

ASSESSING THE RELATIVE ACCURACY OF PLANET AND SENTINEL-2
DERIVED WATER MAPS USING FIELD DATA

by

IAN LANE RAMSEY VAN DUSEN

A THESIS

Presented to the Department of Geography
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Master of Science

June 2023

THESIS APPROVAL PAGE

Student: Ian Lane Ramsey Van Dusen

Title: Assessing the Relative Accuracy of Planet and Sentinel-2 Derived Water Maps Using Field Data

This thesis has been accepted and approved in partial fulfillment of the requirements for the Master of Science degree in Geography by:

Sarah Cooley
Johnny Ryan

Chairperson
Member

and

Krista Chronister

Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2023

© 2023 Ian Lane Ramsey Van Dusen

THESIS ABSTRACT

Ian Lane Ramsey Van Dusen

Master of Science

Department of Geography

June 2023

Title: Assessing the Relative Accuracy of Planet and Sentinel-2 Derived Water Maps Using Field Data

This study compares the accuracy of surface water maps from Sentinel-2 and Planet satellites with 43 shoreline observations on the Tanana and Willamette Rivers. High-precision GNSS rover provided the most precise results, with ~10cm accuracy. Handheld devices (BadElf: ~1m, eTrex: ~2m) were less accurate but still can be used for ground validation of satellite shorelines. For the Tanana River, Planet NDWI-derived water maps (~5m) were slightly more accurate than Sentinel-2 (~6m), despite smaller differences than their spatial resolutions. On the Willamette River, Planet achieved ~3m accuracy and Sentinel-2 ~4m accuracy using NIR-band thresholding due to minimal reflectance difference. The temporal advantage of Planet data was evident, with more clear sky observations, particularly in regions with low orbital convergence and during non-clear sky months. Despite slightly lower spatial accuracy and temporal resolution, the accessibility and reliability of Sentinel-2 data make the datasets comparable.

CURRICULUM VITAE

NAME OF AUTHOR: Ian Lane Ramsey Van Dusen

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Occidental College

DEGREES AWARDED:

Master of Geography, 2023, University of Oregon
Bachelor of Arts, 2020, Occidental College

AREAS OF SPECIAL INTEREST:

Remote Sensing
Hydrology
Geospatial Data Science

PROFESSIONAL EXPERIENCE:

Graduate Employee, University of Oregon, 2021-present
GIS Analyst, Bridger Aerospace, Summer 2021
Remote Sensing Researcher, Byrd Polar and Climate Research Center, 2020-2021

ACKNOWLEDGMENTS

Thank you to all those who communicated with and supported me throughout my time in the Department of Geography at the University of Oregon. Primarily, my advisor Sarah Cooley and committee member Johnny Ryan for their patience and insight into the many steps of a scientific research project. I benefited from watching your research practices bridging personal and scientific curiosities. To my professors and the Geograds for turning me into a geographer and remembering the story behind the data. Thank you to Eric Levenson for leading by example, Theo Ross for invaluable Python expertise, and rest of the lab for camaraderie, even if we have just found our stride in my final month. To my housemates for conversation and compassion, and to the Tanhands community for showing me the joys of Eugene and Western Oregon. Oregon has been an ideal home for trails, rivers, and volcanoes to keep my brain refreshed. Finally, thank you to my family and Lily Johnston for the confidence and foundation through these last two chaotic and rewarding years.

TABLE OF CONTENTS

| Chapter | Page |
|--|------|
| I. INTRODUCTION | 1 |
| II. STUDY SITE | 4 |
| 2.1. Geographic Location | 4 |
| III. METHODS | 7 |
| 3.1. Satellite Data | 7 |
| 3.2. Field Shorelines | 9 |
| 3.3. Comparison of Planet and Sentinel-2 Water Classifications to Ground-Validated Shorelines | 15 |
| 3.4. Multitemporal Analysis of Planet and Sentinel-2 Data | 17 |
| 3.5. Comparative Analysis to Discharge..... | 20 |
| IV. RESULTS | 21 |
| 4.1. Ground Truthing GNSS Devices | 21 |
| 4.2. Ground Comparison with Planet and Sentinel-2 Imagery | 22 |
| 4.3. Temporal Differences between Sentinel-2 and Planet..... | 28 |
| 4.4. Comparison to Discharge..... | 31 |
| V. DISCUSSION | 33 |
| 5.1. Shoreline Walking Validation..... | 33 |
| 5.2. Planet and Sentinel-2 Spatial Resolution..... | 34 |

| | |
|--|----|
| 5.3. Planet and Sentinel-2 Temporal Resolution | 36 |
| 5.4 Planet and Sentinel-2 Accessibility | 37 |
| 5.5. Derived Comparison to Discharge..... | 37 |
| 5.6. Discussion Summary | 40 |
| V. CONCLUSIONS..... | 41 |
| REFERENCES CITED..... | 43 |

LIST OF FIGURES

| Figure | Page |
|--|------|
| 1. Study Site Tanana River | 5 |
| 2. Study Site Willamette River | 6 |
| 3. Radiometric Values of Satellites..... | 9 |
| 4. Shoreline Walking Tanana River..... | 11 |
| 5.. Shoreline Walking Tanana River..... | 12 |
| 6. Handheld Sensor Comparison..... | 13 |
| 7. NDWI Stacked Histograms..... | 18 |
| 8. Flowchart Multi-temporality methods | 20 |
| 9. GNSS Receiver Results | 22 |
| 10. Classifier Accuracy: June Tanana..... | 23 |
| 11. Classifier Accuracy: August Tanana | 25 |
| 12. Classifier Accuracy: September Willamette..... | 27 |
| 13. Classifier Accuracy: November Willamette | 28 |
| 14. Tanana Temporality | 29 |
| 15. Willamette Temporality | 30 |
| 16. Tanana Sentinel-2 and Planet Discharge | 31 |

| | |
|---|----|
| 17. Tanana Sentinel-2 Discharge Calendar..... | 32 |
| 18. Subset Tanana Sentinel-2 Discharge | 39 |

LIST OF TABLES

| Table | Page |
|----------------------------------|------|
| 1. Radiometric Band Values | 8 |
| 2. Total Shoreline Surveys..... | 10 |

INTRODUCTION

1. Introduction

The last decade has brought substantial technological improvements to both governmental and private sectors' satellite technologies. These advancements have substantially improved data collection and accessibility and enhanced the spatial and temporal precision of these observations (Belward & Skøien, 2015). Private companies like Planet now manage constellations of hundreds of CubeSats, providing near-daily observations at 3.7-meter resolution (*Planet Monitoring - Satellite Imagery and Monitoring*, n.d.). In parallel, public-sector contributions such as the European Space Agency's Sentinel-2 satellites offer 10-meter resolution data with a maximum revisit time of 5 days or less when combining data from both satellites in orbit (Delwart, n.d.). This frequency of monitoring and near real-time data accessibility has initiated a paradigm shift in Earth observation. This shift has moved us from infrequent, large-scale observations towards a continuous, high-resolution monitoring system. Instead of waiting for periodic satellite passes, researchers and analysts now have access to a constant stream of high-resolution data, enabling more timely and detailed study of Earth's surface processes (Aragon et al., 2018). This advancement in satellite technology provides superior spatial and temporal resolution compared to earlier iterations of public earth observation satellites. For instance, Landsat 8 and 9, offer 30m resolution and have an 8-day revisit time (Claverie et al., 2018a; Zhu et al., 2019).

Surface water observations have been tracked from optical data for some time (Allen & Pavelsky, 2018; Alsdorf et al., 2007; Pekel et al., 2016). The assimilation of data from Planet and Sentinel-2, each usable since ~2016, has improved our ability to analyze Earth's surface water processes. Research has investigated the seasonality and variability of lakes (Cooley et al., 2017; Kaiser et al., 2021; Mullen et al., 2023; Qayyum et al., 2020). High-resolution remote sensing is beneficial for gaining insight into water quality (Lerch et al., 2005), determining discharge estimates (Bjerklie et al., 2023), and monitoring changes in river structure and behavior (Harlan et al., 2023). This not only

enhances our comprehension of river systems but also facilitates their management, especially in areas where gathering field data is logistically difficult, hazardous, or prohibitively expensive. The value of these satellite observations of surface water is underscored by the limitations inherent in traditional gage-based measurements. Historically, gage-based measurements have been the primary tool for researching fluctuations within rivers and lakes. Water flow estimates from gages form the foundation of our water management and hydrological understanding (Gleason & Durand, 2020), but these single-point measurements only provide direct observation of specific sites within a basin. Furthermore, gage sites are distributed unevenly across the globe, often influenced by national boundaries and proximity to urban centers. This uneven distribution, coupled with the decline in global gages, particularly in Arctic environments (Shiklomanov et al., 2002), limits the scope of large-scale hydrologic analyses. In isolated cold regions where temporal and spatial gaps in field observations are common, these advancements hold immense value in understanding phenomena. High-resolution remote sensing techniques can provide a globally available, consistent data source that can be used to complement and enhance hydrologic understanding from traditional gage-based measurements.

However, despite the high spatial resolution provided by satellite technologies, there are limitations to water classifications in complex environments. Optical imagery-based classification of surface water may struggle in areas with mixed vegetation and mixed pixels, often leading to their exclusion from inundation maps and discharge calculations. Such challenges arise from spectral similarities between different elements within a pixel and shadowing effects (DeVries et al., 2017; Hondula et al., 2021; Jones, 2015). The prevailing classification technique to identify water in high-resolution imagery relies on the Near-Infrared (NIR) band due to the absorptive properties of water. The NIR band is used to create the Normalized Difference Water Index (NDWI), which involves a ratio between the Green and NIR bands (McFEETERS, 1996). Further derivative techniques include Modified NDWI (MNDWI) using NIR and straight NIR-band thresholding (Mondejar & Tongco, 2019). These classification techniques, despite their computational simplicity, are suited for cross-comparison between multiple satellite

platforms and ground-validated shorelines (Liu et al., 2016). Therefore, they provide a trustworthy method for evaluating the precision of shoreline detection by various sensors.

Given the known classification challenges in water environments, ground validation of high-resolution water classifications is essential. While a few studies have used ground data to verify high-resolution water-mapping of map lake areas and ocean shorelines derived from satellite imagery (Mullen et al., 2023; Pilartes-Congo, 2022; Pitcher et al., 2020), there has been limited focus on evaluating the accuracy of high-resolution classifications over rivers. Consequently, developing a standardized protocol and methodology for ground truthing satellite classifications in river systems, where feasible, is of paramount importance.

In this study, we aim to evaluate the accuracy and utility of Planet and Sentinel-2-derived water maps through comparison with in-situ ground data. For this work, we focus on stretches of the Tanana River near Fairbanks, Alaska and the Willamette River near Eugene, Oregon. These diverse locations, with different river forms, management practices, and applications, allow us to assess water classification techniques in heterogeneous locations. To evaluate the functional accuracy of Planet and Sentinel-2-derived classifications, we compared them with in-situ GNSS shoreline surveys. We establish a protocol for data collection by walking along shorelines and developed an adaptive NDWI-based and NIR-based thresholding method for identifying open surface water in river systems. By comparing the accuracy of satellite-derived water maps with in-situ shoreline observations, we assessed the performance of satellite systems and their applicability and relevance in understanding the dynamics of river systems.

II

STUDY SITE

2.1 Geographic Location

This study examines two river reaches in the United States: the Tanana River near Fairbanks, Alaska, and the Willamette River near Eugene, Oregon. Although not representative of all river types, these sites are ideal for their diversity of shoreline environments and accessibility. Both rivers feature road-accessible boat launch sites and are suitable to return for future ground validation campaigns.

Our primary study site for ground validation is the Tanana River. We specifically examine a 14 km stretch located just southwest of Fairbanks and between two gauges, the Tanana at Fairbanks (USGS #15485500) upstream and the Tanana at Nenana (USGS #15515500) downstream. The Tanana peaks at roughly 80,000 CFS in mid-summer (July) and freezes at the surface in the winter months, initially in mid-October before melting out in May. The Tanana's braided nature, wide floodplain, and low banks make it an ideal river for tracking inundation changes. Its high sediment load, originating from upstream glaciers, creates a highly turbid, optically brown river that presents challenges for satellite water classification, despite being unaffected by diurnal melt patterns. The river's high latitude ($\sim 64^{\circ}\text{N}$) is advantageous for sun-synchronous polar orbiting satellites due to increased observations and overlapping flight patterns. Over two field campaigns during Summer 2022, we conducted shoreline surveys in June (when the river discharge was around 55,000 CFS) and surveys in August (45,000 CFS).

