

Toward an efficient framework for remote sensing of river bathymetry: Comparing sensors and algorithms on an inaccessible proglacial river in Alaska

Carl J. Legleiter^a, Christina M. Leonard^b, Paul A. Burger^c, Addison G. Pletcher^c, Paul J. Kinzel^a

^a U.S. Geological Survey, Observing Systems Division, 4620 Technology Drive, Suite 400, Golden, 80403, CO, USA

^b National Park Service, Water Resources Division, 1221 Oakridge Drive, Fort Collins, 80525, CO, USA

^c National Park Service, Alaska Regional Office, 240 West Fifth Avenue, Anchorage, 99501, AK, USA

ARTICLE INFO

Dataset link: <https://doi.org/10.5066/P14F2LE9>

Keywords:

Proglacial river
Depth
Bathymetry
Remote sensing
Aerial orthophoto
Multispectral satellite image

ABSTRACT

Remote sensing can provide reliable information on river depths and this approach might be particularly valuable in areas that are difficult to survey via conventional field methods. In this study, we assessed the potential to map the bathymetry of an inaccessible proglacial river in Alaska from both aerial orthophotos and a multispectral satellite image. In addition, we evaluated a variety of depth retrieval algorithms with different input data requirements, including some methods that require field measurements of water depth for calibration and other techniques that can be applied even when such field data are not available. These approaches might enable more efficient use of remote sensing methods by resource management agencies. Our results suggest that bathymetric mapping along the turquoise-colored river we examined was not only feasible but highly accurate (R^2 up to 0.94) for both types of image data. Algorithms that use paired observations of depth and reflectance to train depth retrieval models were the most accurate, with errors on the order of 15%–20% and little or no bias. Alternative techniques based on hydraulic and statistical concepts also led to strong agreement between predicted and observed depths but were more susceptible to systematic biases toward under- or over-estimation of depth. In contrast to clear-flowing streams, bathymetric mapping in this environment was enabled by a direct relationship between the depth and brightness of the water due to scattering by suspended sediment. In selecting an appropriate depth retrieval method, a compromise might need to be reached between the level of field effort invested and the accuracy of the resulting image-derived bathymetry. Standalone software for implementing these techniques is freely available.

1. Introduction

The increasing use of image-based techniques for characterizing fluvial systems led Marcus and Fonstad (2010) to herald remote sensing as a new subdiscipline of the river sciences. The extensive, high-resolution mapping enabled by this approach is central to the riverscape concept (Carboneau et al., 2012), and Piégay et al. (2020) described how remote sensing could help advance our understanding of human impacts on rivers. By expanding the area of coverage and reducing the need for field-based surveys, remote sensing might also facilitate water management, with potentially substantial improvements in the cost-effectiveness of monitoring efforts (Tamminga et al., 2025). Realizing these objectives will require an ongoing, sustained effort to rigorously test remote sensing methods across a range of river environments,

assess their capabilities and limitations, and make these tools available to a broad user community.

In this study, we focus on a specific application of remote sensing to rivers, inferring water depth from passive optical image data that record how the amount of reflected solar radiation varies both spatially and as a function of wavelength. In this study, we use the terms depth and bathymetry interchangeably, whereas other authors might equate bathymetry with channel bed elevation. For the latter definition, information on both water surface elevation and water depth would be required to obtain bathymetry. Herein, we focus on measuring depth, with the important caveat that obtaining bed elevations would require an additional source of information on water level. In any case, the motivations for mapping river bathymetry are myriad and stem from

* Corresponding author.

E-mail address: cjl@usgs.gov (C.J. Legleiter).

<https://doi.org/10.1016/j.geomorph.2025.110140>

Received 7 August 2025; Received in revised form 15 December 2025; Accepted 18 December 2025

Available online 24 December 2025

0169-555X/© 2026 The US Geological Survey. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

depth's fundamental role as an important aspect of habitat suitability for fish and other aquatic organisms and a key control on hydraulics and sediment transport and thus channel change. In addition, depth information is critical at a more pragmatic level: can people or vehicles safely traverse a given stream? Similarly, what hazards might a particular waterway pose to boats and other vessels?

Extensive reviews of the development of spectrally based depth retrieval methods over the past fifty years, along with a discussion of alternative non-contact means of measuring depth, are provided elsewhere (e.g., He et al., 2024). Herein, we restrict our attention to the practical implementation of this approach in a rugged environment where the advantages of remote sensing relative to conventional field methods are especially pronounced: Alaska, the largest and least accessible state in the U.S. As in other high-latitude regions, the effects of accelerated climate change are acutely felt along Alaskan rivers. For example, Michalek et al. (2025) projected a shift toward earlier, higher magnitude flood peaks in the coming years and Blaskey et al. (2023) showed that discharges have already increased in winter months. Wohl and Scamardo (2022) showed that seasonal aufeis, which forms when groundwater springs continue to provide water to the surface after the river is covered by ice, promotes lateral channel migration and more heterogeneous habitat conditions, but cautioned that these dynamics could be compromised as temperatures rise. Similarly, Rowland et al. (2023) concluded that bank erosion rates are likely to increase on large rivers across the Arctic as the thawing of permafrost reduces bank stability. Geochemistry is also affected: permafrost thaw can accentuate chemical weathering, reduction of iron, and groundwater transport of metals into streams. These processes cause rivers to rust, turning orange in color, and have decreased the abundance and diversity of fish and macroinvertebrates (O'Donnell et al., 2024). The variety and severity of these impacts create a compelling need to monitor rivers throughout Alaska, but the state's vast extent makes field-based data collection logistically challenging, at times dangerous, and often prohibitively expensive (Conaway et al., 2019). All of these factors point to remote sensing as a possible alternative for monitoring streams and numerous investigations have begun to explore this approach (e.g., Bjerklie et al., 2018; Käb et al., 2019; Harlan et al., 2023), including a recent study of the Peace River by Tamminga et al. (2025) that employed some of the same depth retrieval methods we examine herein.

In this study, we build upon this previous research by assessing the potential to characterize the bathymetry of a proglacial river in south-central Alaska. Although spectrally based depth retrieval algorithms have been applied to numerous clear-flowing streams (e.g., Legleiter and Harrison, 2024, and references cited therein), we are not aware of any prior applications of these methods to rivers immediately downstream from glaciers or glacier-dammed lakes. Water in these environments often has a unique turquoise color due to the presence of fine-grained sediment referred to as glacial flour and although this material might be conducive to estimating suspended sediment concentrations, as demonstrated by Vowels et al. (2025), high levels of turbidity could complicate if not preclude bathymetric mapping. To evaluate the feasibility of this approach, we considered two distinct types of remotely sensed data: orthophotos acquired from an aircraft and multispectral images from a satellite. Both kinds of image have proven useful for depth retrieval in more temperate, clear-flowing rivers (Legleiter and Overstreet, 2012; Legleiter, 2013; Legleiter and Harrison, 2019) but their suitability for mapping proglacial streams has not been tested.

Another primary objective is to further vet a suite of depth retrieval algorithms that vary in terms of their input data requirements. The rationale for this analysis is to identify ways to reduce the level of field effort needed to support remote sensing of river bathymetry. Management agencies like the National Park Service (NPS) are tasked with monitoring thousands of km of rivers throughout Alaska (NPS, 2025) and yet difficult access and short field seasons often make this mission difficult to fulfill; any improvements in the efficiency of data collection

would be valuable. The most well-established methods for inferring depth from passive optical image data are fundamentally empirical in nature and rely upon paired observations of depth and reflectance. These techniques include optimal band ratio analysis (OBRA, Legleiter et al., 2009) and a more recent, machine learning (ML)-based method called hyperspectral imaging of river bathymetry using an ensemble of regression trees (HIRBERT, Legleiter et al., 2025b). Flow resistance equation-based imaging of river depths (FREEBIRD) is an alternative approach that incorporates concepts of open channel flow to obviate the need for field measurements of depth (Legleiter, 2015; Leonard et al., 2017). Similarly, image-to-depth quantile transformation (IDQT) provides greater flexibility by linking the frequency distribution of pixel values to that of depth, which relaxes the requirement for precise co-registration between field surveys and image data and makes the technique well-suited to dynamic channels (Legleiter, 2016; Dilbone et al., 2018; Jiang and Rutherford, 2024; Tamminga et al., 2025). In this study, we apply all these techniques to the same Alaskan river to compare their performance and assess whether gains in efficiency might come at the cost of diminished accuracy.

The specific aims of this paper are to:

1. Assess the feasibility of spectrally based depth retrieval in a proglacial river environment with distinctive optical characteristics.
2. Evaluate the utility of two different kinds of passive optical data for this purpose: orthophotos from aircraft and multispectral satellite images.
3. Compare the performance of several depth retrieval algorithms that require varying amounts and types of field data as input, including both methods that rely upon direct measurements of water depth for calibrating image-derived estimates and techniques that can be applied when such field data are not available.

2. Methods

2.1. Study area

This study focused on the uppermost portion of the Mulchatna River, a proglacial stream located in a remote, inaccessible area of south-central Alaska's Lake Clark National Park and Preserve (Fig. 1). The Mulchatna River drains Turquoise Lake, a post-glacial lake that likely formed during the Illiuk Stade approximately 11,000–16,000 years ago (Lanik et al., 2021). Our study area encompassed a 3.84-km segment of the river starting near the outlet from Turquoise Lake and extending downstream to the first substantial tributary entering from the south. Field data were collected within three distinct reaches, numbered sequentially from upstream to downstream (Table 1), where the channel morphology was relatively simple and most of the flow was focused in a well-defined thalweg. The river's distinctive color is shown in Fig. 2. In reaches 1 and 2, the Mulchatna River flows through a Late Wisconsin glacial moraine with an anastomosing planform and a substrate ranging from gravel to cobble-sized material with occasional boulders, typical of morainal deposits. In this area, the channel lacks distinctive banks and a well-defined floodplain and instead gradually transitions from open water to riverine wetlands. The adjacent uplands are characterized by kettle lakes and associated wetlands. In reach 3, the river flows through glacial outwash deposits of Late Wisconsin age and an alluvial fan formed by the tributary that enters at the lower end of our study area. The channel in this area is single-thread, with well-defined banks, floodplain, and active gravel bars. The surrounding uplands feature a broad alluvial fan with numerous distributary channels to the south and kettle lakes with wetlands to the north. The substrate within reach 3 is composed of gravel-size material at the tributary confluence, with cobble and boulders farther downstream (NPS, 2022).

