Permafrost Controls on the Distribution of Lakes and Ponds. *Geophysical Research Letters*, *52*(4), e2024GL112771. https://doi.org/10.1029/2024GL112771

Lima, T. M. A., Martins, V. S., Paulino, R. S., Caballero, C. B., Maciel, D. A., & Giardino, C. (2025). A general spectral bandpass adjustment function (SBAF) for harmonizing landsat-sentinel over inland and coastal waters. *Science of Remote Sensing*, *11*, 100225. https://doi.org/10.1016/j.srs.2025.100225

Lindgren, P. R., Farquharson, L. M., Romanovsky, V. E., & Grosse, G. (2021). Landsat-based lake distribution and changes in western Alaska permafrost regions between the 1970s and 2010s. *Environmental Research Letters*, *16*(2), 025006. https://doi.org/10.1088/1748-9326/abd270

Lyons, E., Sheng, Y., Smith, L., Li, J., Hinkel, K., Lenters, J., & Wang, J. (2013). Quantifying sources of error in multitemporal multisensor lake mapping. *International Journal of Remote Sensing*, *34*, 7887–7905. https://doi.org/10.1080/01431161.2013.827343

Maciel, D. A., Pahlevan, N., Barbosa, C. C. F., de Novo, E. M. L. de M., Paulino, R. S., Martins, V. S., et al. (2023). Validity of the Landsat surface reflectance archive for aquatic science: Implications for cloud-based analysis. *Limnology and Oceanography Letters*, *8*(6), 850–858. https://doi.org/10.1002/lol2.10344

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., & Gascon, F. (2017). Sen2Cor for Sentinel-2. In *Image and Signal Processing for Remote Sensing XXIII* (Vol. 10427, pp. 37–48). SPIE. https://doi.org/10.1117/12.2278218

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. https://doi.org/10.1080/01431169608948714

Mullen, A. L., Watts, J. D., Rogers, B. M., Carroll, M. L., Elder, C. D., Noomah, J., et al. (2023). Using High-Resolution Satellite Imagery and Deep Learning to Track Dynamic Seasonality in Small Water Bodies. *Geophysical Research Letters*, *50*(7), e2022GL102327. https://doi.org/10.1029/2022GL102327

Muster, S., Heim, B., Abnizova, A., & Boike, J. (2013). Water Body Distributions Across Scales: A Remote Sensing Based Comparison of Three Arctic Tundra Wetlands. *Remote Sensing*, *5*(4), 1498–1523. https://doi.org/10.3390/rs5041498

Muster, S., Riley, W. J., Roth, K., Langer, M., Cresto Aleina, F., Koven, C. D., et al. (2019). Size Distributions of Arctic Waterbodies Reveal Consistent Relations in Their Statistical Moments in Space and Time. *Frontiers in Earth Science*, *7*. https://doi.org/10.3389/feart.2019.00005

Nitze, I., Grosse, G., Jones, B. M., Romanovsky, V. E., & Boike, J. (2018). Remote sensing quantifies widespread abundance of permafrost region disturbances across the Arctic and Subarctic. *Nature Communications*, *9*(1), 5423. https://doi.org/10.1038/s41467-018-07663-3

Nitze, Ingmar, Grosse, G., Jones, B. M., Arp, C. D., Ulrich, M., Fedorov, A., & Veremeeva, A. (2017). Landsat-Based Trend Analysis of Lake Dynamics across Northern Permafrost Regions. *Remote Sensing*, *9*(7), 640. https://doi.org/10.3390/rs9070640

Nitze, Ingmar, Cooley, S. W., Duguay, C. R., Jones, B. M., & Grosse, G. (2020). The catastrophic thermokarst lake drainage events of 2018 in northwestern Alaska: fast-forward into the future. *The Cryosphere*, *14*(12), 4279–4297. https://doi.org/10.5194/tc-14-4279-2020

Olthof, I., & Fraser, R. H. (2024). Mapping surface water dynamics (1985–2021) in the Hudson Bay Lowlands, Canada using sub-pixel Landsat analysis. *Remote Sensing of Environment*, *300*, 113895. https://doi.org/10.1016/j.rse.2023.113895

Olthof, I., & Rainville, T. (2022). Dynamic surface water maps of Canada from 1984 to 2019 Landsat satellite imagery. *Remote Sensing of Environment*, *279*, 113121. https://doi.org/10.1016/j.rse.2022.113121

Olthof, I., Fraser, R. H., & Schmitt, C. (2015). Landsat-based mapping of thermokarst lake dynamics on the Tuktoyaktuk Coastal Plain, Northwest Territories, Canada since 1985. *Remote Sensing of Environment*, *168*, 194–204. https://doi.org/10.1016/j.rse.2015.07.001

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

Pahlevan, N., Mangin, A., Balasubramanian, S. V., Smith, B., Alikas, K., Arai, K., et al. (2021). ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. *Remote Sensing of Environment*, *258*, 112366. https://doi.org/10.1016/j.rse.2021.112366

Pastick, N. J., Jorgenson, M. T., Goetz, S. J., Jones, B. M., Wylie, B. K., Minsley, B. J., et al. (2019). Spatiotemporal remote sensing of ecosystem change and causation across Alaska. *Global Change Biology*, *25*(3), 1171–1189. https://doi.org/10.1111/gcb.14279

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

Pi, X., Luo, Q., Feng, L., Xu, Y., Tang, J., Liang, X., et al. (2022). Mapping global lake dynamics reveals the emerging roles of small lakes. *Nature Communications*, *13*(1), 5777. https://doi.org/10.1038/s41467-022-33239-3

Pickens, A. H., Hansen, M. C., Hancher, M., Stehman, S. V., Tyukavina, A., Potapov, P., et al. (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