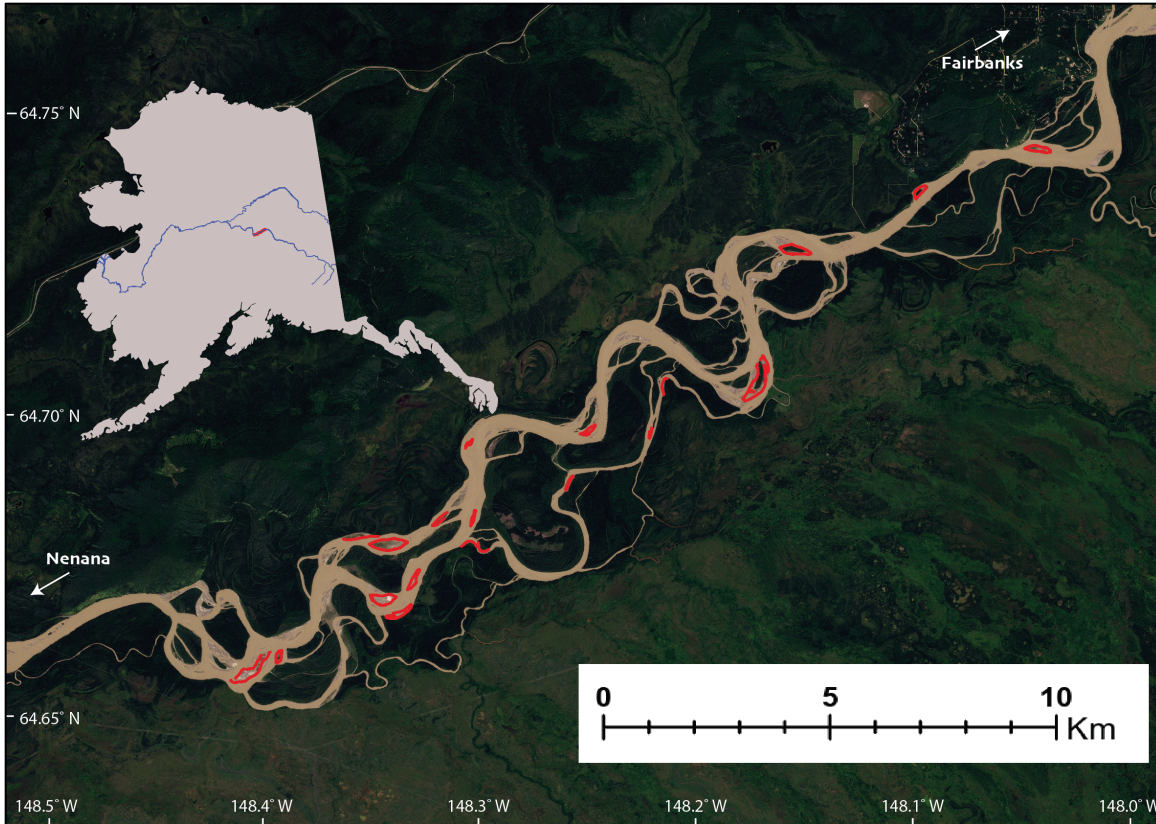


Figure 1. Field Site of the Tanana River in Alaska. Red Paths on the map are the GNSS Rover shoreline surveys from which we compare handheld GNSS receivers and satellite-derived water maps.

The Willamette River serves as the second site for ground truth shoreline validation. Our site was 2.5 kilometers downstream from the confluence with the McKenzie River and located upstream (approx. 8 km) of the USGS gage at Harrisburg (#14166000). We conducted shoreline surveys in September and November, taking advantage of the more reliable cloud-free days during summer and fall. The river's flow ranges from approximately 50,000 CFS at its peak, which can be caused by spring snowmelt or concentrated rain events, to as low as 4,000 CFS in the dry summer season. The river is heavily regulated by upstream dams operated by the US Army Corps of Engineers, which limit sediment transport resulting in an unusually low sediment load and relatively clear river. It is surrounded by cropland and urban areas, providing a very

different environment for testing water classification than the mostly forested Tanana River.

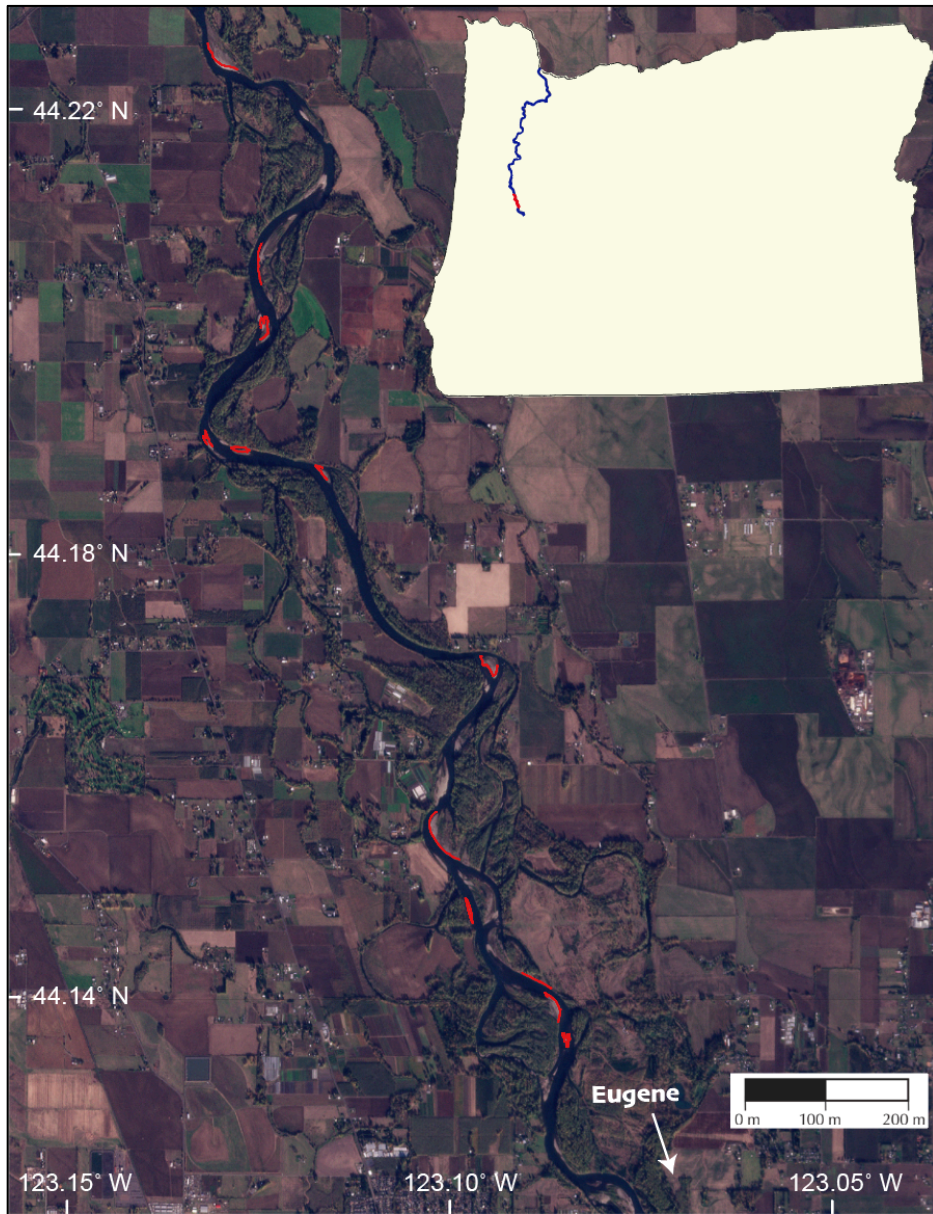


Figure 2. Study site of the Willamette River in Western Oregon. Reaches surveyed are highlighted in red. Most of the area of interest is surrounded by farmland. This section of river is just south of the Harrisburg gage.

III

METHODS

3.1 Satellite Data

Sentinel-2 is a satellite from the European Space Agency (ESA) - Copernicus Earth Observation and Monitoring program. The program includes two active satellites in the constellation, Sentinel-2A, and Sentinel-2B, launched in 2015 and 2017, respectively (Claverie et al., 2018b). A third satellite, Sentinel-2C, is planned to launch in 2024. Operating in a sun-synchronous low Earth orbit at 786km, the satellites have a maximum revisit time of 10 days at the equator, which is reduced to 5 days when combining data from both Sentinel-2A and 2B. The platform has a swath width of 290 km derived from the field of view at 20.6°. The satellites are equipped with a Multi-Spectral instrument (MSI), which capture the following wave bandwidths at 10m resolution: Blue (B2) between 460.7 nm to 524.7 nm, Green (B3) between 542.3 nm to 577.3 nm., Red (B4) between 649.6 nm to 679.6 nm, and Near-infrared between 780.25 nm to 885.35 nm (See Table 1 and Figure 3 for specifics)(Main-Knorn et al., 2017). Although Sentinel-2 also records Vegetation Red Edge and Shortwave Infrared (SWIR) bands at 20m resolution, our analysis focuses exclusively on the 10m bands. We analyze Sentinel-2's atmospherically corrected Level-2A data product, and for cloud masking, we use Sentinel-2's QA60 band to remove clouds, shadows, and other abnormalities (Main-Knorn et al., 2017). Notably, as a public space agency-operated sensor, Sentinel-2 data is freely available to the public.

Table 1. The radiometric values for the bands of interest. For Planet B, G, R, and NIR correlate to bands 1, 2, 3, and 4 at 3 m pixel size. Sentinel-2 data B, G, R, and NIR correlate to bands 2, 3, 4, and 8 at 10 m pixel size. All band values in nanometers.

| Satellite | Blue | Green | Red | NIR |
|------------------------------|---------------|---------------|---------------|---------------|
| Planet Dove Classic: PS2 | 455 - 515 | 400 - 590 | 590 - 670 | 780 - 860 |
| Planet Dove R: PS2.SD | 464 - 517 | 547 - 585 | 650 - 682 | 846 - 888 |
| Planet Super Dove: PSB.SD | 465 - 515 | 547 - 585 | 650 - 680 | 845 - 885 |
| Sentinel-2A: MSI | 460.2 - 525.2 | 542.3 - 577.3 | 649.6 - 679.6 | 780.3 - 885.3 |
| Sentinel-2B: MSI | 459.8 - 524.8 | 541.5 - 576.4 | 649.4 - 680.4 | 780.4 - 885.4 |

We also analyzed the high-temporal and high-spatial-resolution data provided by Planet. Planet operates a vast constellation of over 150 Dove CubeSats, making it the largest Earth observation satellite network globally (*Planet Monitoring - Satellite Imagery and Monitoring*, n.d.). Dove CubeSats, measuring 10cm x 10cm x 30cm, are positioned in sun-synchronous orbits approximately 400km above the Earth's surface. For our specific analysis, we exclusively use the 4-band product. Historically, the imagery came from 3 sources the Dove, Dove-R, and Super Dove satellites, and at present, imagery is solely from the Super Dove satellites (Table 1). The Dove CubeSats provides a ground spatial resolution of 3.7m and a swath width of 25.0km. We use Planet Orthotiles, which are orthorectified and atmospherically corrected by Planet to provide analysis-ready data. The atmospherically corrected surface reflectance product mitigates radiometric inconsistencies between raw Planet imagery, a previously noted difficulty working with the constellation (Claverie et al., 2018b). This is appealing as we can directly compare Planets' 4-band (RGB+NIR) surface reflectance product to Sentinel-2's four 10m resolution bands (RGB+NIR) surface reflectance data. For cloud masking, we use Planet's UDM mask, a data mask produced by Planet that flags pixels with snow, clouds, and smoke (*UDM 2*, n.d.). Planet may sometimes have non-published data, which is predominately due to cloudy conditions, where imagery is unable to locate due to a ground lock (*PlanetScope*, n.d.). Planet is a commercial company, and therefore, its data

is not freely available; the company does maintain partnerships with governments and universities, which enhance scientific accessibility to its dataset.