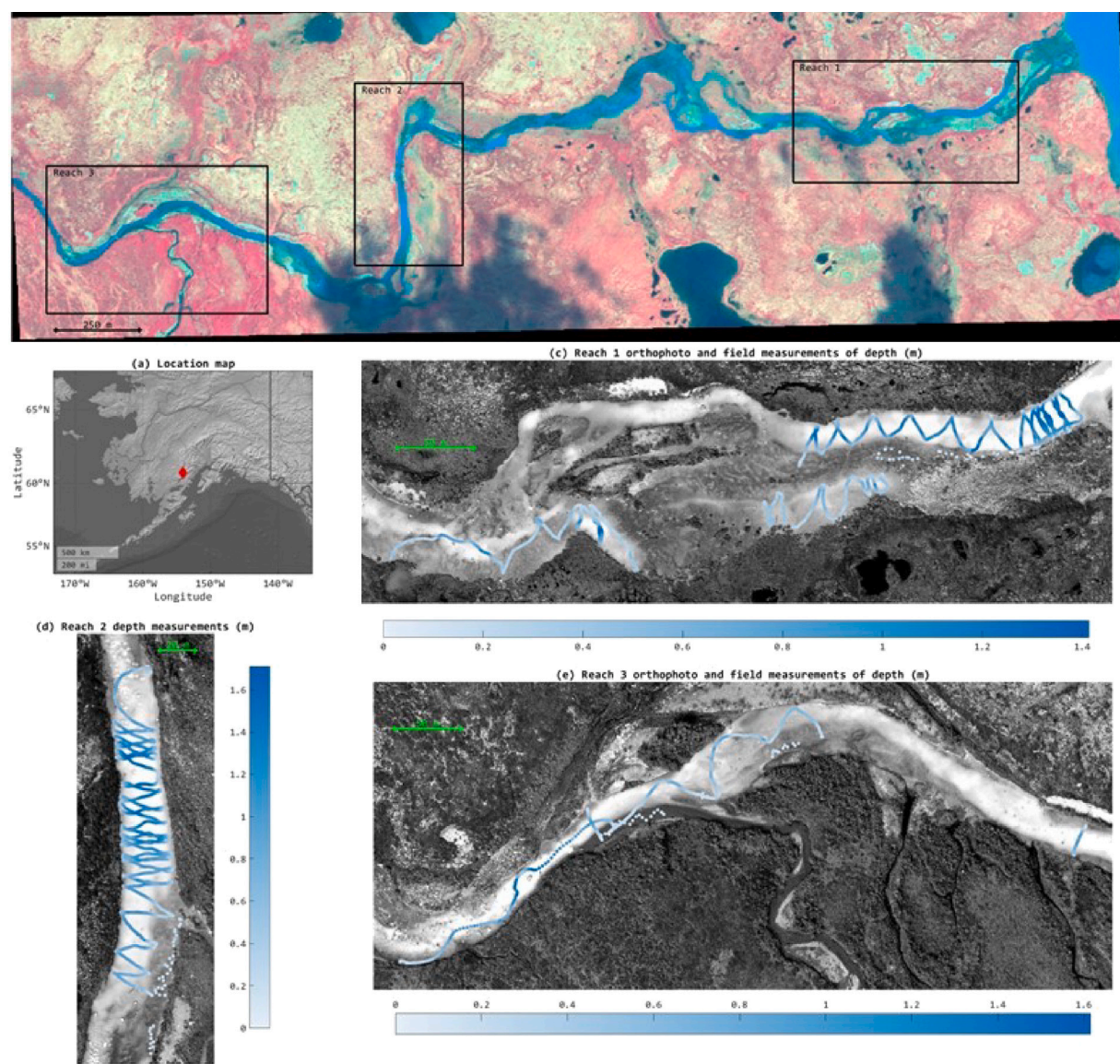


Fig. 1. (a) Map showing location of the Mulchatna River study area in Alaska. (b) WorldView2 (WV) satellite image of the entire study area with rectangles outlining each of the three reaches, shown as a false color composite with a near-infrared band displayed as red, a red band displayed as green, and a green band displayed as blue. (c–e) Orthomosaic subsets for each reach with field measurements of water depth overlain. Within the text and subsequent figures, we refer to the entire study area as reach 4 and the shorter reaches labeled 1–3 are thus sections of reach 4. Basemap in (a) made with (Natural Earth, 2025).

Table 1
Characteristics of study reaches along the Mulchatna River. The reach labeled as 4 encompasses the entire study area, of which the shorter reaches labeled 1–3 are sections of the larger reach.

Reach	Length (m)	Slope	Depth measurements
1	110	0.0035	1696
2	234	0.0004	1529
3	662	0.0026	603
4	3844	0.0021	3828

2.2. Field measurements

All of the field measurements used in this study are available through a data release (Legleiter et al., 2025a), including a metadata file that provides more detailed information on how the data were collected and processed than the brief summary herein.

2.2.1. Water depth

Water depths within each reach were measured July 12–14, 2024, using a SonTek RS5 (Xylem, San Diego, CA) acoustic Doppler current

profiler (ADCP) mounted on a trimaran; a laptop with the SonTek RSQ software was used to control the instrument and record data. The ADCP was deployed by towing the trimaran downstream along a series of zig-zag transects that extended from one bank to the other at an oblique angle. In addition, data were collected along two distinct cross sections (XS) within reach 3 by stretching a rope across the river and then pulling the trimaran back and forth across the channel. For each XS, we made a total of four passes and used these data to fit an overall mean cross section (MCS). The individual measurements were



Fig. 2. Photo taken while collecting field data along reach 1 showing the distinctive turquoise color of the water in the Mulchatna River, Alaska. Photograph by Christina Leonard, National Park Service, 2024.

then projected onto this MCS, which was subdivided into lateral, cross-stream bins 0.5 m in width. All the measurements within each bin were averaged to obtain a single mean depth representative of that 0.5 m-wide portion of the MCS. To complement the ADCP data, additional depth measurements were made by wading shallow areas along the margins of the channel and using a Trimble Geo7X (Westminster, CO) Global Navigation Satellite System (GNSS) receiver to record the location of each depth measurement. These data were originally collected in geographic coordinates but were projected to the same coordinate system as the remotely sensed data so that they could be used to train and test various depth retrieval algorithms. The depth measurements made within each reach are shown in Fig. 1c–e and summarized in Fig. 3. For the study area as a whole, the mean and standard deviation of the measured depths were 0.731 m and 0.405 m, respectively, with a maximum of 2.33 m occurring within reach 3.

2.2.2. River discharge

In addition to water depth, the ADCP also measured flow velocity throughout the water column, and we used the data collected along the two XS in reach 3 to calculate river discharge. This analysis was performed using the U.S. Geological Survey (USGS) QRev software package (Mueller, 2020), which also provided an estimate of the uncertainty associated with the discharge calculation. The mean discharge for the two XS was 15.14 m³/s, with an estimated uncertainty of 4.05%. However, these discharge measurements were not made in accordance with standard USGS protocols (Turnipseed and Sauer, 2010). Compass calibrations, system checks, or temperature checks were not performed, and we assumed that the edges of the channel closer to the banks than the end of each ADCP transect made no contribution to the total discharge.

2.3. Remotely sensed data

The remotely sensed data used in this study are also available from Legleiter et al. (2025a), which complements the following brief description by providing additional detail on instrumentation, flight parameters, etc.

2.3.1. Data acquired from an aircraft

Aerial imagery was acquired by the NPS on August 14, 2024, a clear day with sunny weather and no clouds. The Cessna 206 aircraft was deployed at a target flying height of 731.5 m above ground level and equipped with a Nikon D850 camera (Melville, NY). The three band (red, green, blue, or RGB) images were processed using a standard structure from motion (SfM) workflow that involves aerial triangulation using algorithms for tie point extraction and bundle adjustment. This process resulted in two primary data products: an orthomosaic of georeferenced imagery and topographic data in the form of a digital surface model (DSM, Table 2). We only used the DSM to obtain an estimate of channel slope and did not attempt to perform through-water photogrammetry due to insufficient water clarity. The original orthomosaic and DSM provided by the NPS were spatially subset to our study area along the upper Mulchatna River. A mask highlighting the wetted portion of the river channel was created by digitizing a polygon on the orthomosaic that excluded islands exposed above the water level at the time the images were acquired. The image and mask polygon covering the entire study area, which encompassed all three reaches as well as the river segments between them, were clipped to create separate subset images and channel masks for each of the three reaches.

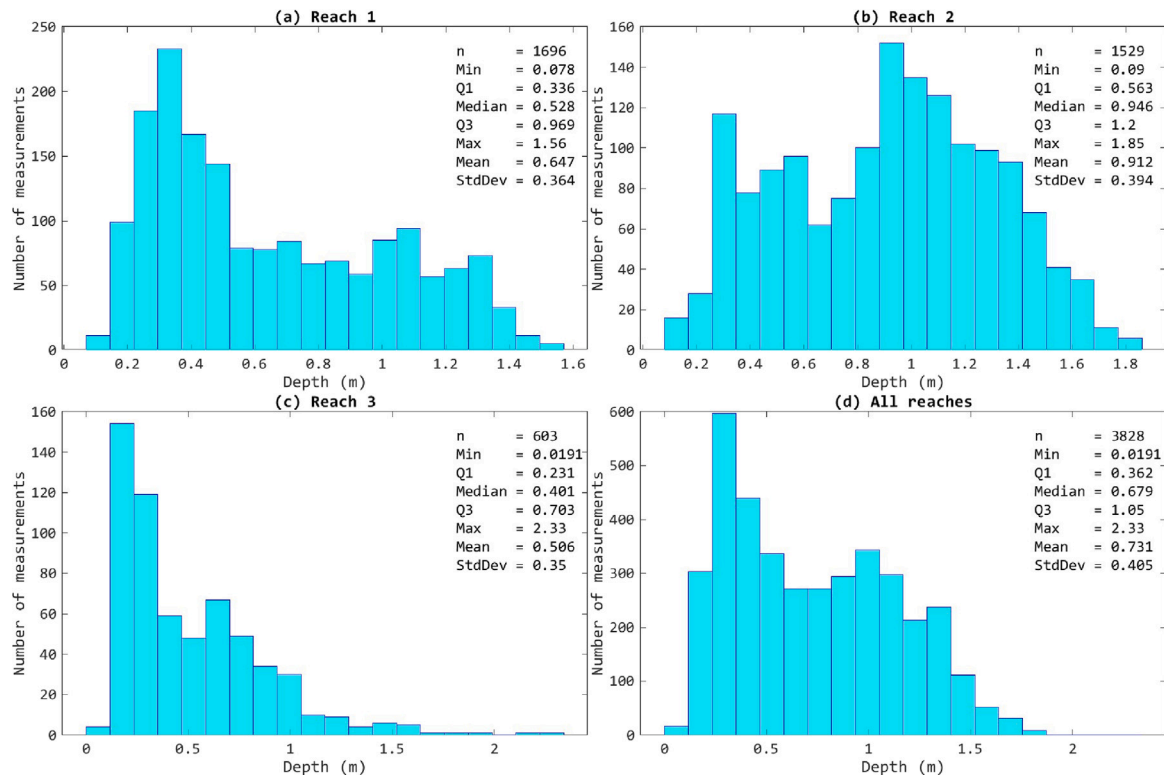


Fig. 3. Frequency distributions and summary statistics for field measurements of water depth measurements (a–c) by reach and (d) for the study area as a whole, aggregated over the three reaches. *n* refers to the number of measurements, *Q1* and *Q3* refer to the first and third quartiles, respectively, and *StdDev* refers to the standard deviation.

Table 2

Remotely sensed data products used in this study. The center wavelength of each spectral band is indicated in parentheses.

Acquisition platform	Data product	Pixel size (m)	Spectral bands (nm)
Aircraft	Orthomosaic	0.1	Blue (450) Green (550) Red (650)
Aircraft	Digital surface model	0.21	N/A
Satellite	Multispectral image	2	Coastal (427.3) Blue (477.9) Green (546.2) Yellow (607.8) Red (658.8) Red edge (723.7) Near-infrared 1 (832.5) Near-infrared 2 (908)

2.3.2. Data acquired from a satellite

We obtained a WorldView2 (WV) satellite image of the Mulchatna River acquired on July 20, 2024, and consisting of eight spectral bands (Table 2). The original tiled data product was combined into a single mosaic from which we produced one image encompassing the entire study area and a subset for each of the three reaches. Initially, the WV images were not aligned well with the aerial orthomosaic or the field measurements. To improve co-registration, tie points, such as isolated small water bodies, trees, or other distinctive features, visible in both the WV images and the aerial orthomosaic were selected and used to parameterize first-degree (rotation/scaling/translation) transformations for warping the WV images. For each of the three reaches, 20 tie points were used for this purpose. For the WV image of the full study area, the warp was based on all 60 tie points, 20 from each reach. Total RMS errors for the geo-referenced images of reaches 1, 2, and 3

and the full study area were 0.98 m, 1.46 m, and 0.76 m, and 1.86 m, respectively.