Pickens, A. H., Hansen, M. C., Stehman, S. V., Tyukavina, A., Potapov, P., Zalles, V., & Higgins, J. (2022). Global seasonal dynamics of inland open water and ice. *Remote Sensing of Environment*, *272*, 112963. https://doi.org/10.1016/j.rse.2022.112963

Qin, Y., Zhang, C., & Lu, P. (2023). A fully automatic framework for sub-pixel mapping of thermokarst lakes using Sentinel-2 images. *Science of Remote Sensing*, *8*, 100111. https://doi.org/10.1016/j.srs.2023.100111

Rajib, A., Khare, A., Golden, H. E., Gupta, B. C., Wu, Q., Lane, C. R., et al. (2024). A call for consistency and integration in global surface water estimates. *Environmental Research Letters*, *19*(2), 021002. https://doi.org/10.1088/1748-9326/ad1722

Riordan, B., Verbyla, D., & McGuire, A. D. (2006). Shrinking ponds in subarctic Alaska based on 1950–2002 remotely sensed images. *Journal of Geophysical Research: Biogeosciences*, *111*(G4). https://doi.org/10.1029/2005JG000150

Roach, J. K., & Griffith, B. (2015). Climate-induced lake drying causes heterogeneous reductions in waterfowl species richness. *Landscape Ecology*, *30*(6), 1005–1022. https://doi.org/10.1007/s10980-015-0207-3

Roy, D. P., Kovalskyy, V., Zhang, H. K., Vermote, E. F., Yan, L., Kumar, S. S., & Egorov, A. (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment*, *185*, 57–70. https://doi.org/10.1016/j.rse.2015.12.024

Runge, A., & Grosse, G. (2019). Comparing Spectral Characteristics of Landsat-8 and Sentinel-2 Same-Day Data for Arctic-Boreal Regions. *Remote Sensing*, *11*(14), 1730. https://doi.org/10.3390/rs11141730

Sharma, R. C., Tateishi, R., Hara, K., & Nguyen, L. V. (2015). Developing Superfine Water Index (SWI) for Global Water Cover Mapping Using MODIS Data. *Remote Sensing*, *7*(10), 13807–13841. https://doi.org/10.3390/rs71013807

Sheng, Y., Song, C., Wang, J., Lyons, E. A., Knox, B. R., Cox, J. S., & Gao, F. (2016). Representative lake water extent mapping at continental scales using multi-temporal Landsat-8 imagery. *Remote Sensing of Environment*, *185*, 129–141. https://doi.org/10.1016/j.rse.2015.12.041

Smith, L. C., Sheng, Y., MacDonald, G. M., & Hinzman, L. D. (2005). Disappearing Arctic Lakes. *Science*, *308*(5727), 1429–1429. https://doi.org/10.1126/science.1108142

Song, C., Fan, C., Ma, J., Zhan, P., & Deng, X. (2025). A spatially constrained remote sensing-based inventory of glacial lakes worldwide. *Scientific Data*, *12*(1), 464. https://doi.org/10.1038/s41597-025-04809-z

Sui, Y., Feng, M., Wang, C., & Li, X. (2022). A high-resolution inland surface water body dataset for the tundra and boreal forests of North America. *Earth System Science Data*, *14*(7), 3349–3363. https://doi.org/10.5194/essd-14-3349-2022

Ulrich, M., Matthes, H., Schirrmeister, L., Schütze, J., Park, H., Iijima, Y., & Fedorov, A. N. (2017). Differences in behavior and distribution of permafrost-related lakes in Central Yakutia and their response to climatic drivers. *Water Resources Research*, *53*(2), 1167–1188. https://doi.org/10.1002/2016WR019267

Vermote, E., Roger, J. C., Franch, B., & Skakun, S. (2018). LaSRC (Land Surface Reflectance Code): Overview, application and validation using MODIS, VIIRS, LANDSAT and Sentinel 2 data's. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 8173–8176). Valencia: IEEE. https://doi.org/10.1109/IGARSS.2018.8517622

Walter Anthony, K., Daanen, R., Anthony, P., Schneider von Deimling, T., Ping, C.-L., Chanton, J. P., & Grosse, G. (2016). Methane emissions proportional to permafrost carbon thawed in Arctic lakes since the 1950s. *Nature Geoscience*, *9*(9), 679–682. https://doi.org/10.1038/ngeo2795

Wang, J., Pottier, C., Cazals, C., Battude, M., Sheng, Y., Song, C., et al. (2025). The Surface Water and Ocean Topography Mission (SWOT) Prior Lake Database (PLD): Lake Mask and Operational Auxiliaries. *Water Resources Research*, *61*(3), e2023WR036896. https://doi.org/10.1029/2023WR036896

Webb, E. E., & Liljedahl, A. K. (2023). Diminishing lake area across the northern permafrost zone. *Nature Geoscience*, *16*(3), 202–209. https://doi.org/10.1038/s41561-023-01128-z

Webb, E. E., Liljedahl, A. K., Cordeiro, J. A., Loranty, M. M., Witharana, C., & Lichstein, J. W. (2022). Permafrost thaw drives surface water decline across lake-rich regions of the Arctic. *Nature Climate Change*, *12*(9), 841–846. https://doi.org/10.1038/s41558-022-01455-w

Yamazaki, D., Trigg, M. A., & Ikeshima, D. (2015). Development of a global ~90m water body map using multi-temporal Landsat images. *Remote Sensing of Environment*, *171*, 337–351. https://doi.org/10.1016/j.rse.2015.10.014

Zhang, H. K., Roy, D. P., Yan, L., Li, Z., Huang, H., Vermote, E., et al. (2018). Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences. *Remote Sensing of Environment*, *215*, 482–494. https://doi.org/10.1016/j.rse.2018.04.031