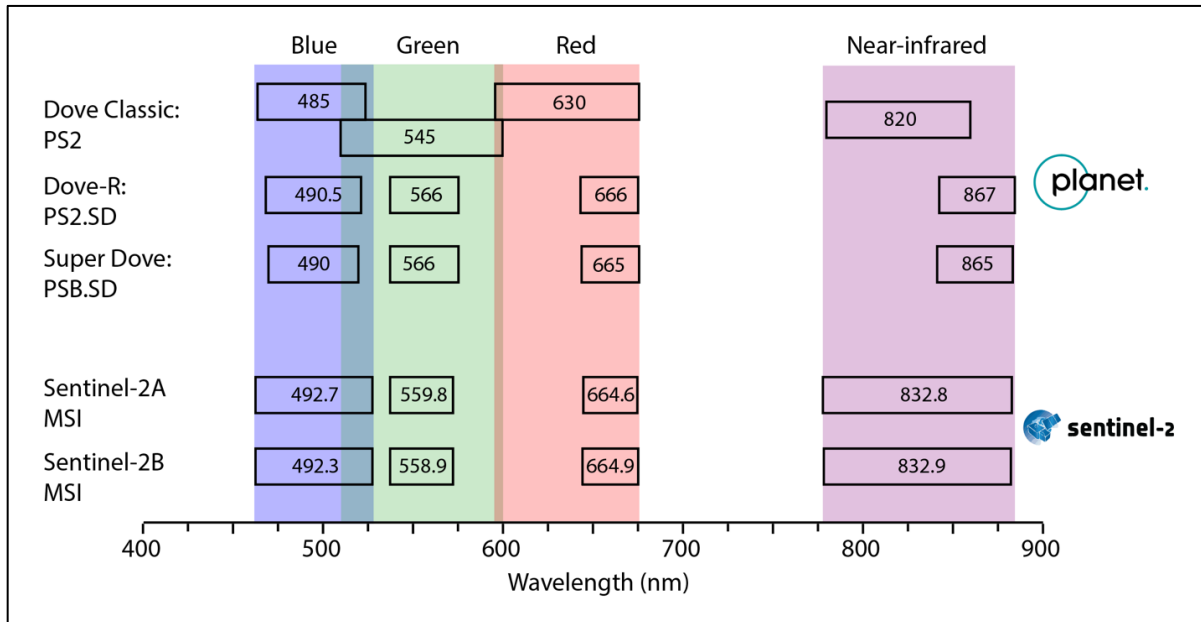


Figure 3. The Radiometric band values for Planet (B1, B2, B3, B4) @3m and Sentinel-2 (B2, B3, B4, B8) @ 10m. Numbers within the boxes indicate the central wavelength, while the box width is proportional to the bandwidth (nm). Since April 2022, Planet has been using the SuperDove dataset, which is a similar product to the Sentinel-2 sensors allowing for harmonization between the two products. The biggest difference between the two platforms occurs in the Near-Infrared Band.

3.2 Field Shorelines

Our field sites for the Tanana and the Willamette Rivers concentrated on reach scales (approximately 10 km in river length), as this scale represents a practical unit in local remote sensing studies and can be extrapolated to global assessments. The 10 km reach length is in reference to centerline river distance. However, there are more than 10 km of shorelines to identify over such a stretch. Ideally, the analysis should encompass a broad islanded river stretch where minor changes in river stage lead to large changes in

inundation detectable via satellites. Both reaches of interest for this study meet the criteria. On the Tanana River, we walked shorelines period from June 22nd to 24th and again from August 26th to 28th (Figure 4). The June survey, we walked 13 shorelines, for a total of 9.09 km of comparable data, while discharge was at ~50,000 cfs. In August we surveyed 15 shorelines, over 10.62 km, with a discharge of ~48,000 cfs. On the Willamette River, we surveyed shoreline locations on September 27th and November 18th (Figure 5). The September field data, we gathered 7 shorelines totaling 3.78 km at a discharge of 5500 cfs. The November data totaled 8 shorelines for 4.19km at 6370 cfs (Table 2).

Table 2. Total Number of shorelines surveyed over the Willamette and the Tanana Rivers and the total cumulative length at each field site used for comparison.

| Field Site | Total Shorelines Surveyed | Cumulative Length |
|----------------------|----------------------------------|--------------------------|
| June Tanana | 13 | 9.09 km |
| August Tanana | 15 | 10.62 km |
| September Willamette | 7 | 3.78 km |
| November Willamette | 8 | 4.19 km |



Figure 4. Shoreline walking environments for the survey on the Tanana River. Quality of shoreline observations correlated with sediment size and saturation of the sediment.



Figure 5. Shorelines on the Willamette River for survey. Increased sediment size made a clear shoreline, while in vegetated areas, we could not walk directly over the open water interface.



Figure 6. Handheld (black and yellow) and High-precision receivers (green and orange), we had a consistent methodology, including survey poles and walking directly above the wet/dry interface.

The first goal of our research is to establish a consistent and accurate protocol for river shoreline surveying to enable reliable ground truthing comparisons with satellite data. We designed our in situ surveying methods from previously published work mapping lake surface area and coastal shorelines (Pilartes-Congo, 2022; Pitcher et al., 2020). We evaluated three different GNSS receivers for shoreline surveying: a high precision rover (the Septentrio Altus NR3), with a cost of approximately \$15,000, a weight of 1.16kg (excluding survey pole), and a stated accuracy of less than 10cm; the handheld Bad Elf GNSS Surveyor, priced at \$650, weighing approximately 90 grams, and possessing a stated kinematic accuracy of 2.5 meters; and the handheld Garmin Etrex, priced at roughly \$110, weighing 141 grams, and offering a stated accuracy of 15

meters. A consistent protocol was applied to all GNSS devices to ensure a direct comparison between various sensor types (Figure 6). The receivers were affixed to survey poles, with handheld instruments secured by screws or duct tape, and maintained level directly above the shoreline interface zone (between wet and dry//between water and land). Before survey commencement, devices were activated to operate for a minimum of 15 minutes to establish a stable satellite connection. Throughout the shoreline survey, devices consistently logged data at 1-second intervals. In instances where shoreline traversal was not feasible, we documented the timing where we could not exactly survey the shoreline and subsequently clipped during data processing.

Shorelines were selected based on length, accessibility, and safety. River shorelines were surveyed on islands and banks with clear walkable reaches of at least 50m in length while avoiding woody banks with overhanging vegetation that might pose risks during data collection. Ideally, field observations were conducted between 11 AM and 2 PM to align as closely as possible with sun-synchronous orbit observation times. When multiple users were involved in data collection, we maintained consistency by following the footsteps or route of the person with the high-precision rover (who walked first). The primary goal was to assess the accuracy of different GNSS sensors rather than evaluate errors associated with different individuals' interpretations of walking shorelines.

Upon completing data collection, we adhered to specific processing steps to maintain consistency between GNSS devices. We waited at least one week after data collection to allow for the appropriate RINEX conversion updates. The Septentrio Rover records data in Receiver Independent Exchange Format (RINEX) and requires post-processing, with free services offered by the Government of Canada through CSRS-PPP. Subsequent output CSV files were reprojected to UTM Zone 6N (EPSG:32606). We converted the GNSS outputs to SHP format and reprojected them into UTM Zone 6N. In instances where we were unable to walk a stretch of shoreline safely and accurately, such as due to the presence of large woody debris or overhanging banks, we noted the time. We removed all recorded points on all devices from those locations to ensure they did not affect our shoreline calculations.

To assess the performance and accuracy of shoreline mapping from different GNSS receiver types, we conducted a comparison of mapped shoreline positions. Operating under the assumption that the high-precision rover (the Septentrio Altus NR3) is the most accurate device, given its stated error margin of approximately 15 cm, we followed a series of steps for each shoreline. First, we converted the rover point SHP into a line to account for differences in internal logging speeds between the devices. Second, we calculated the average distance from each receiver point to the rover line, considering both the average distance (error from the rover) and the total number of points. Finally, we computed the weighted cumulative average for the receivers across all islands in the respective field season. This gave us a mean average error for each receiver. In total, we walked 28 shorelines on the Tanana River using the Septentrio rover, BadElf, and Etrex devices. The shorelines we walked ranged in length from 100m to 2km. On the Tanana River, we surveyed a total length of 19.71 km. On the Willamette River, we walked 15 shorelines totaling 6.97 km, solely using the BadElf, as we completed the field effort after validation of device accuracy from the Tanana River.

3.3 Comparison of Planet and Sentinel-2 Water Classifications to Ground-Validated Shorelines

We obtained imagery acquired as close to our field surveys as possible. We compared our ground validations for June surveys to imagery from June 24th, as both platforms had clear sky imagery on this day. However, in August, there were no clear sky days coincident with when fieldwork occurred. Thus, we used the closest clear-sky observations from August 20th for Sentinel-2 (45,800 cfs) and August 23rd (43,300 cfs) for Planet. Over our August field survey days, discharge in the river ranged from 47,600 – 48,500 cfs. On the Willamette River, we conducted fieldwork on September 27th and November 18th and analyzed Sentinel-2 and Planet data from the exact same dates to coincide with satellite observations. Our objective was to assess the 'best fit' between satellite classifications and in situ shorelines to determine the functional accuracy of classifications compared to ground data. Since we specifically chose to analyze clear sky

days, we did not need to use a cloud mask for this part of the analysis, avoiding potential cloud masking effects that would limit the accuracy of the comparison.

We chose to use Normalized Difference Water Index (NDWI)-based classifications for surface water due to its compatibility with both Planet and Sentinel-2, as well as its prevalent use in contemporary research (Cooley et al., 2017; Huang et al., 2018; McFEETERS, 1996; Mullen et al., 2023; Sogno et al., 2022). On the Tanana River, we first tested an NDWI classification (Green-NIR)/(Green+NIR). To capture the entire land water boundary, we generated shorelines from the NDWI band by applying binary thresholds in 0.01 increments between -0.2 and 0.2 NDWI values. To reduce noise, we sieved the values using a threshold of 5 and converted the TIFFs to polygon shapefiles. Finally, we converted the polygons into lines, with each line corresponding to a shoreline classification based on the selected thresholds.

On the Willamette River, Planet NDWI histograms did not return a consistent 2-peaked histogram as idealized in water classification studies (McFEETERS, 1996), representing water and land (Figure 5), meaning the NDWI water maps classification was unreliable. Upon careful examination of the imagery, we think this is likely due to radiometric inconsistencies deriving from the low sensitivity of Planet's Green band. We, therefore, instead tested water classification via thresholding in just the NIR band. Owing to the strong absorptive properties of water in the NIR, thresholding of the NIR band is another common water classification technique (Li et al., 2012; Liu et al., 2016) and has also been used with both Sentinel-2 and Planet (Islam & Ahamed, 2023). To find the best threshold for each image, we calculated the reflectance thresholds between 0.0 and 0.40 at 0.01 increments to find the threshold most accurate to our shoreline observations, following the same comparison steps as on the Tanana River.

To estimate the error compared to ground data for each satellite classification, we calculated the average error from the satellite to the in situ GNSS survey shorelines using the following steps. First, we computed the distance from each individual GNSS survey point to the nearest shoreline (24 total shorelines for the Tanana). We capped the maximum distance (error) at 20 meters to prevent the inclusion of extreme outliers

caused by entirely missed in-river islands. By averaging the total distance from every GNSS survey point, we derived the mean average error in the accuracy of each classification. This procedure was repeated for every classification threshold (40 for each sensor, in both NDWI and NIR) to identify the most accurate threshold in comparison with our in-situ shorelines.

3.4 Multitemporal Analysis of Planet and Sentinel-2 Data

To assess the functional temporal resolution of each dataset, we also analyzed imagery covering the entire 2022 open water seasons for both the Tanana and Willamette Rivers. We evaluated the multi-temporality of these datasets within the same SWORD reaches that we used for ground-validated shorelines. For the Tanana River, the ice-free time period spans roughly from May 8th to October 31st, 2022, while on the Willamette River, we analyzed the entire year, from January 1st to December 31st, 2022.

Sentinel-2 data collection and cleaning were performed in Google Earth Engine. The process began by merging same-date imagery and selecting the clearest pixels. Cloudy and shadowed areas were masked out using the QA60 band (Ming Wang, 2020). The final image comprised a masked 4-band TIFF (RGB + NIR) clipped to the reach, with cloudy data masked out as no data. This resulted in an image for every day that Sentinel-2 collected data over the designated reach.

Planet imagery was acquired through Planet Explorer and subsequently processed and cleaned locally. All data overlapping our specified date range and region of interest (ROI) were downloaded. We used the unusable Data Mask (UDM) from Planet (*UDM 2*, n.d.) to remove erroneous pixels not of clear ground, assigning a no-data value to masked-out portions of the image. Images from the same date were then mosaicked together, prioritizing clear image pixels. The product consisted of a mosaicked, cloud-masked 4-band TIFF (RGB + NIR) every day when a Planet observation was available over the ROI.