2.3.3. Surface water and ocean topography (SWOT) water level data

Due to weather and other logistical issues, field measurements and image acquisition were not coincident in time. The field survey was completed six days before the satellite image was acquired and 31 days before the orthophoto flight. To assess whether water levels in Turquoise Lake might have changed during this time period, we used SWOT pixel cloud (PIXC) data from July 7, July 28, and August 18, 2024, a period that encompasses our field measurements and the acquisition of both types of remotely sensed data used in this study. The SWOT_L2_HR_PIXC granules for these three dates were obtained from the National Aeronautics and Space Administration (NASA) Earth-Data portal (SWOT, 2023) and filtered to include only open water

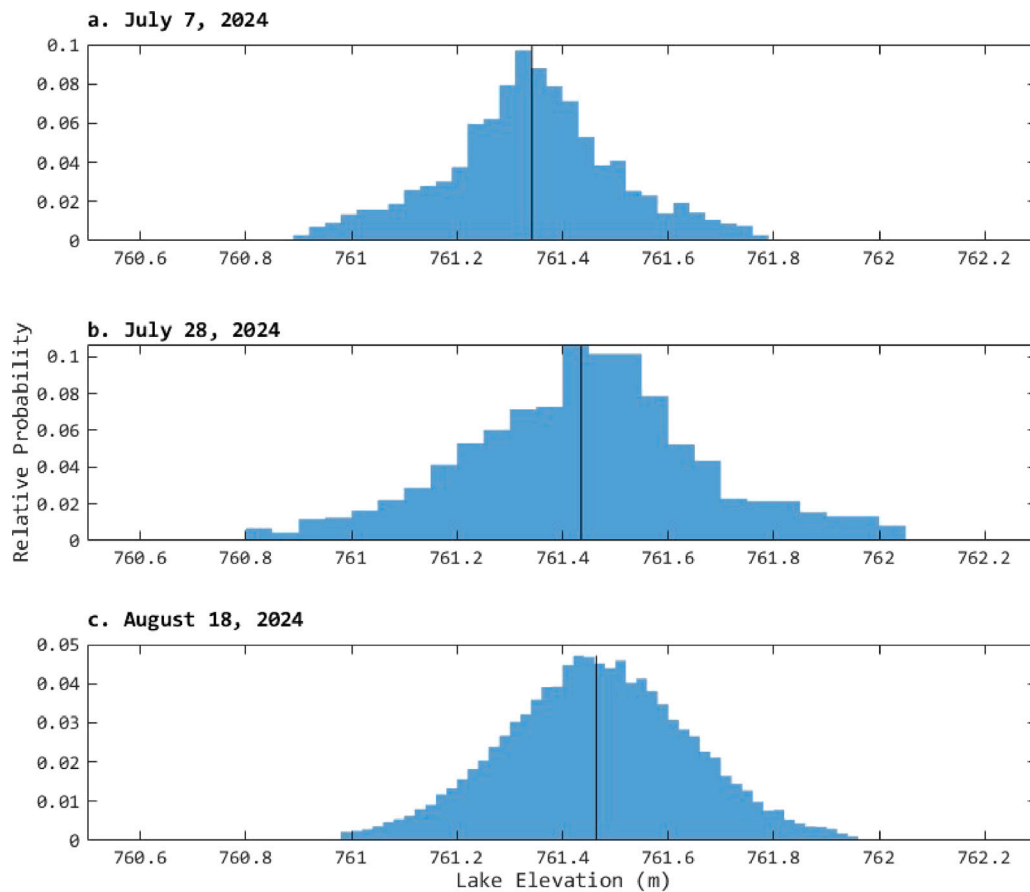


Fig. 4. Water surface elevations in Turquoise Lake for a time period that encompasses both field data collection and the acquisition of the satellite images and aerial orthophotos used in this study. Histograms show the distribution of the Surface Water and Ocean Topography (SWOT) pixel cloud (PIXC) returns within the lake filtered to remove outliers for visualization. Vertical black lines are the median lake elevation of all PIXC returns on each date. Elevations are referenced to the EGM2008 vertical datum.

returns with quality flags less than 65,536 for the geolocation, sig0 (a normalized radar cross-section that serves as an estimate of the radar backscatter), interferogram, and classification variables and cross track values between 10,000 and 60,000. We then clipped the filtered returns to the extent of Turquoise Lake and calculated the median of the water surface elevations within the lake for each date (Fig. 4). SWOT data indicate that the median lake level varied by 0.12 m between passes on July 7, July 28, and August 18. However, the median lake elevation on each date remained within the interquartile range of the elevations recorded on the other two dates, suggesting that the differences between the three dates were relatively minor compared to the variability in lake elevation measurements from each of the individual SWOT passes. As described in Appendix A, we calculated how the change in median lake level translated into variations in river depth. Based on this analysis, we assumed that because Turquoise Lake is the headwater of the Mulchatna River and thus controls streamflow in the river downstream of the lake, water depths along the river did not change substantially between the time the field data were collected and when the images were acquired.

2.4. Depth retrieval algorithms

This section provides a general description of the four depth retrieval algorithms we evaluated. Further information on how these techniques were implemented within the context of this study is provided in Section 2.5.

2.4.1. Optimal band ratio analysis (OBRA)

This well-established method requires field measurements for calibration but has a sound physical basis in radiative transfer theory (Mobley, 1994; Legleiter et al., 2004). The rate at which sunlight is attenuated within the water column varies as a function of wavelength and in clear-flowing, relatively shallow streams, depth d is the primary control on the radiance signal (Legleiter et al., 2009). OBRA seeks to isolate this effect by relating d to an image-derived quantity X , defined as a log-transformed band ratio:

$$X = \log \left[\frac{R(\lambda_1)}{R(\lambda_2)} \right], \quad (1)$$

where R refers to radiance, reflectance, or raw digital numbers and λ_1 and λ_2 are the wavelengths of the numerator and denominator bands, respectively. Taking as input paired observations of d and $R(\lambda)$, OBRA calculates X for all possible band combinations and performs a regression of X against d for each version of X . The optimal band ratio is that which yields the highest coefficient of determination R^2 and the corresponding regression equation can be applied throughout an image to produce a bathymetric map. Although the initial version of OBRA assumed a linear relation between X and d , Legleiter and Harrison (2019) showed that an exponential formulation can provide a better fit and avoid negative depth estimates in shallow areas. In this study, we performed OBRA based on the more recent exponential model:

$$d = b_0 e^{b_1 X}, \quad (2)$$

where b_0 and b_1 are coefficients that can be estimated by linearizing Eq. (2) prior to regression (Equation 7 of Legleiter and Harrison, 2019).

2.4.2. Hyperspectral imaging of river bathymetry using an ensemble of regression trees (HIRBERT)

This workflow is similar to OBRA in that paired observations of d and $R(\lambda)$ are required as input but uses a more advanced machine learning (ML) algorithm to train depth retrieval models. Building upon previous research by Kwon et al. (2024), Legleiter et al. (2025b) showed that using an ensemble of regression trees, also known as a random forest (Breiman, 2001), can yield more accurate depth estimates over a broader range of depths by capturing complex, nonlinear relationships between depth and reflectance across multiple spectral bands, not just the single pair of wavelengths used in OBRA. Further discussion of the advantages of this approach and additional details on its implementation are available in Legleiter et al. (2025b) and only a brief summary is provided here.

A regression tree ensemble yields robust estimates by integrating the output from many individual trees, each of which is based on a distinct random subset of the data and potential predictor variables, which in this case are the individual bands of a multi- or hyperspectral image. For a given tree, the predictors are partitioned into nodes, analogous to the branches of a tree, based on values of the output variable, depth. The training process involves iteratively adjusting the splitting rules for all nodes to maximize the reduction in the mean squared error (MSE) between predicted and observed depths. The final predicted depth is the mean of the predictions from all the trees in the ensemble, which helps mitigate against overfitting. Another factor to consider when developing an ensemble of regression trees is parsimony; models based on a carefully selected subset of bands can lead to depth estimates that are just as accurate as those based on a full hyperspectral data set (Kwon et al., 2024). To guide variable selection, we used a metric called out-of-bag permuted predictor importance estimates by permutation that measures the influence of each individual band in a model on depth predictions. Ideally, an ML-based depth retrieval model would also provide some insight regarding the underlying relationships between depth and reflectance that inform the model's estimates. To facilitate this type of interpretation, we calculated Shapley values, which quantify how much a given band contributes to the predicted depth when the band is included in the model, as opposed to when the model does not include the band, averaged over all possible combinations of bands. Shapley values also provide information on how predicted depths vary as reflectance changes for each of the bands in the model (Legleiter et al., 2025b).

2.4.3. Flow resistance equation-based imaging of river depths (FREEBIRD)

Unlike OBRA and HIRBERT, this approach does not require field measurements to calibrate some image-derived quantity X to water depth. Instead, this connection is established using principles of open channel flow. The primary inputs to FREEBIRD are an image consisting of X values, channel aspect (width/depth) ratio, water surface slope, and discharge. The core of this method is a flow resistance equation adapted from Bjerklie et al. (2005):

$$\bar{v} = k\bar{d}^{0.67}S^{0.33}, \quad (3)$$

where $\bar{v}$ is the cross-sectional average velocity, k is a conductance coefficient, $\bar{d}$ is the mean flow depth, and S is the water surface slope. We used the single, constant value of $k = 7.14$ from Bjerklie et al. (2005).

The original formulation of FREEBIRD assumed a linear relation between X and d given by

$$d = b_0 + b_1 X, \quad (4)$$

where b_0 and b_1 are the intercept and slope coefficients, respectively. The critical connection between the image and channel hydraulics is established by substituting Eq. (4) into Eq. (3). Legleiter (2015) showed how the product of the image-derived depth estimate and the velocity calculated from Eq. (3) based on that estimate can be used to calculate a unit discharge q for each pixel along a XS extracted from the image.

The resulting q values can then be integrated across the channel to yield the total discharge $\hat{Q}$:

$$\hat{Q} = \Delta w k S^{0.33} \sum_{j=1}^M (b_0 + b_1 X_j)^{1.67}, \quad (5)$$

where Δw is the incremental width represented by each image pixel and j indexes the M pixels along the XS.

Plausible, XS-specific estimates of b_0 and b_1 can be established by linking the shallowest depth d_0 along the i th of N XS to either the minimum or maximum value of X , denoted by $X_{0,i}$, along the corresponding image transect. Two cases can be distinguished, depending on whether X is directly ($b_1 > 0$) or inversely ($b_1 < 0$) related to d :

$$X_{0,i} = \begin{cases} \min(X_j), & \text{if } b_1 > 0 \\ \max(X_j), & \text{if } b_1 < 0 \end{cases} \quad (6)$$

Legleiter (2015, Equations 4-8) provides a derivation of b_0 and b_1 for the case where X and d are linearly related.

In this study, we also considered an exponential relation between X and d , which leads to the following expressions for the exponent b_1 and coefficient b_0 appearing in Eq. (2):

$$b_{1,i} = \frac{\log(\bar{d}_i) - \log(d_0)}{(\bar{X}_i - X_{0,i})} \quad (7)$$

$$b_{0,i} = \exp[\log(d_0) - b_{1,i} X_{0,i}] \quad (8)$$

The resulting values of $b_{0,i}$ and $b_{1,i}$ can be substituted into either the linear (4) or exponential (2) X vs. d relation to estimate depths that in turn can be used to obtain a discharge estimate $\hat{Q}_i$ for each XS via Eq. (5).

The next step in FREEBIRD is to calibrate the image-derived depth estimates so as to match a known discharge by modifying the coefficients of the X vs. d relation. This process involves using the initial values of $b_{1,i}$ and $b_{0,i}$ given by Eqs. (7) and (8), or the analogous expressions from Legleiter (2015) for the linear case, to estimate $\hat{Q}_i$ for each XS along a reach. The N estimated discharges are compared to the known discharge and used to calculate the root mean square error (RMSE) as

$$\epsilon_Q = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{Q}_i - Q)^2} \quad (9)$$

This metric of disagreement between the known Q and that based on image-derived depth estimates serves as the objective function for an optimization algorithm that minimizes ϵ_Q by iteratively adjusting $b_{0,i}$ and $b_{1,i}$.

For the linear case, Eq. (4) can yield $d < 0$ if mixed pixels along the banks include both water and terrestrial features, or if $b_{0,i}$ and $b_{1,i}$ are mis-specified. FREEBIRD avoids negative depth estimates by assigning a velocity of 0 to any pixels for which the current values of $b_{0,i}$ and $b_{1,i}$ result in $d < 0$, which prevents those areas from contributing to $\hat{Q}_i$. To prevent the algorithm from returning coefficients that would result in a large number of pixels with $d < 0$, the user must also specify the fraction of depth estimates allowed to be negative. For each iteration of the optimization routine, the objective function (9) is set to a very large value if the current coefficients yield an unacceptably high proportion of negative depth estimates, which forces the optimization toward coefficients that result in $d > 0$.

2.4.4. Image-to-depth quantile transformation (IDQT)

In terms of field data requirements, the final depth retrieval algorithm we evaluated is intermediate between OBRA and HIRBERT, which rely upon paired observations of depth and reflectance, and FREEBIRD, which does not require depth measurements for calibration. Instead, the primary input to IDQT is the frequency distribution of depths within the reach, which can then be linked to the frequency distribution of pixel values. In essence, IDQT uses the distributions

of d and X to perform an aspatial calibration, with the image itself providing information on the spatial pattern of depths within the reach. This approach provides greater flexibility in the timing of field data collection relative to image acquisition, which is an important consideration for sites that are difficult to access and also eliminates the need for precise co-registration between field surveys and image data. Another advantage of IDQT is that a functional form of the relationship between X and d need not be specified. Instead, complex, non-linear associations between the two can be captured as long as X and d are monotonically related to one another. This conversion from image pixel values to depth is accomplished by matching quantiles of the distributions of X and d via a quantile transformation procedure that takes as input the cumulative distribution function (CDF) value for an image pixel and returns the corresponding quantile of the distribution of d . The IDQT framework is summarized below, but a more thorough discussion of the approach is available from Legleiter (2016).

The first step in the IDQT process is to scale both X and d by their mean values within the reach, denoted as $\langle X \rangle$ and $\langle d \rangle$, to obtain non-dimensional quantities $X/\langle X \rangle$ and $d/\langle d \rangle$ that represent proportions of the reach-averaged pixel value and depth, respectively. This standardization accounts for the different magnitudes and units of X and d , but the image and field data must encompass the same spatial extent (i.e., both distributions must be sampled from the same reach “population”). Legleiter (2016) described several ways of obtaining CDFs for the two scaled variables, denoted by $F_X(X/\langle X \rangle)$ and $F_d(d/\langle d \rangle)$, but found that the IDQT algorithm was insensitive to these implementation details. In this study, we considered two specific cases. First, both distributions can be specified empirically by using all the pixel values within the water-only image and all the depth measurements within the reach. The second approach involves using a hypothetical distribution for $d/\langle d \rangle$. The gamma distribution has been shown to provide a reasonable approximation to the observed frequency distribution of depths within stream reaches (e.g., Segura and Pitlick, 2015, and other references cited in Legleiter, 2016). The shape and scale parameters, denoted by $a > 0$ and $b \geq 0$, respectively, of a gamma distribution's can be specified solely in terms of the reach-averaged mean depth and the variance of the scaled depths:

$$a = \frac{1}{\text{Var}[d/\langle d \rangle]} \quad (10)$$

$$b = \text{Var}[d/\langle d \rangle] \quad (11)$$

This approach is advantageous because a full, plausible distribution of depths can be generated based on limited field data.

Having defined the CDF $F_X(X/\langle X \rangle)$ empirically and the CDF $F_d(d/\langle d \rangle)$ either empirically or as a gamma distribution parameterized via Eqs. (10) and (11), the IDQT algorithm proceeds by determining the CDF value for each pixel in the image as

$$p(s) = F_X(X(s)/\langle X \rangle), \quad (12)$$

where s denotes the location of the pixel and $X(s)$ represents the value of that pixel. Eq. (12) thus yields a probability value, $p(s)$, for each pixel. Next, $p(s)$ is inserted into the inverse CDF of $d/\langle d \rangle$ to obtain the scaled depth at that location:

$$d(s)/\langle d \rangle = F_d^{-1}(p(s)) \quad (13)$$

Whereas the CDF of $X/\langle X \rangle$ takes a pixel value as input and returns a probability, the inverse CDF of $d/\langle d \rangle$ takes that probability as input and returns a scaled depth. The IDQT algorithm thus links the CDF of the image to that of depth. After applying the quantile transformation, the resulting values of $d/\langle d \rangle$ are multiplied by $\langle d \rangle$ to create the final depth map.

2.5. Data processing

The purpose of this section is to explain how we applied each of the four depth retrieval algorithms described in Section 2.4 to the Mulchatna River. The workflows outlined below and summarized graphically in Fig. 5 were executed for both the aerial orthomosaic and the WV satellite image for each of the three reaches and for the study area as a whole, for a total of eight analyses (two types of data times four spatial extents). The code we developed to complete this processing is available from Legleiter (2025). Also note that although including figures to illustrate various intermediate and final outputs for all combinations of sensor, site, and algorithm would be impractical, such content is included in a set of slides provided as supporting information (SI). The slides are organized into sections by reach, image type, and depth retrieval method, and we refer to several specific examples in the remainder of this paper.

2.5.1. Pre-processing

We prepared the images and field measurements for depth retrieval by performing a common set of operations across all four algorithms that served as a precursor to method-specific aspects of the analysis. For each of the eight combinations of sensor and site, these procedures included: (1) importing the image, field data, and digitized channel mask; (2) clipping the depth measurements to the channel polygon; (3) applying the mask to the image; and (4) smoothing the masked water-only image using a Wiener spatial filter (Lim, 1990) with a 3- X 3-pixel moving window. After initial pre-processing was complete, the workflow diverged for the various depth retrieval algorithms. However, because some of the intermediate outputs generated when implementing one method could also be used by one or more of the other methods, we performed the processing in the order presented below to take advantage of these efficiencies. This approach allowed us to avoid unnecessary repetition and reduce computational run time and storage requirements.

2.5.2. OBRA processing

The next step in the processing chain focused on the OBRA algorithm and involved: (1) identifying all of the depth measurements located within each pixel of the masked, filtered image and averaging them to obtain pixel-scale mean depths; (2) extracting image spectra from each of these pixels, examining the spectra, and discarding obvious outliers; (3) splitting the paired observations of depth and reflectance into separate subsets, with 80% of the available data used to calibrate a relation between X and d and the remaining 20% set aside for accuracy assessment; (4) performing OBRA using an exponential form of the X vs. d relation and producing a calibration plot and matrix; and (5) applying the X vs. d relation for the optimal band ratio throughout the image to produce a bathymetric map; and (6) assessing the accuracy of the resulting image-derived depth estimates using the test subset of the pixel-scale mean depths. To make these procedures less abstract, two steps from this phase of the processing are illustrated below using data from reach 1 as a representative example. Fig. 6 shows spectra for each decile of the distribution of the training subset of pixel-scale mean depths for both the RGB orthomosaic and the eight-band multispectral image. Similarly, the OBRA calibration plots in Fig. 7a and c show the fitted relation between X and d and the matrices shown in Fig. 7b and d provide a means of visualizing how the strength of the X vs. d relation varied for different combinations of numerator and denominator bands.

Also note that the test subset of the pixel-scale mean depths generated during this phase of the analysis was also used to evaluate the performance of the FREEBIRD and IDQT methods, which provided a common basis for comparison across the different techniques.

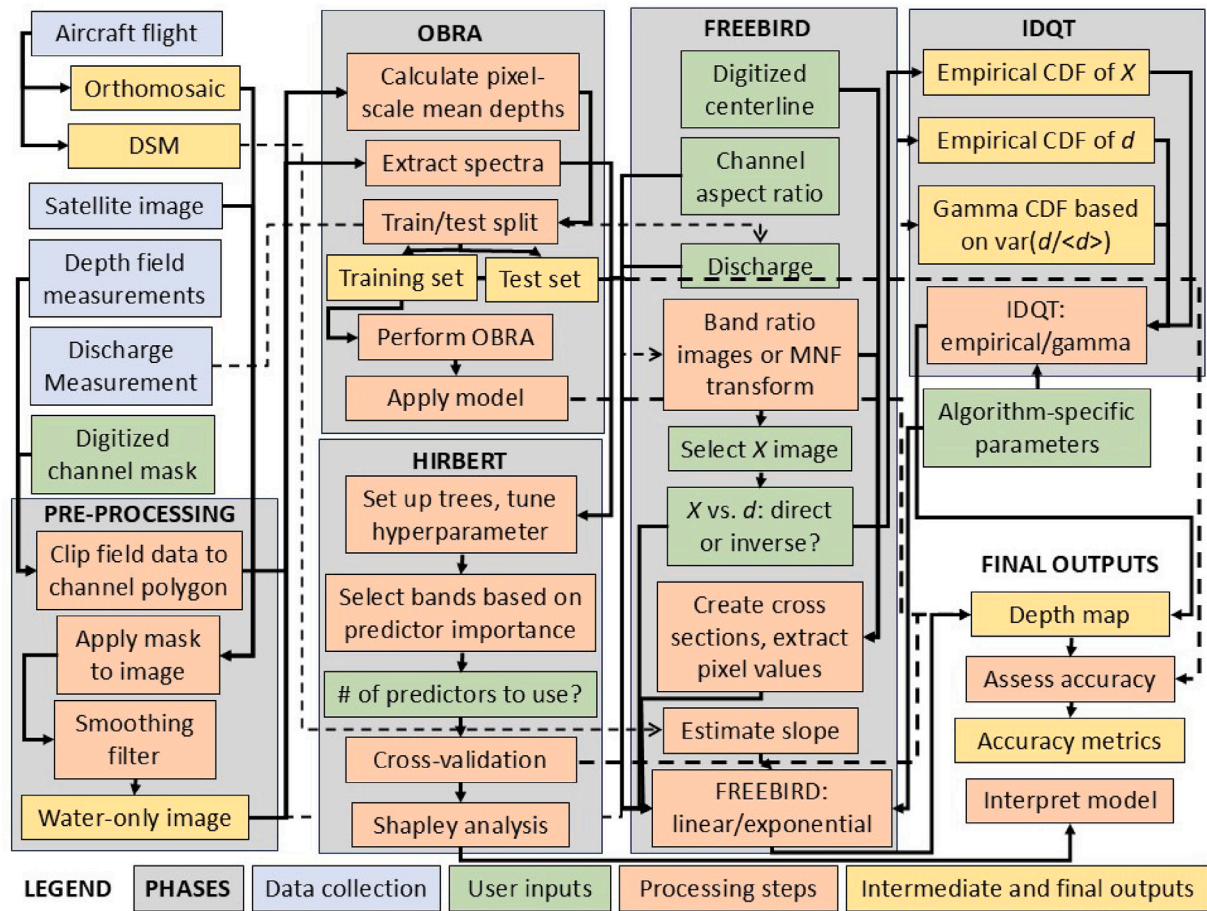


Fig. 5. Flowchart summarizing the various depth retrieval algorithms. To clarify some of the connections that are obscured by other graphical elements, note that: (1) the same pre-processed water-only image serves as input to all of the algorithms; and (2) the training set established during optimal band ratio analysis (OBRA) is also used in image-to-depth quantile transformation (IDQT) and the test subset is used to assess the accuracy of depth estimates derived via OBRA, flow resistance equation-based imaging of river depths (FREEBIRD), and IDQT. DSM = digital surface model, MNF = minimum noise fraction, CDF = cumulative distribution function.