We established a river extent larger than the maximum observed extent, which we

used as the bounding limit and clipped all of our S2 and Planet imagery too. This meant that our histograms for NDWI data would display NDWI bimodal displays, as they have a closer to equal balance of Land and Water pixels in the imagery. Unlike our shoreline comparison, for our multitemporal analysis, we do not use a standardized threshold. Initial testing and previous work demonstrated that empirically selected thresholds can vary significantly due to image radiometric values, regions, and seasonality (Brown et al., 2022; Claverie et al., 2018b; Frazier & Hemingway, 2021), leading to overestimation or underestimation of water area. Thus, instead, we use a normalized adaptive technique when using the same classification on more than one date. Adaptive thresholding is an effective tool for comparing images to classify water (Brown et al., 2022; Cooley et al., 2017). Generally, NDWI histograms over water bodies are bimodally peaked (Figure 7), with each representing land and water (Zhang et al., 2018).

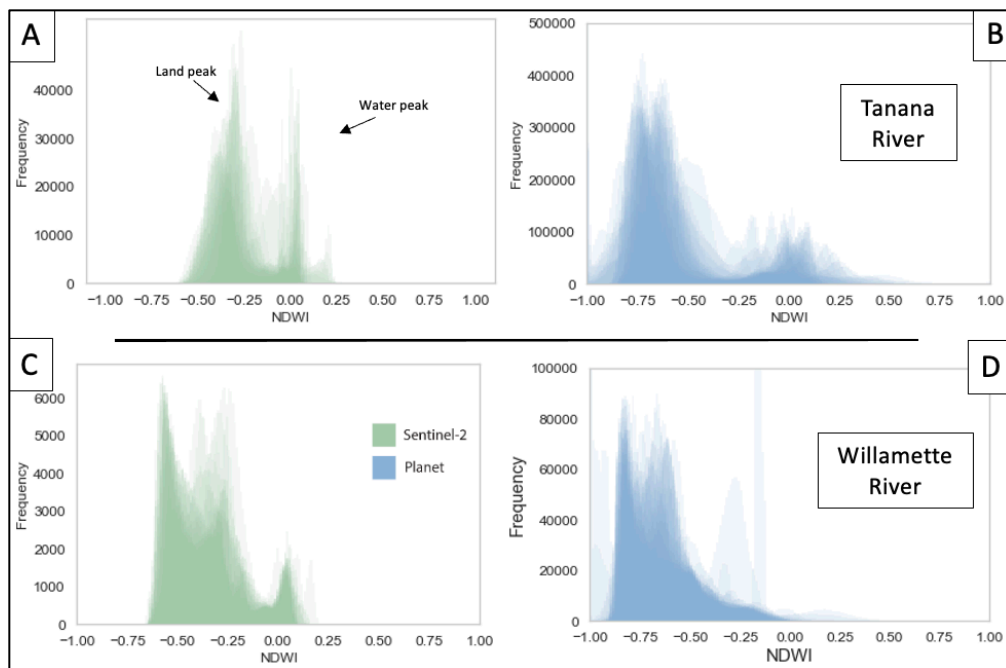


Figure 7. A+B: NDWI Histograms over the Tanana River from all data over the study area. We find distinctive two peaked histograms, although there is a significant variance between images on both platforms. C+D: NDWI histograms over the Willamette River, as you can see, Planet data does not display a clear second peak, while the Sentinel-2 imagery is not as distinctive as on the Tanana.

For all images from both platforms, we classified water using normalized NDWI and NIR thresholding techniques, enabling consistent and accurate comparisons of satellite data throughout the year. We calculated the NDWI band, which we then normalized between 0-100, with 0 representing the top of the land peak and 100 representing the top of the water peak. The same normalization technique was applied with NIR data on the Willamette. However, the band peaks are inversed relative to NDWI, with water as the left peak (0) and land as the right peak (100). Given the radiometric variability between different sensors and observations, this normalization facilitates direct image comparison and generates more consistent results than employing fixed NIR or NDWI thresholds. We classified a chosen normalized value (e.g., 70 for NDWI) as water if greater than 70 and land if less than 70, producing a binary classification for a water/non-water map. We performed the same analysis for NIR. Both Planet and Sentinel-2 data were filtered to include only images with at least 75% clear river data, discarding images with less than 75% clear data. This means that the observations must cover at least 75% of the reach while having usable data; otherwise, they are deemed ineffective for rating curve creation or island comparison.

We evaluated the temporal frequency of Sentinel-2 and Planet observations to understand each dataset's usability. Each day was classified into three categories for each dataset: No observation, Usable observation, and Unusable observation. No observation means that an image was not collected by that sensor on that day. We define a useable observation as when the reach was covered 75% usable (i.e., not cloudy, not missing) data on a given day for a given sensor. We analyzed all data that overlapped with the ROI, even if it was just a slight corner, so because of that, unusable data consisted of both cloudy days and clear days where the imagery did not sufficiently cover the ROI. The steps to this process are seen in Figure 8.

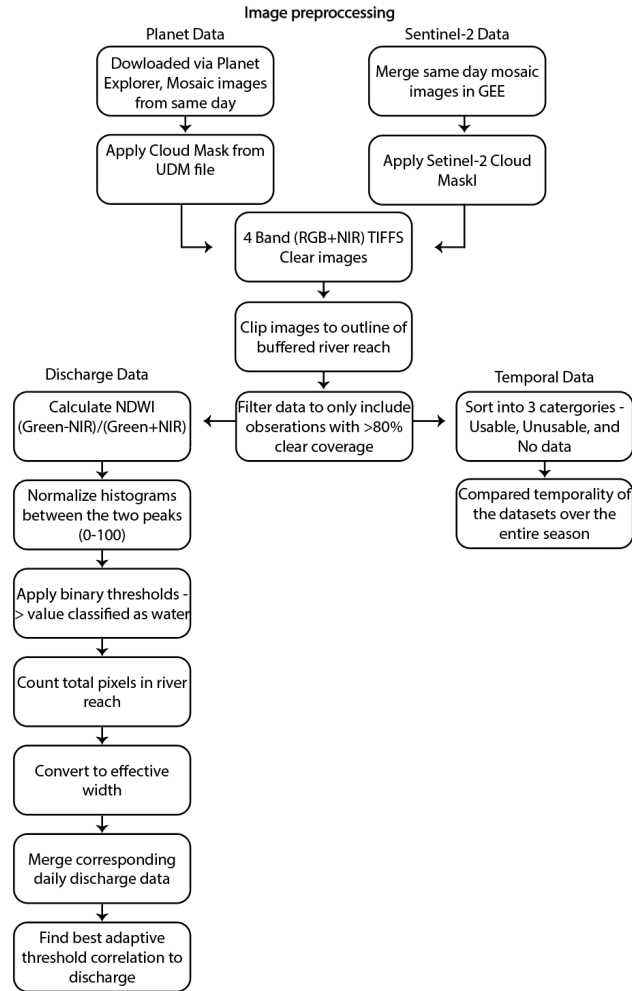


Figure 8. Flow chart to explain the process for computing utility of temporal data and comparing with discharge data at the reach scale.

3.5 Comparative Analysis of Discharge

We also assessed the comparative performance of Planet and Sentinel-2 satellite data by examining the correlation between observed inundation and river discharge obtained from USGS gages. To perform this correlation, we converted the total water pixels in water classification to effective width and assigned this value to each observation over the region of interest. Discharge data from the USGS gauge directly upstream of the reach were then compared to the effective width values. We established a rating curve to delineate the relationship between discharge and effective width.

IV

RESULTS

4.1 Ground Truthing GNSS Devices

Through comparison of high precision vs. handheld GNSS surveying devices, we find that the relatively affordable, handheld Bad Elf and Etrex GNSS receivers exhibit satisfactory performance when compared to the high-precision rover, indicating their suitability for shoreline mapping with an acceptable accuracy of approximately 1-2 meters. Over 24 shorelines on the Tanana River, the Bad Elf GNSS receiver had an average error of roughly 0.6m, while the Etrex GNSS receiver exhibited an accuracy of approximately 1.25 meters. It is essential to employ proper usage techniques, such as utilizing a survey pole, ensuring consistency, and activating the devices before initiating measurements, to attain these accuracies in the X and Y dimensions. Both Etrex and Bad Elf devices exhibit superior spatial accuracy compared to all but the highest-resolution satellites, rendering them appropriate for ground-truthing. Moreover, their portability, accessibility, and affordability render them advantageous over the Septentrio Rover when high precision (i.e., < 50 cm) is not required. We did not quantify the error in the Z-dimension.

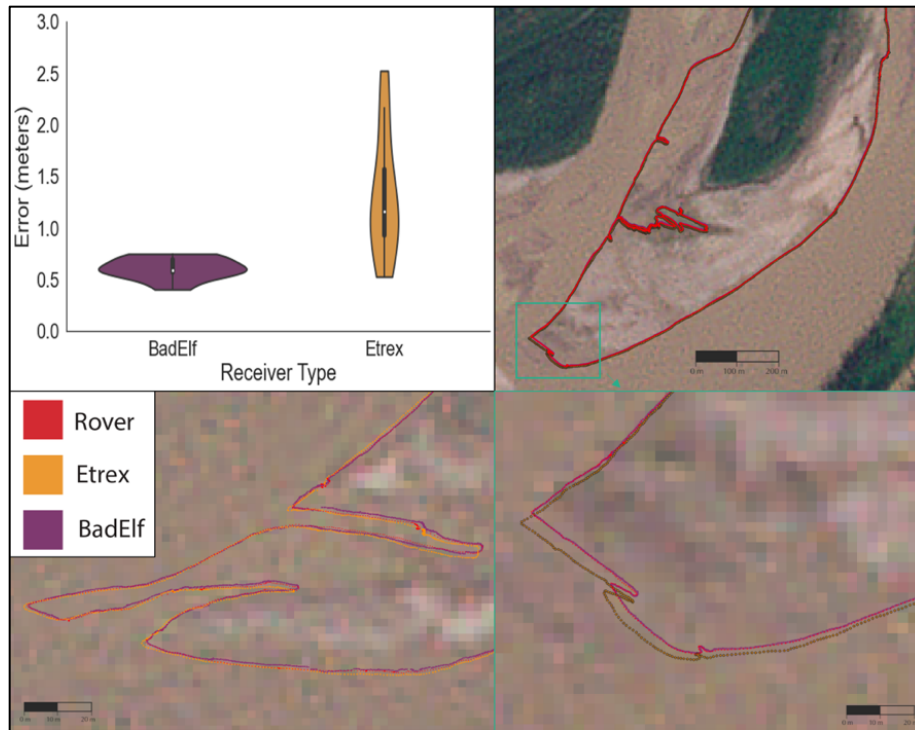


Figure 9. Comparison of the spatial accuracy of BadElf and Etrex GNSS devices relative to the Septentrio Rover. The BadElf has less variance and a lower mean error than the Etrex. Both devices are accurate to the rover on the island scale, but zooming in the Bad Elf has spatial accuracy on the sub-meter scale, while the Etrex has slightly more variance.

4.2 Ground Comparison with Planet and Sentinel-2 Imagery

On the Tanana River in June, we found that the NDWI classification method using both Planet and Sentinel-2 satellites yielded consistently accurate results. For Planet observations, we determined that the optimal NDWI threshold was -0.01, which resulted in an error compared to the GNSS shoreline retrievals of approximately 5.5 meters. This is slightly larger than Planet's stated ground accuracy of 3.7 meters. The optimized NDWI threshold for Sentinel-2 imagery was 0.03 with an error of around 6.5 meters, which is less than Sentinel-2's stated ground sampling distance of 10 meters. Our results suggest that applying the same NDWI threshold to both Sentinel-2 and Planet imagery does not lead to accurate results, even when using atmospherically corrected

imagery. Additionally, we found that the NIR classification method was not effective for the Tanana River, likely due to the high sediment content that prevents the formation of clear land/water peaks. In June, the river exhibited a pattern of diminishing discharge, resulting in saturated shorelines along the areas we surveyed.