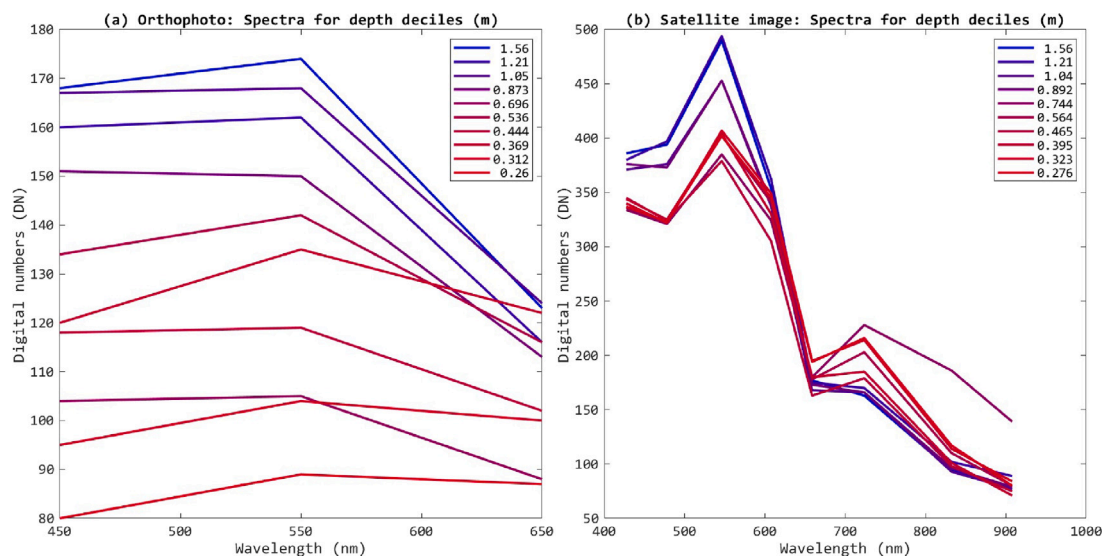


Fig. 6. Spectra extracted from the (a) orthophoto and (b) satellite image for each decile of the distribution of the training subset of pixel-scale mean depths for reach 1.

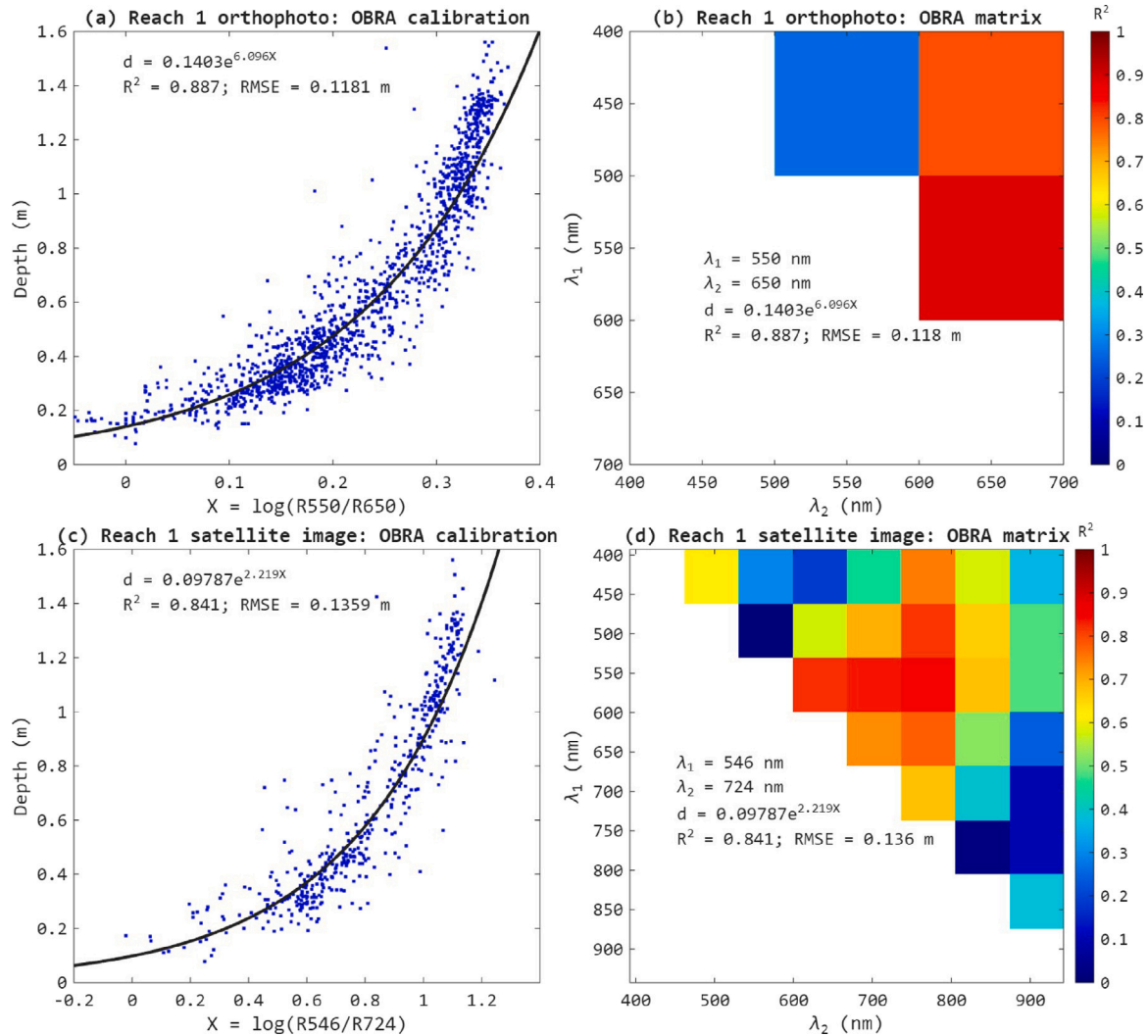


Fig. 7. Example optimal band ratio analysis (OBRA) outputs derived from the orthophoto and satellite image of reach 1 include (a and c) plots illustrating calibration of the X vs. d relation and (b and d) matrices that show how the strength of the X vs. d relation varies for different combinations of numerator and denominator bands.

2.5.3. HIRBERT processing

Because the HIRBERT workflow creates its own partition of the input data as part of a cross-validation procedure, the separate training and testing subsets created during the OBRA processing were recombined and the same overall set of pixel-scale mean depths and corresponding image spectra were used for HIRBERT. We then implemented this algorithm as follows. (1) Use all of the available predictors (bands) to set up an ensemble of regression trees and tune the hyperparameter for the maximum number of decision splits (i.e., branch nodes) to allow for each tree. (2) Select bands to retain in a reduced model by calculating a metric of predictor importance (see slide 36 of the SI for an example), sorting the predictors in order of importance, fitting models with different numbers of bands, and then using these models to estimate depths for a test subset of the data. (3) Decide how many predictors to retain based on a plot of the accuracy of the image-derived depth estimates against the number of bands (SI slide 37). (4) Assess accuracy by performing a five-fold cross-validation of the regression tree ensemble based on the selected bands. Of the models for the five folds, that which yields the highest observed versus predicted R^2 is used as the final model. (5) Apply this model to the pre-processed image to produce a bathymetric map (SI slide 38). (6) Produce a Shapley summary plot to quantify the importance of each predictor and facilitate interpretation of the depth retrieval model (SI slide 41).

2.5.4. FREEBIRD processing

As described in Section 2.4.3, FREEBIRD does not require field measurements of water depth. The only inputs to this algorithm were the pre-processed image, a digitized channel centerline, the DSM, and the parameters listed in Table 3. The first step in the FREEBIRD workflow is to identify an appropriate image to use as X . For the RGB orthophotos, we generated a log-transformed band ratio image for each of the three possible combinations of numerator and denominator bands and visually inspected them to identify which one appeared to be most strongly correlated with water depth based on our field experience at each site (Fig. 8a–c). For the eight-band satellite images, examining all 28 possible band ratios was impractical so we distilled the multispectral data into a reduced number of bands using the minimum noise fraction (MNF) transform introduced by Green et al. (1988) and described in Legleiter (2016). In this study, we found that the first MNF band distinguished the water portion of the masked image from the background and the second MNF band was correlated with water depth; we thus used the latter band as X . We then visually inspected the X image to assess whether X was directly related to d , in which case brighter pixels were associated with deeper water, or inversely related to d , with darker areas of the image corresponding to greater depths. For all cases considered in this study, the relationship between X and d was direct, such that a plot of X vs. d would have a positive slope.

Table 3
Depth retrieval algorithm parameters.

Algorithm	Parameter name	Value
FREEBIRD	Cross section spacing	10 m
	Channel aspect ratio	41.04
	Minimum depth	0.01 m
	Known discharge	15.14 m ³ /s
	Negative depth fraction	0.05
IDQT	Distribution censoring	0.01

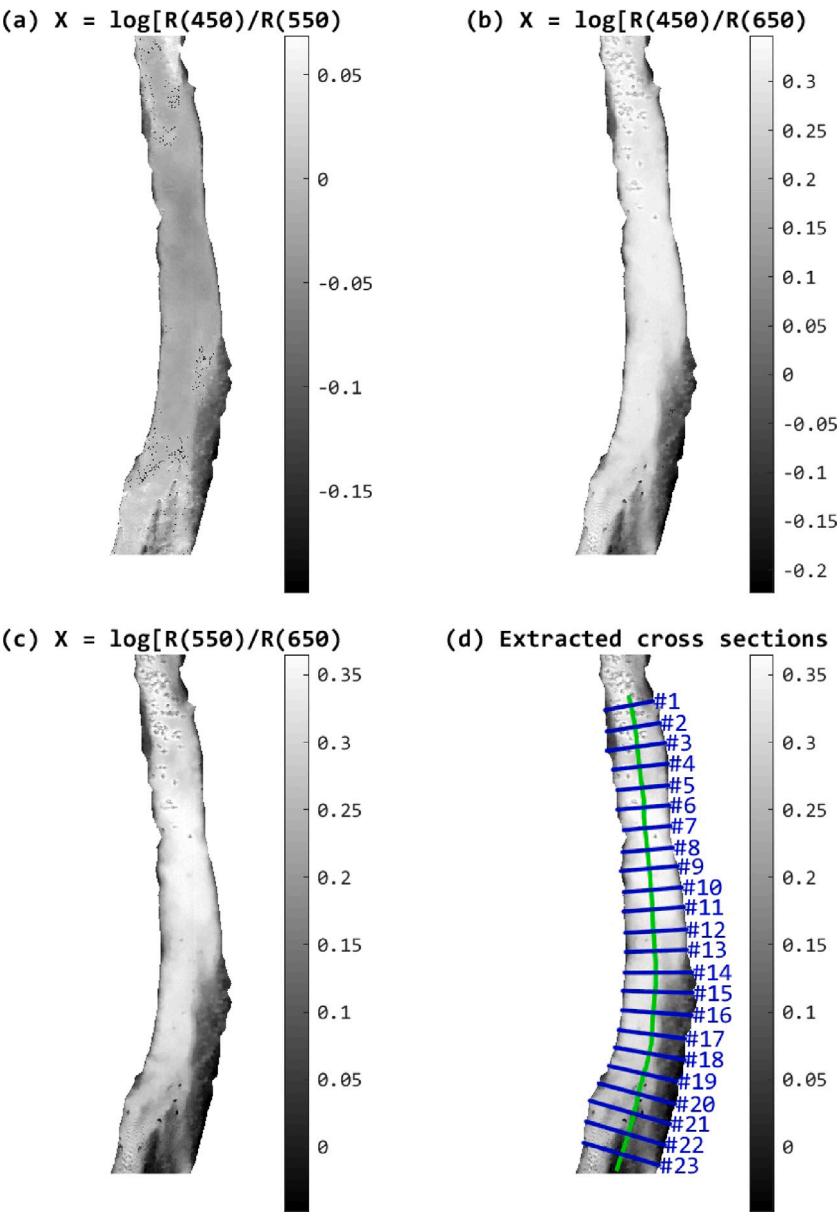


Fig. 8. (a–c) Log-transformed band ratio images generated from the orthophoto for reach 2 for each combination of numerator and denominator bands. (d) Cross sections extracted from the selected *X* image along a digitized channel centerline.