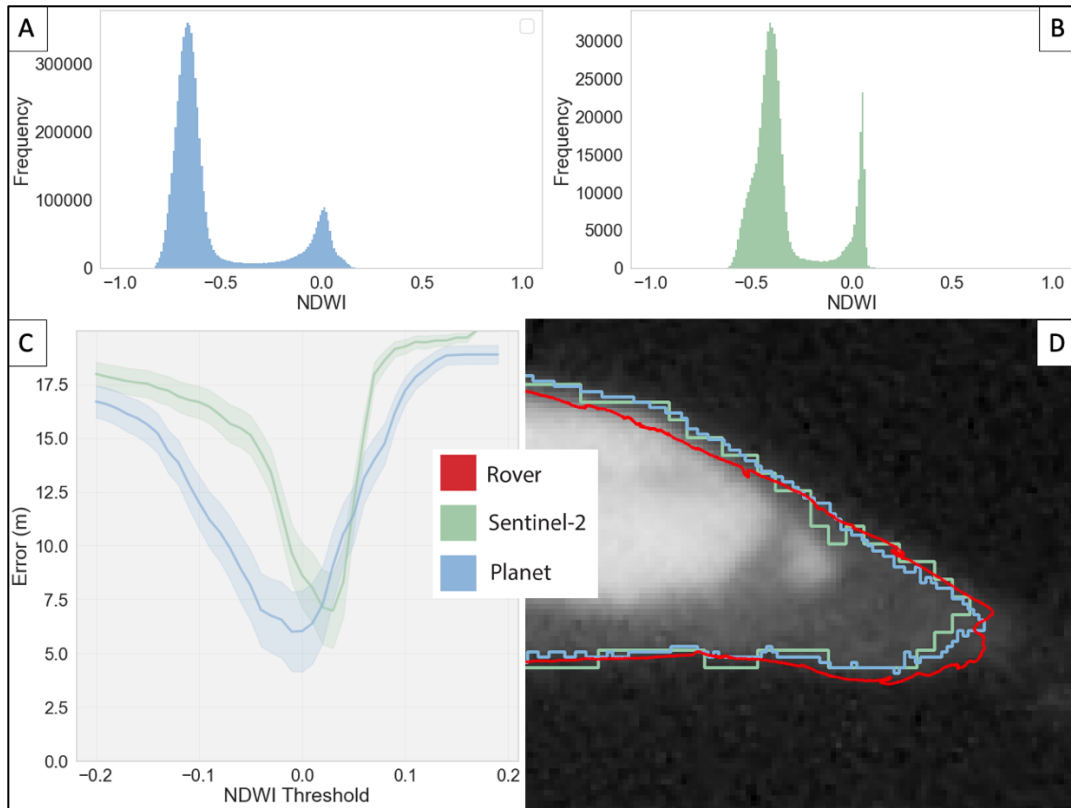


Figure 10 – A+B. NDWI histograms for Planet and Sentinel-2 data from coincident observations on June 24th. C: Planet data is slightly more accurate in identifying the shorelines than Sentinel-2 relative to our ground shoreline surveys indicated by the Rover. Both images have a small threshold area where data classification accuracy is increased. D: Map view of the data, can see the increased resolution of the Planet data (blue) relative to that of Sentinel-2.

For our August surveys, our survey window from the 26th to the 28th was disrupted by persistent cloud cover, preventing us from acquiring clear imagery.

Consequently, we had to use the closest clear-sky imagery: Planet's observation from August 23rd and Sentinel-2's from August 20th. Although the Sentinel-2 image was captured farther from our survey dates, it better resembled the river's water level during our fieldwork. In contrast, the Planet image was taken when the water level was slightly lower, leading to variations in shoreline locations. Furthermore, the river's water level was rising or consistent in August, while it was receding in June. As a result, our August observations revealed fewer muddy inundated zones, making satellite identification more straightforward compared to the conditions we encountered in June. In August, the best agreement threshold was -0.01 for Sentinel-2, resulting in an error of approximately 3 meters, while the threshold for Planet was -0.07, resulting in an error of around 5.5 meters. Both these thresholds were lower than those applied in June, signifying the need for flexible threshold values when comparing different date imagery. This result in threshold values for maximum accuracy underscores the importance of avoiding a one-size-fits-all approach when comparing imagery across dates.

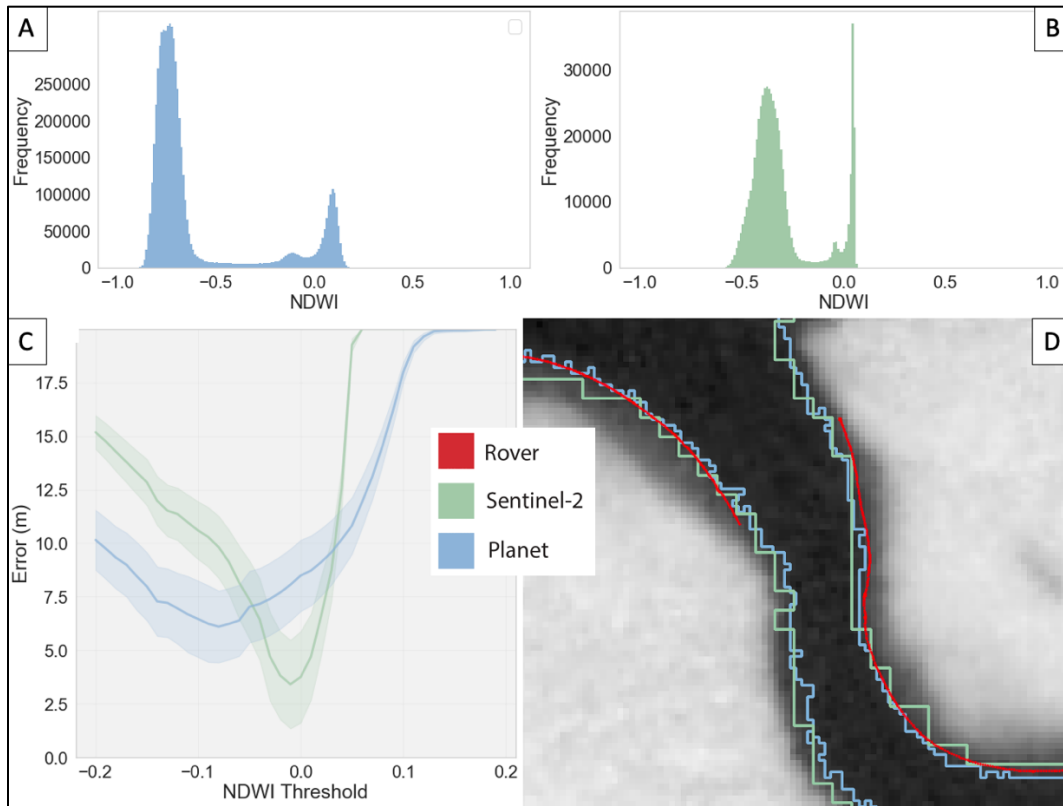


Figure 11 – A+B: The NDWI histograms over the Tanana River in August 2022. Both Planet and Sentinel-2 show a second water hump in areas with turbid water (Liu et al., 2016). C: It must be noted that we did not have coincident imagery, and although Planet was closer to the date of observation, the river was at a more similar level for Sentinel-2. A rising river likely led to more distinctive identifiable shorelines. D: In the shoreline classification, it is clear to see the difference in pixel size between Sentinel-2 and Planet observations.

In contrast to the Tanana River, the NDWI classification on the Willamette River did not produce two distinct land/water peaks in Planet imagery (Figure 4). This result is potentially attributable to the radiometric sensitivity of the Planet Green band, as the NDWI data over the river exhibited considerable noise. Sentinel-2 data yielded a two-peaked NDWI histogram, albeit not as distinct as over the Tanana. Both Planet and Sentinel-2 data generated two-peaked histograms in NIR classification, but establishing a consistent threshold for the entire year proved difficult due to seasonal variations. For the

field data from September 27th, with a discharge at the Harrisburg gage at ~5,500 cfs, we find the optimal NIR thresholds to be 0.15 for Planet and 0.23 for Sentinel-2. These thresholds highlight the radiometric disparities between the two satellite systems, specifically when using the NIR band (Figure 3). The error of Planet data was approximately 3.0 meters, while that of Sentinel-2 was around 3.5 meters. Our second field survey date was on November 18th, discharge of 6,400 cfs. We used coincident satellite imagery observations. We find a NIR threshold of 0.16 for Planet and 0.23 for Sentinel-2. With a mean average error of 3.5 and 4m, respectively. The discharge in November was higher than in September but was during a falling period as Discharge peaked at 19,000 cfs on November 5th and declined steadily til our survey date. In this example, NIR performed superior to NDWI, and the field data demonstrated it to be a more representative classification scheme.

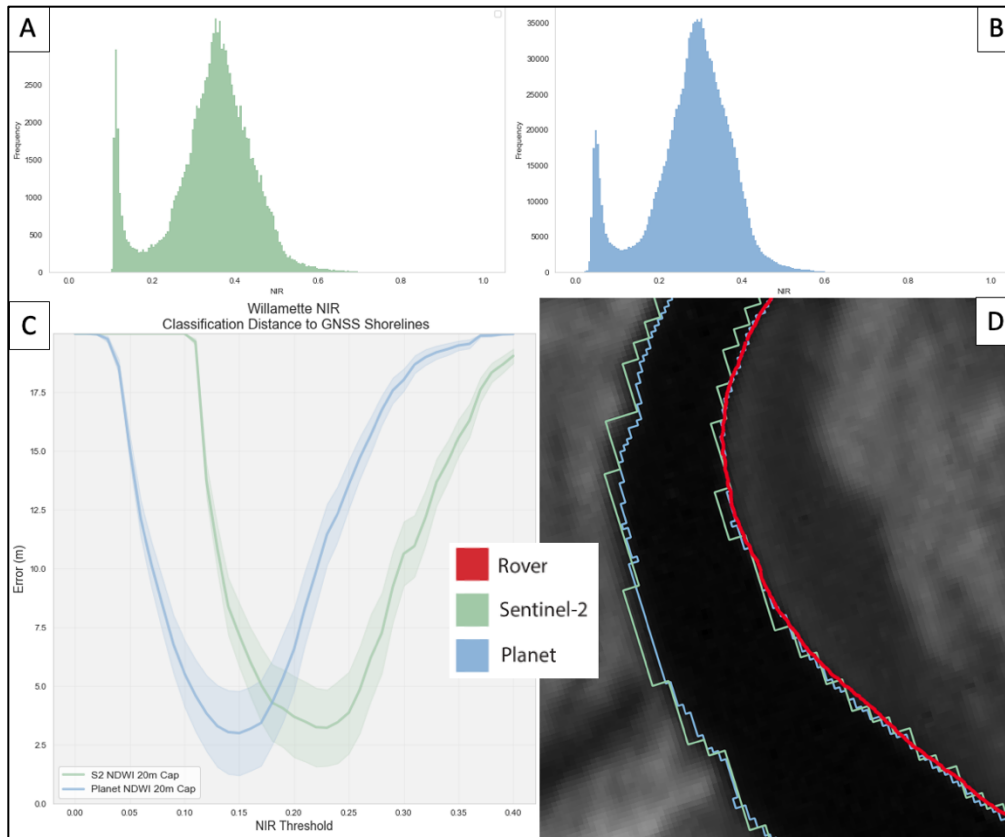


Figure 12 – A+B: Histograms for the NIR band surface reflectance values over the Willamette River on September 27th. Sentinel-2 and Planet NIR values do not begin at the same reflectance origin, and the thresholds are not directly comparable to one another. C: Accuracy for both satellites is significantly improved to just over 3m on the Willamette River, where there are more defined shorelines. D: The image in the bottom right is NIR based, with the black area representing water.