The next stage of the FREEBIRD process involved creating a series of regularly spaced XS along a digitized channel centerline, as shown in Fig. 8d. We then extracted pixel values from the *X* image for each of the resulting XS to serve as input to the core FREEBIRD algorithm. In addition, we used the same set of XS, along with the DSM, to estimate the channel slope by: (1) extracting water surface elevation (WSE) values from the DSM where each XS intersects the digitized channel polygon on each side of the river (SI slide 137); (2) averaging the WSE values from the two banks for a given XS; (3) performing a

linear regression of the laterally averaged WSE against the streamwise distance along the centerline for each XS (SI slide 138); the slope coefficient of this regression was used as an estimate of the average water surface slope for the reach. Given the cross sections of *X* values, the estimated slope, and the other parameters listed in Table 3, we then applied both the original linear and new exponential formulations of the FREEBIRD algorithm. For each of these two variants, we produced a bathymetric map and assessed its accuracy by comparing image-derived depth estimates to

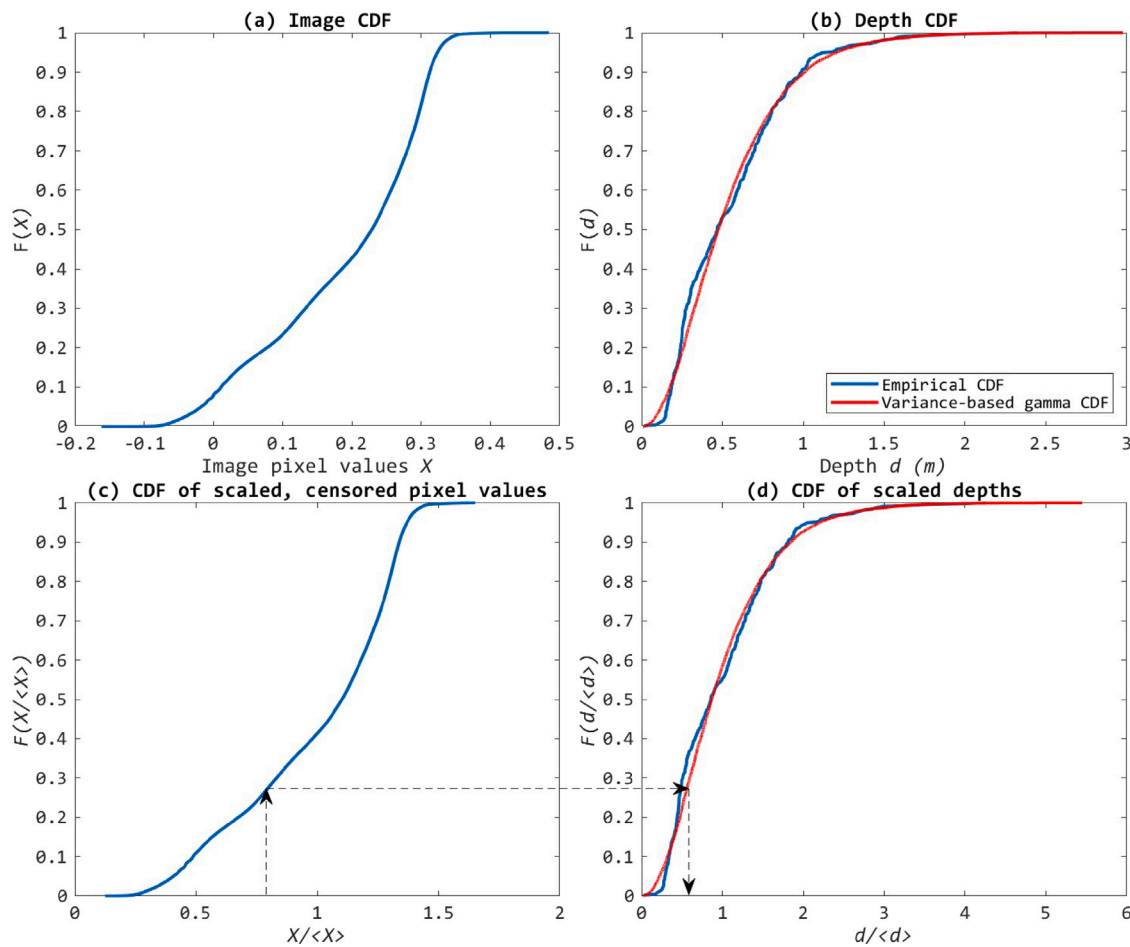


Fig. 9. Example of the image-to-depth quantile transformation (IDQT) process from the orthophoto for reach 3, with the distribution of depths defined either empirically or as a gamma distribution parameterized based on the mean depth and the variance of the scaled depths $d/\langle d \rangle$.

the test subset of pixel-scale mean depths established during the OBRA phase (Section 2.5.2, step 3).

2.5.5. IDQT processing

To evaluate IDQT, the final depth retrieval algorithm, we used the same X image as for FREEBIRD, either a log-transformed band ratio image produced from the RGB orthophoto or the second MNF band derived from the WV image. We then used the training subset of the pixel-scale mean depths established during the OBRA processing as input to two different versions of IDQT. In both cases, we used the empirical distribution of X pixel values, censored by removing the top and bottom 1% of values to avoid outliers associated with mixed pixels along the banks (Table 3), but the distribution of depths was established differently. For the first variant, we used the full, empirical distribution of depths as input to the quantile transformation. In the second, we instead established a hypothetical gamma distribution based on the mean depth and the variance of the scaled depths $d/\langle d \rangle$, as described in Section 2.4.4. An example from the orthophoto for reach 3 that illustrates the two distributions of depths, as well as the quantile transformation process, is shown in Fig. 9. For each of the IDQT variants, we produced a depth map, and assessed its accuracy by comparing image-derived depth estimates to the same test subset of pixel-scale mean depths used to quantify the performance of the OBRA and FREEBIRD methods.

2.5.6. Progressively truncated input depths (PTID) analysis

In addition to the preceding analyses, which were conducted for all four sites (three reaches and the full study area) and involved six

different algorithms, we also performed two complementary types of analysis that focused on reach 3 and a subset of the depth retrieval methods. First, we evaluated whether the Mulchatna River might exceed the maximum depth detectable by the camera deployed from the aircraft and/or the multispectral sensor onboard the satellite using progressively truncated input depths (PTID) (Legleiter et al., 2018). This approach involves iteratively repeating the depth retrieval process for different subsets of the data created by excluding observations that exceed a sequence of specified depth cutoffs. The PTID technique thus provides a means of quantifying bathymetric mapping performance across a range of depths and Legleiter et al. (2018) hypothesized that an inflection point on a plot of observed vs. predicted (OP) R^2 against cutoff depth, beyond which accuracy begins to decline, could be used to infer a site-specific maximum detectable depth. Conversely, if no such inflection point is apparent, presumably the imaging system is capable of detecting the full range of depths present in the reach of interest. For this study, we performed PTID analyses for the OBRA and HIRBERT algorithms for both the orthophoto and WV image.

2.5.7. Cross-validation OBRA

The second type of analysis we conducted exclusively for reach 3 was intended to assess the potential of a recently introduced OBRA variant to enable depth retrieval when only a very limited number of field observations, all from a single cross section, are available for calibration (Legleiter et al., 2024). This cross-validation approach involves subdividing the available data into separate training and testing subsets, called folds. For a given number of points, n , and a specified number of folds, k , the data are randomly partitioned into k

disjoint subsamples, each of which has approximately the same number of observations. For this study, we set $k = 5$ and used the isolated discharge measurement cross section located at the upstream (east) end of reach 3 (Fig. 1(e)) as input, consisting of $n = 49$ pixel-scale mean depths and corresponding spectra for the orthophoto but only $n = 15$ for the coarser-resolution WV image. For each of the 5 folds, we performed OBRA based on the 80% ($1 - 1/k$) of the total sample size (39 points for the orthophoto or 12 for the WV image) in the training partition. We then applied the resulting X vs. d relation to the test subset to predict depths from image pixel values. The accuracy of these estimates was quantified by comparing them to the known test depths and calculating the root mean squared error (RMSE). This process was repeated for each of the k folds and the RMSE values tabulated for each fold. The fold that yielded the smallest RMSE was considered optimal and the corresponding OBRA model was used as the final X vs. d relation. This equation was then applied throughout the image to produce a bathymetric map that could be compared to that produced by the standard OBRA workflow based on all of the field data available for reach 3.

2.6. Evaluating depth retrieval performance

We quantified the accuracy of depth estimates derived using all combinations of image type and retrieval algorithm in terms of four primary metrics. First, we calculated depth retrieval errors ϵ as the difference between the pixel-scale mean depths derived from the field data, d_f , and the depths inferred from the image, d_i :

$$\epsilon_j = d_{f,j} - d_{i,j}, \quad (14)$$

where j indexes the n_{test} observations in the subset of the data reserved for testing. Positive values of ϵ thus imply that depths were under-predicted (i.e., shallower than measured in the field), whereas an $\epsilon < 0$ indicates over-prediction (i.e., the image-derived estimate was deeper than that observed in the field). The first summary metric quantified the bias associated with image-derived depth estimates and was calculated by dividing the mean of ϵ by that of d_f :

$$NBias = \frac{\frac{1}{n_{test}} \sum_{j=1}^{n_{test}} \epsilon_j}{\frac{1}{n_{test}} \sum_{j=1}^{n_{test}} d_{f,j}} \quad (15)$$

Normalizing by the mean depth in this manner allowed for comparison across sites. To obtain a similar measure of the typical magnitude of depth retrieval errors, irrespective of their sign, we also calculated the normalized mean absolute error as

$$NMAE = \frac{\frac{1}{n_{test}} \sum_{j=1}^{n_{test}} |\epsilon_j|}{\frac{1}{n_{test}} \sum_{j=1}^{n_{test}} d_{f,j}} \quad (16)$$

We obtained an index of precision by applying the same normalization to the RMSE calculated for each site:

$$NRMSE = \frac{\sqrt{\frac{1}{n_{test}} \sum_{j=1}^{n_{test}} \epsilon_j^2}}{\frac{1}{n_{test}} \sum_{j=1}^{n_{test}} d_{f,j}} \quad (17)$$

We also used the test subset of the data to regress image-derived depth estimates (d_i) against pixel-scale mean depths based on field measurements (d_f). The coefficient of determination R^2 for these observed versus predicted (OP) regressions provided a concise summary of the overall level of agreement between d_i and d_f . Ideally, the OP regression line would coincide with the one-to-one line of perfect agreement, resulting in an R^2 of 1 and a regression equation with a slope of 1 and intercept of 0 indicating unbiased estimates (Pineiro et al., 2008). Finally, we also used histograms to summarize distributions of ϵ and produced maps of depth retrieval errors to visualize spatial variations in depth retrieval performance.