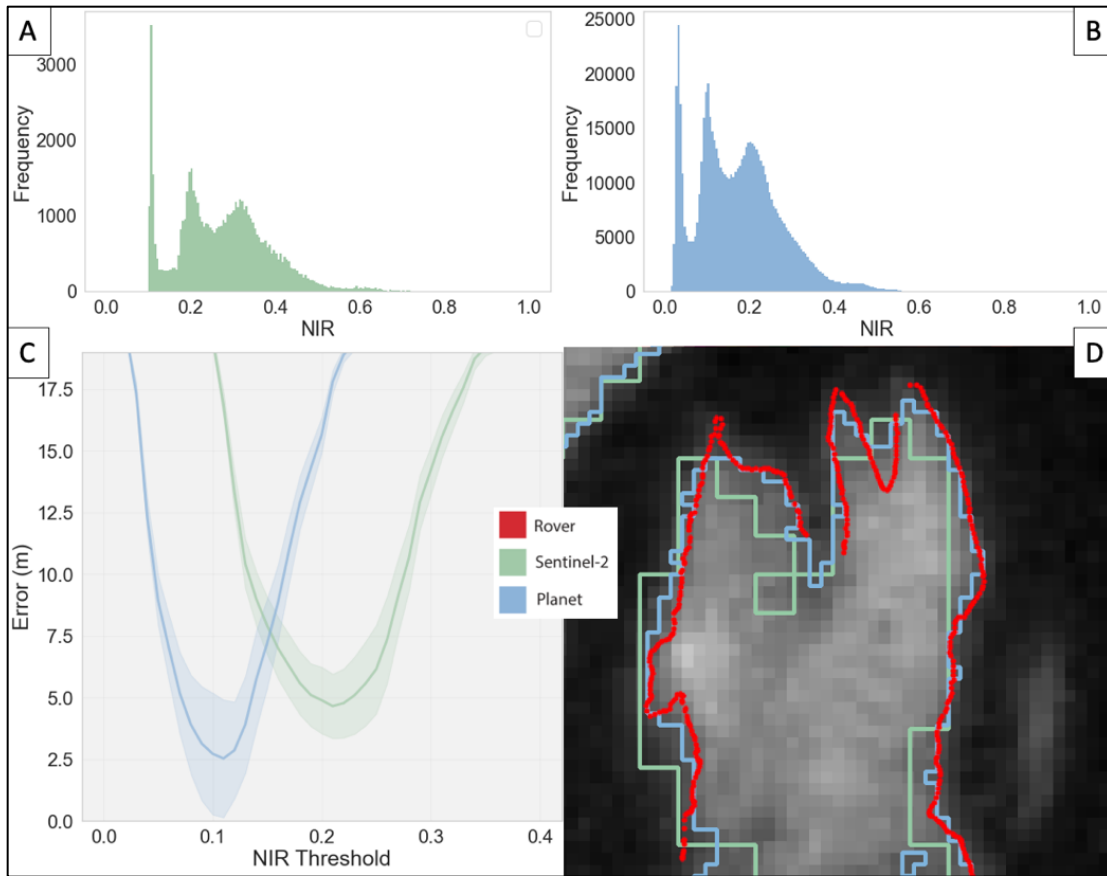


Figure 13 -A-B: Histograms over the Willamette River on November 18th. The survey varies slightly from that in September, which explains the differences in the histograms relative to September but emphasizes the necessity of normalized thresholding. C: A Planet NIR threshold at 0.11 and Sentinel-2 threshold at 0.21 led to the minimum error relative to the Rover of 2.5 and 4.5 meters, respectively.

4.3 Temporal Differences between Sentinel-2 and Planet

Although the differences in spatial accuracy between Planet and Sentinel-2 satellites are minor, our analysis reveals that Planet offers a superior temporal resolution. During our 172-day observation period on the Tanana River, we managed to identify 32 Planet observations and 25 Sentinel-2 observations that were usable. Here, 'usable' refers to an image that provides clear pixels covering at least 75% of the river's reach. While Planet collected imagery over 96% of the total time, only 18.5% of the year had usable

observations. In contrast, Sentinel-2 collected imagery for approximately 60% of the dates and yielded usable data 14.5% of the time. Despite a more evenly distributed record for Planet, Sentinel-2 had a higher percentage of usable observations, likely due to more consistent data quality, complete reach coverage, and a more effective cloud masking product.

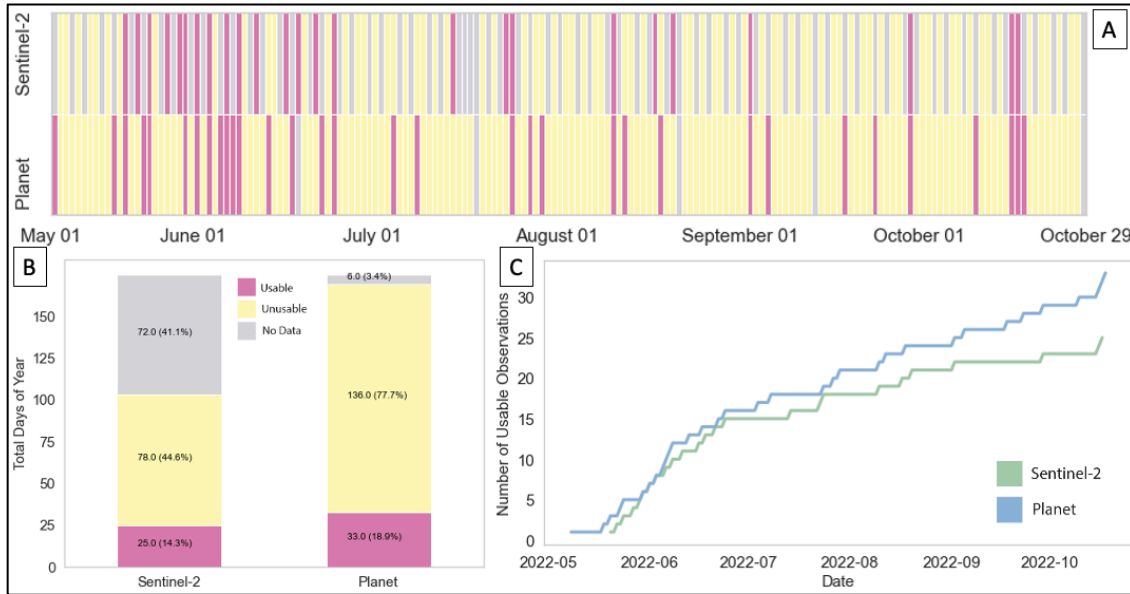


Figure 14 – A: The temporality of observations between Planet and Sentinel-2o over the Tanana River ice-free time period between May and October. The most usable observations in June coincided with clear sky days. B: Total cumulative observations; even though Planet has significantly more observations, a similar amount is usable, likely due to full area coverage and data quality. C: The rate of observations over time; Planet has more Usable observations spread throughout the year, while Sentinel-2 is mostly in June.

Over the Willamette River, we noticed that both Planet and Sentinel-2 satellites provided less temporal coverage than they did over the Tanana River. However, Planet outperformed Sentinel-2 in terms of the ratio of observed data to usable data. Throughout a 365-day period covering the 2022 calendar year, Planet made observations of the river on 253 days, equivalent to roughly 70% of the year, yet only 84 of these days, or about

23%, yielded usable data. This fell short of Planet's advertised daily observation frequency. A possible reason for the lower total observations was an inconsistency in data collection from April to June, during which there were the highest occurrences of no observations. Sentinel-2 had observation days totaling 145, or 40% of the year, with only 38 days, approximately 10.4%, deemed usable. Sentinel-2 observation schedule maintained a consistent schedule over the year. The distinct weather patterns of the Willamette River, characterized by extremes of either very cloudy or completely clear conditions, notably with clear dry summers and cloudy winters, contrasted with the Tanana River's mixed cloud patterns. These weather patterns influenced the variance of usable observations, as few clouds can cover significant portions of the reach of interest, rendering the data unusable for these purposes.

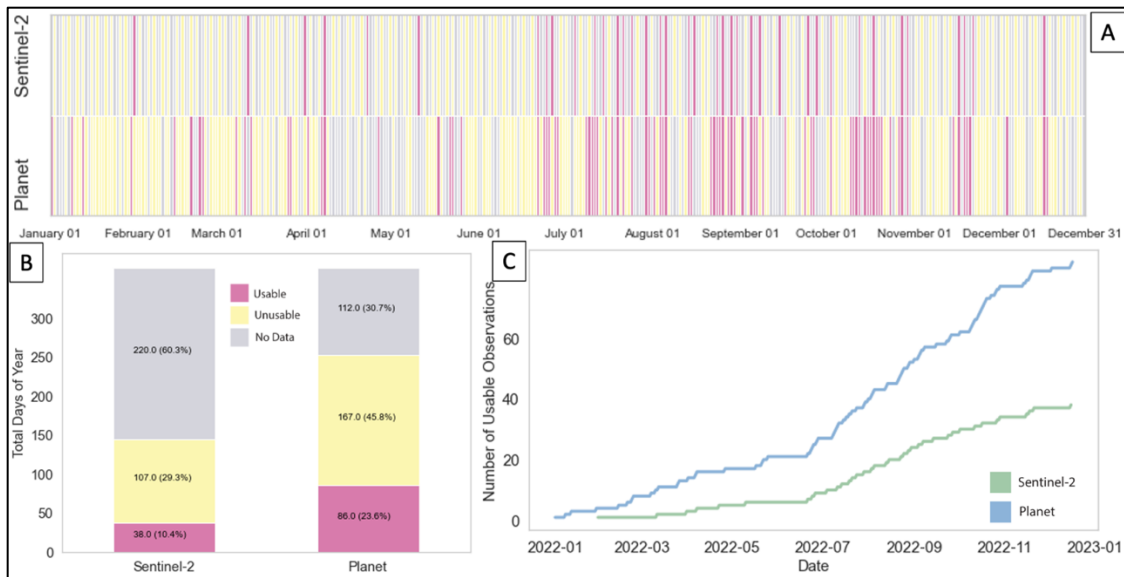


Figure 15 – A: 2022 temporal statistics for the Willamette River. While both platforms accumulate the most usable observations of the summer, predominately cloud-free months, Planet has more consistent data throughout the entire year. B: Planet has roughly 2/3 more observations than S2 but almost double the # of usable images. C: Planet has an increased usable observation over the entire year.

4.4 Comparison to Discharge Results

Our initial comparisons from inundation to effective width to discharge yield promising results. Over the Tanana River, at the reach scale, we find the Sentinel-2 data to correlate at 0.78 and Planet 0.83 (Figure 16). We not find considerable difference in correlation between Sentinel-2 and Planet. Further analysis of the results demonstrates that these derived calculations, it is clear that this data may not be consistent nor correct. It represents a strong seasonality in the data and can track discharge on a monthly scale but not daily nor from image to image. We see this inconsistency in both Planet and Sentinel-2 results. This likely is an error in the adaptive thresholding process and the scaling of the effective width data.

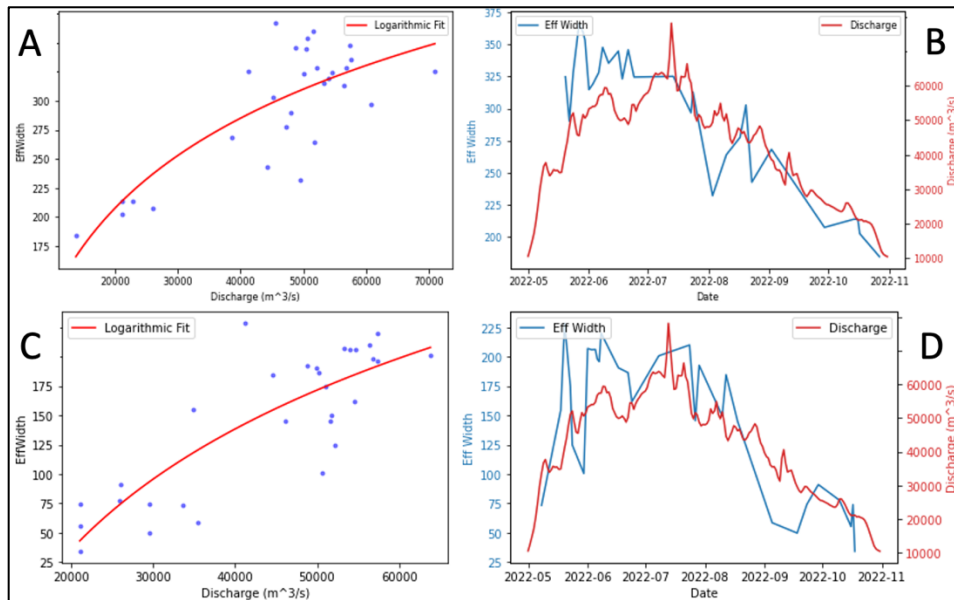


Figure 16 – Using the adaptive thresholding methodology to compare an effective width value derived from inundation against discharge. A logarithmic fit is applied to Sentinel-2 (A and B) data with a correlation of 0.78, while Planet has 0.83(C and D). This data, which does demonstrate a seasonality that reflects Discharge, is unreliable on smaller scales and returns invalid Effective Width (m) data.

The methodology currently creates a water map for every image and filters these images based on the % of clear pixels as defined in the methods. In Figure 17, you can

see the result of the filtered data. At the scale, the river appears to shrink in size with the usable water maps. This may be due to a changing threshold-based calculation that does not effectively model the river changes. It also appears that some images are missing data over the river (e.g., 08-03) and may be an artifact of the cloud masking step.



Figure 17 – An adaptive thresholding approach to identify discharge over the Tanana River. This example shows S2 imagery using a subset of dates over the reach of interest. Red days are not included in the analysis due to not enough clear data pixels over the reach.

DISCUSSION

5.1 Shoreline Walking Validation

Shoreline walking is a valuable tool for interpreting ground surface features captured by water classifications and, crucially, identifying surfaces they fail to detect. Our in-situ shoreline surveys on the Tanana River revealed that satellite classifications often do not perfectly capture the "wet vs. dry" transition, particularly on sandbars where discerning between inundated sediment and shallow mucky water is challenging using images alone. We found that cost-effective handheld GNSS receivers perform adequately compared to satellites, making shoreline walking an affordable ground validation approach. Our findings also indicate that human error in identifying and traversing shorelines exists but is generally limited to less than one meter. The ~one-meter error we find represents a combination of sensor and human error, implying that human error is relatively minor in numerous environments. However, we note that while the field surveying team attempted to walk the same shorelines, they exhibited slight discrepancies in their perception of shoreline locations, resulting in minor variations in data collection. Factors including individual step size and walking style may also marginally influence the data. The human survey error likely exists on different values for different shoreline types. A rocky shore, with even footing, is much easier and clearer to survey directly above the interface zone. We find supporting evidence that the difficulty of classification increases as grain size decreases (Cabezas-Rabadán et al., 2021). At the same time, a cut bank with slightly sandy surface at the water level is extremely difficult to monitor. Areas where remote sensing classifications struggle due to vegetation and adverse conditions often coincide with locations where traditional surveying techniques are most likely to encounter difficulties. By recognizing these potential error sources, we can work to mitigate their impact on our analysis, thereby enhancing the overall quality of shoreline data.

While shoreline ground truthing offers valuable insights into the performance of

classifications compared to actual ground conditions, it exhibits some inherent limitations. First, shoreline walking necessitates accessibility, as it can only be conducted in areas with safe water access, excluding remote and dangerous waterways. Remote sensing often uniquely enables analysis of remote regions, but many locations cannot be validated because of their inaccessibility. Furthermore, shoreline walking is physically demanding, involving navigation through challenging terrain and time-consuming data collection. Sampling bias may arise due to the varying ease and safety of different shorelines, potentially skewing sample areas. Locations with obstacles, such as large woody debris, mucky zones, or other blockages, were not viable for ground truthing. Moreover, the state of the river significantly impacts the accessibility of different shoreline types. During a falling river phase, muddy shores that were recently submerged pose challenges for sensors. In contrast, during a rising river phase, there is a more distinct demarcation between dry land and water, making it easier for satellites (and humans) to differentiate. Thus, while informative, shoreline ground-truthing should be considered in the context of these limitations. Ground truthing through traditional surveys is effective where possible; the limitations should not prevent the necessity of shoreline observations. High-precision Drones can be used as an additional ground-validating data source and bridge the gap between ground survey and remote sensing imagery (Levenson & Fonstad, 2022).

5.2 Planet and Sentinel-2 Spatial Resolution

Our analysis indicates that Planet exhibits a slight advantage in terms of functional spatial accuracy in comparison to field data, though in specific circumstances, it may perform less effectively than Sentinel-2 due to issues related to the radiometric resolution. If assessed solely based on stated resolution, one would expect Sentinel-2 to exhibit an error ranging between 5-10 meters (given its 10x10-meter pixel size/ground sampling distance) and Planet to exhibit an error between 1.5 and 3.7 meters (given its 3.125x3.125m pixel size/ground sampling distance). However, our findings reveal that Planet demonstrates a higher error in comparison to ground data than would be expected

given its stated pixel size, while Sentinel-2 does not.

Our findings indicate that inconsistencies between surface reflectance imagery from multiple dates affect both satellites; however, the effects are more pronounced and variable with Planet sensors, ultimately influencing the derived classifications. For both sensors, NDWI calculations were performed suitably on the Tanana River, where the high sediment content creates a distinct two-peaked histogram. In contrast, the Willamette River's green band sensitivity led to significant inconsistencies in sensor results, rendering NDWI ineffective. NIR performed well on the Willamette River as an alternative to using an NDWI adaptive technique. We observed that the maximum accuracy of both NDWI and NIR-derived classifications is similar for both sensors, although they are required in different river environments. Normalized thresholding for the effectiveness of both sensors remains challenging, even after atmospheric corrections, due to factors such as seasonality of observations, water quality, and other elements that limit the consistency of results.

It is important to acknowledge the variability in the threshold performance of the mentioned satellites. When considering the NDWI classifier accuracy figures, it is observed that Planet's derived accuracy metric exhibits a slightly wider range compared to Sentinel-2. This suggests that when utilizing Planet data, even if the desired threshold is missed, the impact on shoreline accuracy will be relatively lower compared to Sentinel-2. This characteristic becomes especially valuable considering the challenges associated with accessing ground data and the subsequent calibration of derived water maps. The observed discrepancies in the histogram peaks between water and land can be attributed to the differences in the NIR and Green bands utilized by the satellites. Notably, the disparity in peak values is more pronounced in Planet data than in Sentinel-2.

Our selected error metric for assessing the accuracy of satellite-derived water maps was the Mean Average Error. However, there are limitations to this method, particularly with respect to pixel size. Throughout our analysis, we consistently observed that the errors associated with Sentinel-2 were smaller than their corresponding pixel size. This finding suggests that the GNSS-identified shorelines and the shorelines derived

from water maps likely overlapped within the same pixel.

5.3 Planet and Sentinel-2 Temporal Resolution

While Planet's spatial resolution advantages over Sentinel-2 are minor, we do find that Planet exhibits a notable advantage in terms of temporal resolution. We observed that both satellites register a significantly higher number of observations over the Tanana compared to the Willamette, with Planet having over two times the useful photos of Sentinel-2 on the Willamette. Over the Willamette, Planet provides an observation on just under 70% of the total days; however, a significant portion of the missing data occurs over a two-week stretch. This is likely, however not for certain, because Planet data was unable to acquire ground geolocation due to extremely cloudy conditions, and thus does make the data available for download. Excluding this data gap, the relative frequency of observations compared to Sentinel-2 would be higher, but such data inconsistencies should be expected when working with Planet, given the constant changes to their satellite constellation. Despite Planet's higher temporal frequency, its imagery often fails to cover the entire ROI, limiting its usefulness and advantage over Sentinel-2, which almost always records the entire ROI. Additionally, Sentinel-2's larger swath size increases the likelihood of observing the entire reach within one image, compared to Planet, where multiple images may need to be composited. Lastly, we note that Sentinel-2 observation patterns over the Willamette and Tanana are more consistent (i.e., acquiring data approximately two-three days over the Tanana and once every three days over the Willamette) compared to the Planet's less predictable albeit increased (daily to sub-daily) sampling frequency. It is important to note that the greatest source in the usability of observations stems from the cloud masks generated by each respective platform. Improved cloud masking would likely increase the percentage of usable data for each platform and each respective reach.

5.4 Planet and Sentinel-2 Accessibility

A crucial advantage of Sentinel-2 imagery over Planet lies in its global accessibility. Sentinel-2 imagery is publicly available and easily accessible, as it is natively integrated within Google Earth Engine (GEE), promoting open-source code collaboration and constantly improving the Sentinel-2 product (Gorelick et al., 2017). In contrast, Planet imagery is not freely available, although partnerships with space agencies and some universities make the data more broadly accessible to researchers. Planet imagery can be best accessed through Planet Explorer, which allows batch ordering and querying, but the system does not integrate as seamlessly with GEE as Sentinel-2 does. Moreover, the mosaicking and cloud masking products provided by Sentinel-2 and accessible can be performed tasks online, while Planet's (rapidly improving) mosaicking and masking are at times performed locally using the necessary UDM files. Although the synergy between the Landsat and Sentinel-2 systems has been extensively studied, including the development of a harmonized data product (Claverie et al., 2018b; Vanhellemont, 2020), less attention has been given to the differences and potential integration between Sentinel-2 and Planet data. This study shows that the systems are more similar in spatial accuracy, but there is not a clear need as Planet has superior temporal data collection. Planet may be used as a complement to Sentinel-2 data when in need of increased temporal resolution data.

5.5 Derived Comparison to Discharge

Interestingly, our analysis of the comparison between remotely sensed effective width and discharge data highlights that scaling up classification methods based solely on shoreline comparison may not be the most effective approach. The relationship between inundation and discharge is neither consistent nor uniform, as subsurface hydrodynamics remain undetectable by optical satellites.

Using NDWI-derived water classifications to minimize error relative to in situ shoreline observations may not optimally classify changes occurring throughout the entire river system. In challenging locations like the Tanana River, where sediment

content and watercolor vary substantially, it is crucial not to overemphasize the water-sediment interface in calculations. In a previous analysis of lake areas using Planet and Sentinel-2, the two platforms produced minimal differences in area detection when both platforms identified lakes present, but Planet identified significantly more small lakes (Mullen et al., 2023). Differences in shoreline identification become less significant as the region of interest (ROI) expands, emphasizing the need to capture broader hydrological changes rather than scaling to specific shorelines that may not be representative of the entire region. This relationship varies across rivers, and optimizing these relationships could involve focusing on ROIs smaller than the reach scale. Aligning classification with ground-truth observations prioritizes a limited period over the study duration, which may not necessarily result in a superior rating curve for the river.

To combat this issue in our discharge calculation, we tried calculating inundation from a smaller portion of the river (Figure 18). We used the same methods for the full-reach normalization. We chose this area because of the islands with long sandy bars that are highly variable to slight changes in discharge. For this analysis, we manually filtered to only use 100% clear images of the river so to assure the quality of the input data. We can conclude that this specific methodology in our research fails to relate to discharge, but that does not mean other adaptive thresholding techniques may not identify the phenomena more accurately.

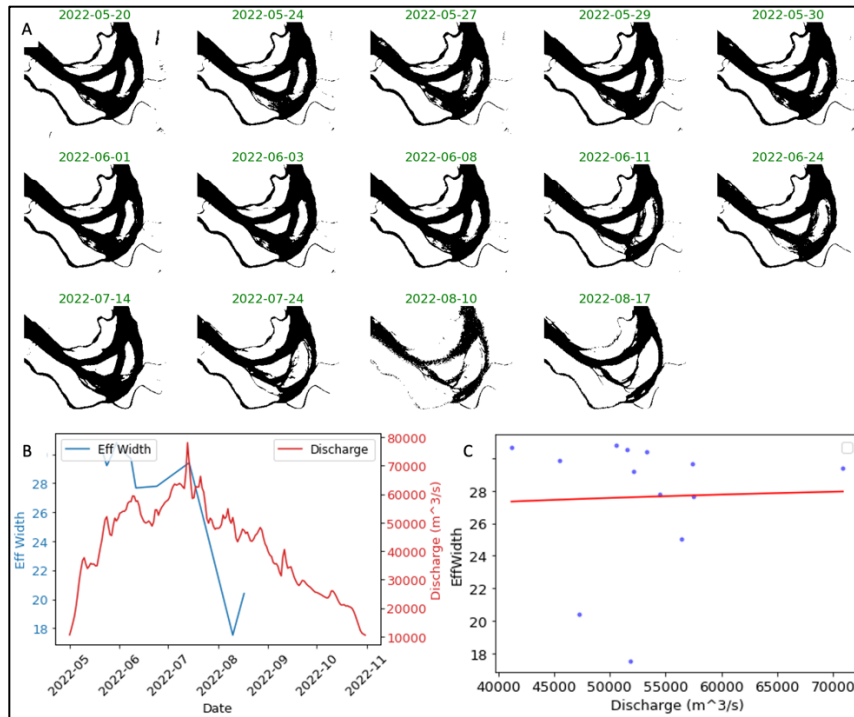


Figure 18 – A: Example of using clear Sentinel-2 images over the Tanana River to map inundation on a subset of the reach, B: Discharge compared with effective width returns incorrect data, and C: The correlation is non-existent between Discharge and Effective Width. The error found from these images has more to do with the classification scheme than input quality.

While our inundation comparisons are not accurate at this current phase, they show promise in capturing the seasonality of the river system. Previous scientific studies (Cooley et al., 2017) demonstrate that an adaptive threshold can be used to relate to discharge. However, the discharge comparison could be increased with the inclusion of physical hydrology measurements. The recent satellite remote sensing study of river discharge (Bjerklie et al., 2023) recommends that optical data is just one piece of the larger puzzle in deriving discharge from space. Optical can still be useful in identifying phenomena and contributing to data from the newly launched NASA Satellite Water and Ocean Topography (S.W.O.T.) satellite.

5.6 Discussion Summary

In summary, while Planet does offer slightly higher spatial and temporal resolution than Sentinel-2, the differences are small enough that, when combined with data accessibility issues, Sentinel-2 data is likely suitable for many analyses requiring a moderate combination of high spatial and temporal resolution. For consistent and accurate classification, Sentinel-2 is likely preferable to Planet. The difference in accessibility is particularly notable by the ease of Sentinel-2 data. Planet's main advantage is its temporal resolution; if approximately weekly observations are acceptable, the ease of accessibility and improved, more consistent radiometric resolution make Sentinel-2 a better option for most applications.