3. Results

To assess the feasibility of mapping water depths along the proglacial Mulchatna River and, more generally, support the effective use of remote sensing, we performed an extensive analysis involving multiple image data types and several methods for inferring bathymetry. More specifically, we evaluated RGB orthophotos acquired from an aircraft and multispectral images from a satellite. Six different depth retrieval algorithms were considered: OBRA, HIRBERT, linear and exponential formulations of FREEBIRD, and IDQT based on empirical and variance-based gamma distributions of depth. We applied each of these techniques to both types of data for each of four spatial extents: three distinct reaches and the full study area. This study design resulted in a total of 48 separate analyses: 4 sites $\times$ 2 data types $\times$ 6 algorithms. We also conducted additional PTID and cross-validation OBRA analyses focused on reach 3, as described in Sections 2.5.6 and 2.5.7. Presenting detailed results for more than 50 different scenarios would be unwieldy, so we instead provide a representative example of the output for one particular combination of site, sensor, and algorithm in the following subsection, synthesize our accuracy assessments in Section 3.2, and summarize the PTID and cross-validation OBRA results in Sections 3.3 and 3.4. In addition, the supporting information includes figures analogous to those for the example provided within the main text for all the other cases. To facilitate viewing and comparison, this content is compiled in a set of slides organized by reach, image type, and method.

3.1. Example depth retrieval outputs

As a representative example of the 48 depth retrieval analyses we conducted, this section focuses on the site farthest upstream along the Mulchatna River (reach 1), the image data type with the smallest number of bands (an RGB orthophoto), and the most well-established depth retrieval algorithm (OBRA). The spectra extracted from the image at the locations of the field measurements shown in Fig. 6a indicated that depth was directly proportional to reflectance. The more reddish-colored lines in this figure represent shallower water (i.e., the first few deciles of the distribution of pixel-scale mean depths) and tend to have smaller digital numbers (i.e., lower reflectances). Conversely, the bluer lines representing deeper areas (i.e., larger deciles of the depth distribution) generally correspond to brighter areas in the image, particularly for the shorter-wavelength blue and green bands. Deeper water was thus associated with higher reflectance and this relationship was quantified via the calibrated exponential relation between X and d shown in Fig. 7a. The ratio of the green (550 nm) to the red (650 nm) band yielded the strongest correlation, with an R^2 of 0.887, but the OBRA matrix in Fig. 7b indicates that using the blue (450 nm) band would have produced an R^2 nearly as high. Of the three possible band combinations for this RGB image, the only one that did not yield an X strongly related to d was $R(450)/R(550)$.

Producing a log-transformed band ratio image using the optimal pair of wavelengths, $R(550)/R(650)$, and then applying the calibrated X vs. d relation from Fig. 7a to each pixel resulted in the depth map shown in Fig. 10a. The spatial pattern of bathymetry is consistent with our field experience at this site: deeper flow along the north bank through the upper portion of the reach and in a thalweg closer to the center of the channel farther downstream to the west. Much of this reach is very shallow, with depths less than 0.2 m over the large bar in the middle of Fig. 10a. Comparison of image-derived depth estimates with the test subset of the pixel-scale mean depths calculated from the field data indicated that depths inferred via OBRA were unbiased on average, with mean and median errors on the order of 0.01 m and a standard deviation of 0.105 m (Fig. 10b). The distribution of errors was slightly skewed toward positive values, with a maximum of nearly 0.35 m, suggesting that depths were more likely to be under-estimated relative to the field measurements. The corresponding map of errors

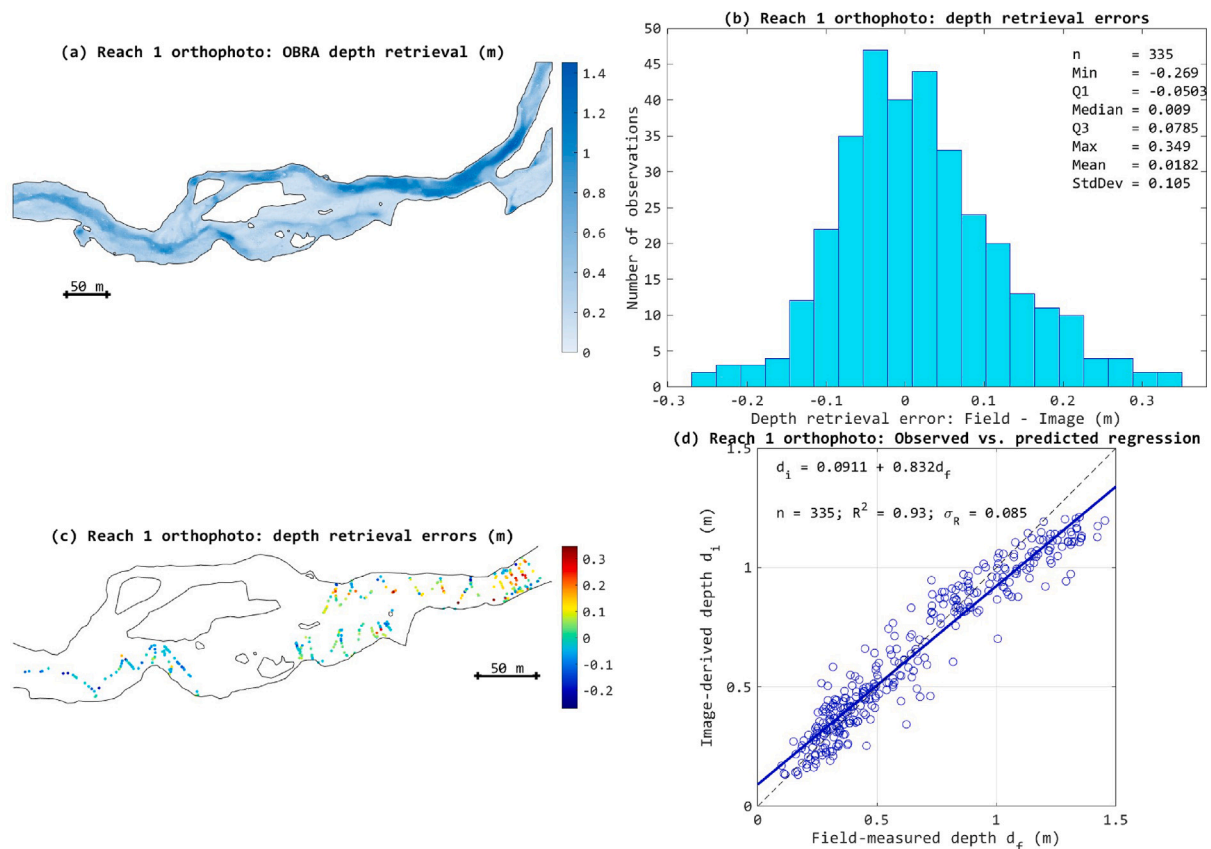


Fig. 10. (a) Depth map produced from the orthophoto of reach 1 via optimal band ratio analysis (OBRA). (b) Distribution of depth retrieval errors and associated summary statistics. (c) Corresponding map of depth retrieval errors. (d) Regression of observed (measured in the field) versus predicted (estimated from the image) depths, with the dashed line representing perfect agreement.

indicated that the largest under-predictions of depth occurred toward the upper (east) end of the reach along the south bank just downstream of the confluence of a side channel with the main thread, where the greatest depths were observed in the field. Conversely, the negative errors in the lower (west) portion of the reach imply that depths tended to be over-predicted in this area. Overall, however, agreement between observed and predicted depths was very strong, with an R^2 of 0.93 for the test subset (Fig. 10d). The OP regression equation had an intercept greater than 1 and a slope less than 1, implying that depths tended to be over-predicted in shallow areas and under-predicted in the deepest portions of the river.

3.2. Accuracy assessment summary

We performed similarly detailed analyses for 47 other combinations of site, sensor, and algorithm and plots analogous to those shown in Fig. 10 are provided as supporting information. To distill these results into a more concise summary and enable comparisons across these three factors, we focused on the four performance metrics defined in Section 2.6: normalized values of the mean bias, mean absolute error, and RMSE, and the OP R^2 . We produced bar graphs for each metric, with each group of bars representing a site (the full study area is coded as reach 4) and different combinations of sensor and algorithm represented by distinct bar colors. To keep this graphical presentation manageable, we generated three separate figures 11–13 of this kind, each comparing two related algorithms. However, note that for a given metric these three figures have different vertical scales.

Fig. 11 compares the two methods that require paired observations of depth and reflectance, OBRA and HIRBERT. These results suggest that both methods performed well overall, with OP R^2 ranging from

0.72 to 0.93 across the four sites and two sensors, but the new ML-based HIRBERT technique tended to provide slightly more reliable depth estimates than the more established OBRA framework. Depth estimates tended to have a small positive bias, implying under-predictions of depth, but in all cases the bias was less than 8% of the reach-averaged mean depth. Similarly, mean absolute errors for both methods were less than 20% across all sites and sensors. Depth estimates were also reasonably precise, with normalized RMSE values typically on the order of 0.15–0.2 and in all cases less than 0.3. Differences in depth retrieval performance between orthophotos acquired from an aircraft and multispectral images from the WV satellite were small and inconsistent, implying that either sensor could be used to map this river environment.

Accuracy metrics for the linear and exponential variants of the FREEBIRD workflow, which does not require depth measurements for calibration, are depicted in Fig. 12. The performance of these two methods was more variable across sites and sensors, with OP R^2 ranging from 0.61 for the satellite image of reach 3 to 0.93 for the orthophoto of reach 1. Differences among reaches were also more pronounced, with consistently positive normalized biases up to 0.43 in reach 1, whereas this metric was negative but smaller in absolute value (up to 0.13) in reach 2 for both versions of FREEBIRD and both sensors; the normalized bias was a mix of positive and negative values for reach 3 and the full study area. The exponential formulation of FREEBIRD introduced in this study tended to yield larger normalized mean absolute errors and RMSE values than the original version based on a linear relation between X and d , particularly for the WV image of reach 3. Although comparison of the two sensors led to mixed results, applying the FREEBIRD technique to the orthophotos tended to yield more reliable depth estimates than when WV images were used as input. Using an exponential X vs. d relation to process WV images led to the poorest