CONCLUSIONS

In conclusion, our study emphasizes the usefulness of ground observations in understanding the functional accuracy of satellite-derived water classification. Despite the accessibility and physical challenges, shoreline walking presents a cost-effective approach enabled by affordable handheld GNSS receivers. We find that in-situ observations can help make remote sensing classification decisions to more accurately depict that actual ground story. It's important to note, however, that the state of the river—whether rising or falling—can significantly influence the accuracy of shoreline classification. Our investigation also revealed the substantial variability of threshold-based classification techniques across different rivers and within each watershed's seasonal patterns. Therefore, the need for adaptive thresholding proves that reflectance value comparison is paramount. Without this normalization, data inconsistencies could lead to inaccurate interpretations. Ultimately, this comparative analysis refines our techniques and guides informed decision-making regarding satellite data sources and classification approaches, bolstering our capacity to monitor and study dynamic river systems.

We found that the spatial accuracy discrepancies between Sentinel-2 and Planet data were less pronounced than initially assumed. While Sentinel-2 data displayed a minor reduction in functional spatial accuracy and a less advanced temporal resolution, it remains an appealing option due to its ease of use and widespread accessibility. However, the true strength of Planet data lies in its superior temporal observation capabilities, which offer significant advantages in certain applications.

This research invites additional questions on the capabilities of high-resolution sensors. There is compelling potential in the development of a hybrid product that combines Sentinel-2 and Planet data. This product would use Sentinel-2 as the foundational base, with the option to scale up to Planet in specific scenarios. Achieving this would likely require consistency in cloud masking across both platforms.

Additional questions include assessing the effectiveness of machine learning

techniques in identifying the shoreline interface or in predicting discharge measurements (Claverie et al., 2018b; Hondula et al., 2021; McAllister et al., 2022; Mullen et al., 2023; Toure et al., 2019).

The exploration of spatial and temporal accuracy for remote sensing of lakes remains a critical area of study. Additionally, analogous studies could evaluate the functional accuracy of commercial radar data, which has the advantage of being unaffected by cloud cover.

REFERENCES

- Allen, G. H., & Pavelsky, T. M. (2018). Global extent of rivers and streams. *Science*, *361*(6402), 585–588. <https://doi.org/10.1126/science.aat0636>
- Alsdorf, D. E., Rodríguez, E., & Lettenmaier, D. P. (2007). Measuring surface water from space. *Reviews of Geophysics*, *45*(2). <https://doi.org/10.1029/2006RG000197>
- Aragon, B., Houborg, R., Tu, K., Fisher, J. B., & McCabe, M. (2018). CubeSats Enable High Spatiotemporal Retrievals of Crop-Water Use for Precision Agriculture. *Remote Sensing*, *10*(12), Article 12. <https://doi.org/10.3390/rs10121867>
- Belward, A. S., & Skøien, J. O. (2015). Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*, *103*, 115–128. <https://doi.org/10.1016/j.isprsjprs.2014.03.009>
- Bjerklie, D. M., Durand, M., Lenoir, J., Dudley, R. W., Birkett, C. M., Jones, J. W., & Harlan, M. (2023). Satellite remote sensing of river discharge: A framework for assessing the accuracy of discharge estimates made from satellite remote sensing observations. *Journal of Applied Remote Sensing*, *17*(1), 014520. <https://doi.org/10.1117/1.JRS.17.014520>
- Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., & Tait, A. M. (2022). Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, *9*(1), Article 1. <https://doi.org/10.1038/s41597-022-01307-4>
- Cabezas-Rabadán, C., Pardo-Pascual, J. E., & Palomar-Vázquez, J. (2021). Characterizing the Relationship between the Sediment Grain Size and the Shoreline Variability Defined from Sentinel-2 Derived Shorelines. *Remote Sensing*, *13*(14), Article 14. <https://doi.org/10.3390/rs13142829>
- Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., & Justice, C. (2018a). 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>
- Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., & Justice, C. (2018b). 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., Smith, L., 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>
- Delwart, S. (n.d.). *ESA Standard Document. 1.*
- DeVries, B., Huang, C., Lang, M. W., Jones, J. W., Huang, W., Creed, I. F., & Carroll, M. L. (2017). Automated Quantification of Surface Water Inundation in Wetlands Using Optical Satellite Imagery. *Remote Sensing*, 9(8), Article 8. <https://doi.org/10.3390/rs9080807>
- Frazier, A. E., & Hemingway, B. L. (2021). A Technical Review of Planet Smallsat Data: Practical Considerations for Processing and Using PlanetScope Imagery. *Remote Sensing*, 13(19), Article 19. <https://doi.org/10.3390/rs13193930>
- Gleason, C., & Durand, M. (2020). Remote Sensing of River Discharge: A Review and a Framing for the Discipline. *Remote Sensing*, 12(7), 1107. <https://doi.org/10.3390/rs12071107>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Harlan, M. E., Gleason, C. J., Flores, J. A., Langhorst, T. M., & Roy, S. (2023). Mapping and characterizing Arctic beaded streams through high resolution satellite imagery. *Remote Sensing of Environment*, 285, 113378. <https://doi.org/10.1016/j.rse.2022.113378>
- Hondula, K. L., DeVries, B., Jones, C. N., & Palmer, M. A. (2021). Effects of Using High Resolution Satellite-Based Inundation Time Series to Estimate Methane Fluxes From Forested Wetlands. *Geophysical Research Letters*, 48(6), e2021GL092556. <https://doi.org/10.1029/2021GL092556>
- Huang, C., Chen, Y., Zhang, S., & Wu, J. (2018). Detecting, Extracting, and Monitoring Surface Water From Space Using Optical Sensors: A Review. *Reviews of Geophysics*, 56(2), 333–360. <https://doi.org/10.1029/2018RG000598>
- Islam, Md. M., & Ahamed, T. (2023). Development of a near-infrared band derived water indices algorithm for rapid flash flood inundation mapping from sentinel-2 remote sensing datasets. *Asia-Pacific Journal of Regional Science*. <https://doi.org/10.1007/s41685-023-00288-5>
- Jones, J. W. (2015). Efficient Wetland Surface Water Detection and Monitoring via Landsat: Comparison with in situ Data from the Everglades Depth Estimation Network. *Remote Sensing*, 7(9), Article 9. <https://doi.org/10.3390/rs70912503>

- Kaiser, S., Grosse, G., Boike, J., & Langer, M. (2021). Monitoring the Transformation of Arctic Landscapes: Automated Shoreline Change Detection of Lakes Using Very High Resolution Imagery. *Remote Sensing*, *13*(14), Article 14. <https://doi.org/10.3390/rs13142802>
- Lerch, R. N., Kitchen, N. R., Kremer, R. J., Donald, W. W., Alberts, E. E., Sadler, E. J., Sudduth, K. A., Myers, D. B., & Ghidry, F. (2005). Development of a conservation-oriented precision agriculture system: Water and soil quality assessment. *Journal of Soil and Water Conservation*, *60*(6), 411–421.
- Levenson, E. S., & Fonstad, M. A. (2022). Characterizing coarse sediment grain size variability along the upper Sandy River, Oregon, via UAV remote sensing. *Geomorphology*, *417*, 108447. <https://doi.org/10.1016/j.geomorph.2022.108447>
- Li, Y., Wang, Q., Wu, C., Zhao, S., Xu, X., Wang, Y., & Huang, C. (2012). Estimation of Chlorophyll a Concentration Using NIR/Red Bands of MERIS and Classification Procedure in Inland Turbid Water. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(3), 988–997. <https://doi.org/10.1109/TGRS.2011.2163199>
- Liu, Z., Yao, Z., & Wang, R. (2016). Assessing methods of identifying open water bodies using Landsat 8 OLI imagery. *Environmental Earth Sciences*, *75*. <https://doi.org/10.1007/s12665-016-5686-2>
- Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., & Gascon, F. (2017). Sen2Cor for Sentinel-2. *Image and Signal Processing for Remote Sensing XXIII*, *10427*, 37–48. <https://doi.org/10.1117/12.2278218>
- McAllister, E., Payo, A., Novellino, A., Dolphin, T., & Medina-Lopez, E. (2022). Multispectral satellite imagery and machine learning for the extraction of shoreline indicators. *Coastal Engineering*, *174*, 104102. <https://doi.org/10.1016/j.coastaleng.2022.104102>
- 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>
- Ming Wang, Z. L. (2020). Comparisons of Image Cloud Detection Effect based on Sentinel-2 Bands/Products. *Remote Sensing Technology and Application*, *35*(5), 1167–1177. <https://doi.org/10.11873/j.issn.1004-0323.2020.5.1167>
- Mondejar, J. P., & Tongco, A. F. (2019). Near infrared band of Landsat 8 as water index: A case study around Cordova and Lapu-Lapu City, Cebu, Philippines. *Sustainable Environment Research*, *29*(1), 16. <https://doi.org/10.1186/s42834-019-0016-5>

- Mullen, A. L., Watts, J. D., Rogers, B. M., Carroll, M. L., Elder, C. D., Noomah, J., Williams, Z., Caraballo-Vega, J. A., Bredder, A., Rickenbaugh, E., Levenson, E., Cooley, S. W., Hung, J. K. Y., Fiske, G., Potter, S., Yang, Y., Miller, C. E., Natali, S. M., Douglas, T. A., & Kyzivat, E. D. (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>
- 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>
- Pilartes-Congo, J. A. (2022). *Evaluation of Different GNSS Solutions and SfM Software Workflows for Surveying Shorelines and Remote Areas using UAS* [M.Sc., Texas A&M University - Corpus Christi]. <https://www.proquest.com/docview/2681480555/abstract/C85DEF31876C4C02PQ/1>
- Pitcher, L. H., Smith, L. C., Cooley, S. W., Zaino, A., Carlson, R., Pettit, J., Gleason, C. J., Minear, J. T., Fayne, J. V., Willis, M. J., Hansen, J. S., Easterday, K. J., Harlan, M. E., Langhorst, T., Topp, S. N., Dolan, W., Kyzivat, E. D., Pietroniro, A., Marsh, P., ... Pavelsky, T. M. (2020). Advancing Field-Based GNSS Surveying for Validation of Remotely Sensed Water Surface Elevation Products. *Frontiers in Earth Science*, *8*, 278. <https://doi.org/10.3389/feart.2020.00278>
- Planet Monitoring—Satellite Imagery and Monitoring*. (n.d.). Planet. Retrieved December 6, 2022, from <https://www.planet.com/products/monitoring/>
- PlanetScope*. (n.d.). Retrieved May 30, 2023, from <https://developers.planet.com/docs/data/planetscope/>
- Qayyum, N., Ghuffar, S., Ahmad, H., Yousaf, A., & Shahid, I. (2020). Glacial Lakes Mapping Using Multi Satellite PlanetScope Imagery and Deep Learning. *ISPRS International Journal of Geo-Information*, *9*(10), 560. <https://doi.org/10.3390/ijgi9100560>
- Shiklomanov, A. I., Lammers, R. B., & Vörösmarty, C. J. (2002). Widespread decline in hydrological monitoring threatens Pan-Arctic Research. *Eos, Transactions American Geophysical Union*, *83*(2), 13. <https://doi.org/10.1029/2002EO000007>
- Sogno, P., Klein, I., & Kuenzer, C. (2022). Remote Sensing of Surface Water Dynamics in the Context of Global Change—A Review. *Remote Sensing*, *14*(10), Article 10. <https://doi.org/10.3390/rs14102475>

- Toure, S., Diop, O., Kpalma, K., & Maiga, A. (2019). Shoreline Detection using Optical Remote Sensing: A Review. *ISPRS International Journal of Geo-Information*, 8(2), 75. <https://doi.org/10.3390/ijgi8020075>
- UDM 2. (n.d.). Retrieved May 23, 2023, from <https://developers.planet.com/docs/data/udm-2/#new-metadata-fields>
- Vanhellemont, Q. (2020). Automated water surface temperature retrieval from Landsat 8/TIRS. *Remote Sensing of Environment*, 237, 111518. <https://doi.org/10.1016/j.rse.2019.111518>
- Zhang, F., Li, J., Shen, Q., Ye, H., Wang, S., & lu, Z. (2018). A simple automated dynamic threshold extraction method for the classification of large water bodies from landsat-8 OLI water index images. *International Journal of Remote Sensing*, 39, 3429–3451. <https://doi.org/10.1080/01431161.2018.1444292>
- Zhu, Z., Wulder, M. A., Roy, D. P., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., Healey, S. P., Schaaf, C., Hostert, P., Strobl, P., Pekel, J.-F., Lymburner, L., Pahlevan, N., & Scambos, T. A. (2019). Benefits of the free and open Landsat data policy. *Remote Sensing of Environment*, 224, 382–385. <https://doi.org/10.1016/j.rse.2019.02.016>