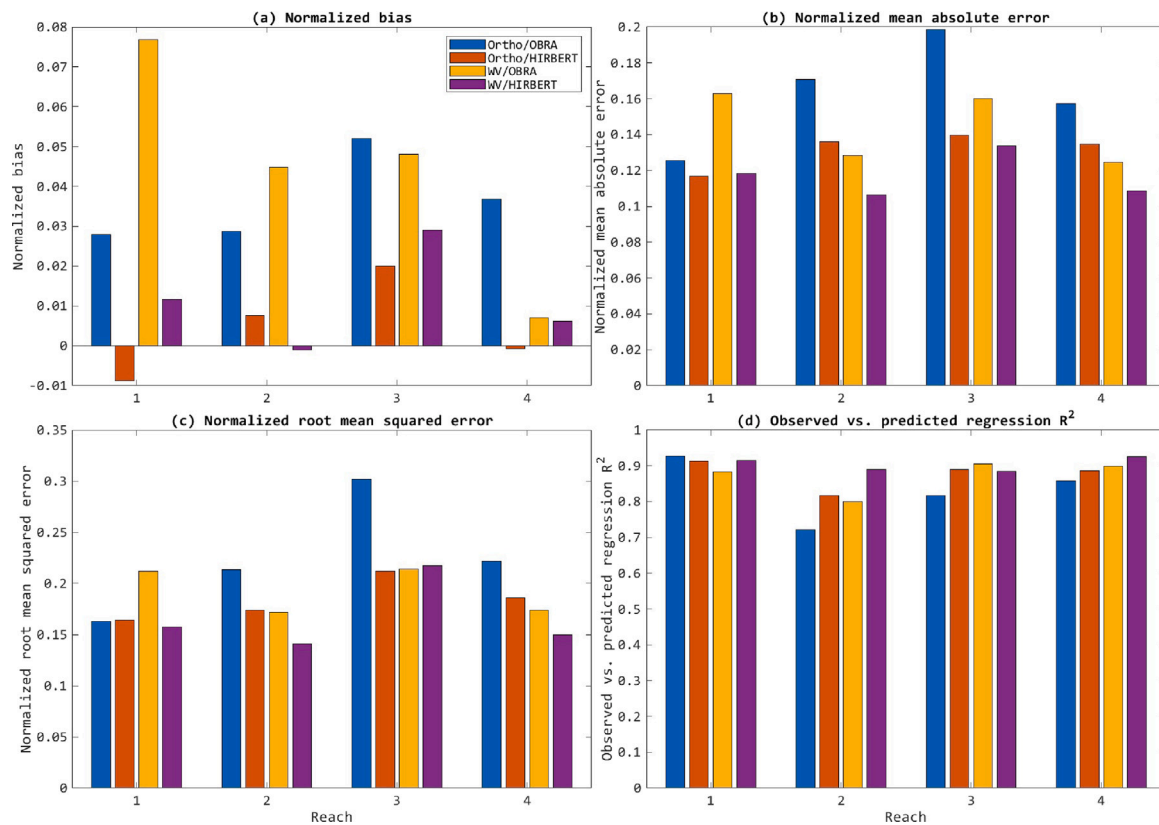


Fig. 11. Accuracy assessment summary for the optimal band ratio analysis (OBRA) and hyperspectral imaging of river bathymetry using an ensemble of regression trees (HIRBERT) depth retrieval methods for the three reaches and the full study area, which is coded as reach 4, and for the orthophoto (ortho) acquired from an aircraft and the multispectral image from the WorldView2 (WV) satellite in terms of four metrics of performance: (a) normalized bias, (b) normalized mean absolute error, (c) normalized root mean square error, and (d) observed versus predicted (OP) regression R^2 .

performance across all four metrics we evaluated. These results imply that the new exponential version did not provide an improvement over the linear formulation and that FREEBIRD might be better-suited to aerial than satellite images.

Accuracy metrics for the two versions of IDQT we evaluated, based on empirical or variance-based gamma distributions of depth, are shown in Fig. 13. In terms of OP R^2 , this approach performed well across the 16 combinations of site, sensor, and algorithm, with values ranging from 0.70 for the orthophoto from reach 2 to 0.94 for the WV image of reach 3. However, the high level of agreement between observed and predicted depths masked a substantial bias in the bathymetry mapped via this method. Normalized biases were consistently negative in reaches 1 and 2 and for the full study area, typically on the order of 40% of the reach-averaged mean depth, implying that depths tended to be over-predicted for these sites. The bias was much smaller, less than 0.06 in absolute value, for reach 3. Similarly, normalized MAE and RMSE values were around 40% for the first two reaches and the full study area but about half as large for reach 3. Differences in performance between the two ways of specifying the distribution of depths and between the two kinds of images were small and inconsistent, implying that IDQT was insensitive to implementation details and input data type.

As a final synthesis, we enabled a more direct comparison across the six depth retrieval algorithms by averaging each performance metric over site and sensor for a given method. The results of this analysis are summarized in Fig. 14, with error bars representing the standard deviation of the metric across the four sites and two sensors. Overall, this analysis indicated that the two methods that require paired observations of depth and reflectance for training, OBRA and HIRBERT, provided the most reliable bathymetric information. Although all six algorithms we evaluated yielded mean OP R^2 values greater than

0.78, and up to 0.89 for HIRBERT, in some cases a strong correlation between observed and predicted depths obscured biases in the image-derived estimates. For example, whereas HIRBERT provided depth estimates that were essentially unbiased on average, both versions of IDQT led to large negative biases on the order of 30% of the mean depth, implying substantial over-predictions of depth by this method. For FREEBIRD, the bias tended to be negative for the linear version of the algorithm and positive, implying under-predictions of depth, for the new exponential formulation. The two methods that require little (IDQT) or no (FREEBIRD) field data also led to larger values of the normalized MAE and RMSE than the two techniques (OBRA and HIRBERT) that rely upon more extensive training data. For the latter two algorithms, the small error bars associated with each metric imply that performance was not highly variable across the four sites and two sensors we considered in this study.

3.3. Depth retrieval performance across a range of depths

To assess whether depths along the Mulchatna River exceeded the maximum that could be detected by the camera deployed from an aircraft or the multispectral sensor mounted on the WV satellite, we conducted a progressively truncated input depths (PTID) analysis focused on reach 3. We applied a series of cutoff depths from 0.3 m to 2.3 m in 0.1-m increments and then performed OBRA and HIRBERT at each iteration to quantify how depth retrieval accuracy varied across this range of depths. For the initial shallow depth cutoffs, OBRA yielded low R^2 values due to the small number of observations available for calibration when most of the field measurements were discarded (Fig. 15). As the depth cutoff increased and more, deeper observations were made available, performance improved considerably, reaching a plateau around 0.85 for both calibration and OP R^2 for the orthophoto.

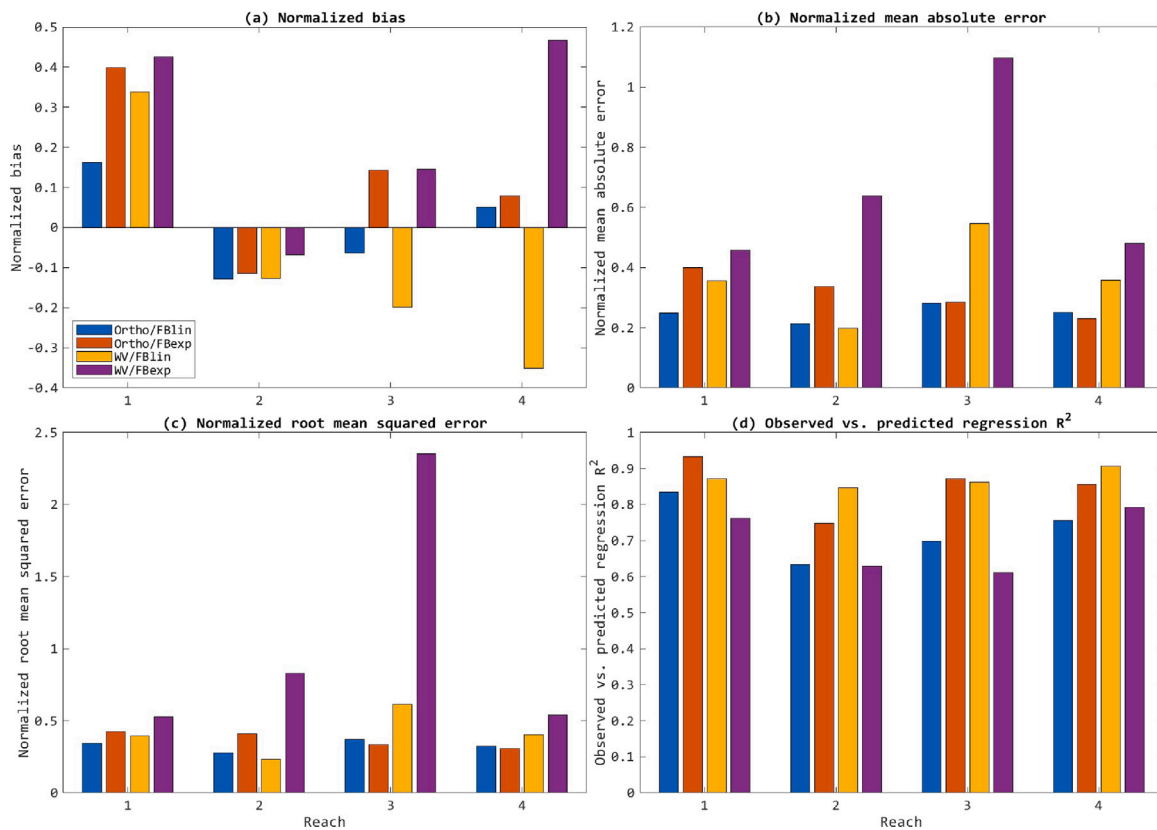


Fig. 12. Accuracy assessment summary for the flow resistance equation-based imaging of river depths (FREEBIRD) depth retrieval method based on linear and exponential relations between X and d , denoted as FBlin and FBexp, for the three reaches and the full study area, which is coded as reach 4, and for the orthophoto (ortho) acquired from an aircraft and the multispectral image from the WorldView2 (WV) satellite in terms of four metrics of performance: (a) normalized bias, (b) normalized mean absolute error, (c) normalized root mean square error, and (d) observed versus predicted (OP) regression R^2 .

For the satellite image, the OBRA calibration R^2 leveled off around 0.72 but the OP R^2 continued to rise, up to 0.87, as the depth cutoff increased and more measurements were used to calibrate the exponential relationship between X and d . The greater sensitivity to cutoff depth for the satellite image than for the orthophoto was a consequence of the coarser spatial resolution of the WV data, which led to a smaller number of pixel-scale mean depths for a given cutoff. The ML-based HIRBERT technique was less affected by sample size and provided more consistent, stronger performance across the full range of cutoff depths, with a cross-validation R^2 of approximately 0.9 for both the orthophoto and satellite image. Depth retrieval performance did not deteriorate for either sensor or algorithm as deeper field observations were included in the training data set. These results imply that the maximum detectable depth was not exceeded in this study.

3.4. Depth retrieval based on a single cross section

To assess whether spectrally based depth retrieval could be based on a minimal amount of field data, we used depth measurements from a single XS in reach 3 as input to the cross-validation version of OBRA described in Section 2.5.7. The results of this analysis suggest that such a limited data set could yield depth estimates that were just as accurate as those obtained when all the field measurements from reach 3 were used as input to the standard OBRA workflow. For the orthophoto, the OP R^2 was 0.83 for cross-validation OBRA and 0.82 for standard OBRA. The depth maps produced via the two versions of OBRA were virtually indistinguishable (compare slides 195 and 367 in the SI) and led to similar distributions of errors (slides 196 and 368). However, the effect of using data from a single XS as input to OBRA was more pronounced

for the satellite image, which had coarser pixels and thus fewer pixel-scale mean depths, only 15 for the XS we used. Whereas standard OBRA based on the entire field data set from reach 3 led to an OP R^2 of 0.90, applying the cross-validation approach to measurements from a single XS resulted in much weaker agreement between observed and predicted depths ($R^2 = 0.53$). Relative to the depth map produced using standard OBRA, the depth map generated via the cross-validation workflow was more homogeneous and failed to capture depth variations within the reach (compare slides 215 and 374 in the SI); the mean and standard deviation of depth retrieval errors were also much larger for the latter method (slides 216 and 375). To provide a more graphical summary, we extracted a XS located near the apex of the bend to the south in the middle of reach 3 (approximately 185 m downstream from the discharge measurement XS we used as input to the cross-validation OBRA algorithm) from the image-derived depth maps produced via the standard and cross-validation versions of OBRA for both the orthophoto and the WV image (Fig. 16). For the higher-resolution aerial image, the XS based on the full data set is nearly identical to that based on depth measurements from a single transect, with the cross-validation approach leading to slightly shallower depth estimates over the bar on the left side of the channel. Although this channel morphology was also captured in the XS extracted from the depth map produced from the WV image using standard OBRA (orange line in Fig. 16), depths were overestimated for the bar when the cross-validation method was applied using a much smaller number of field measurements. These results imply that although field data from a single transect could support depth retrieval in some cases, this approach might be less reliable when applied to images with lower spatial resolution.

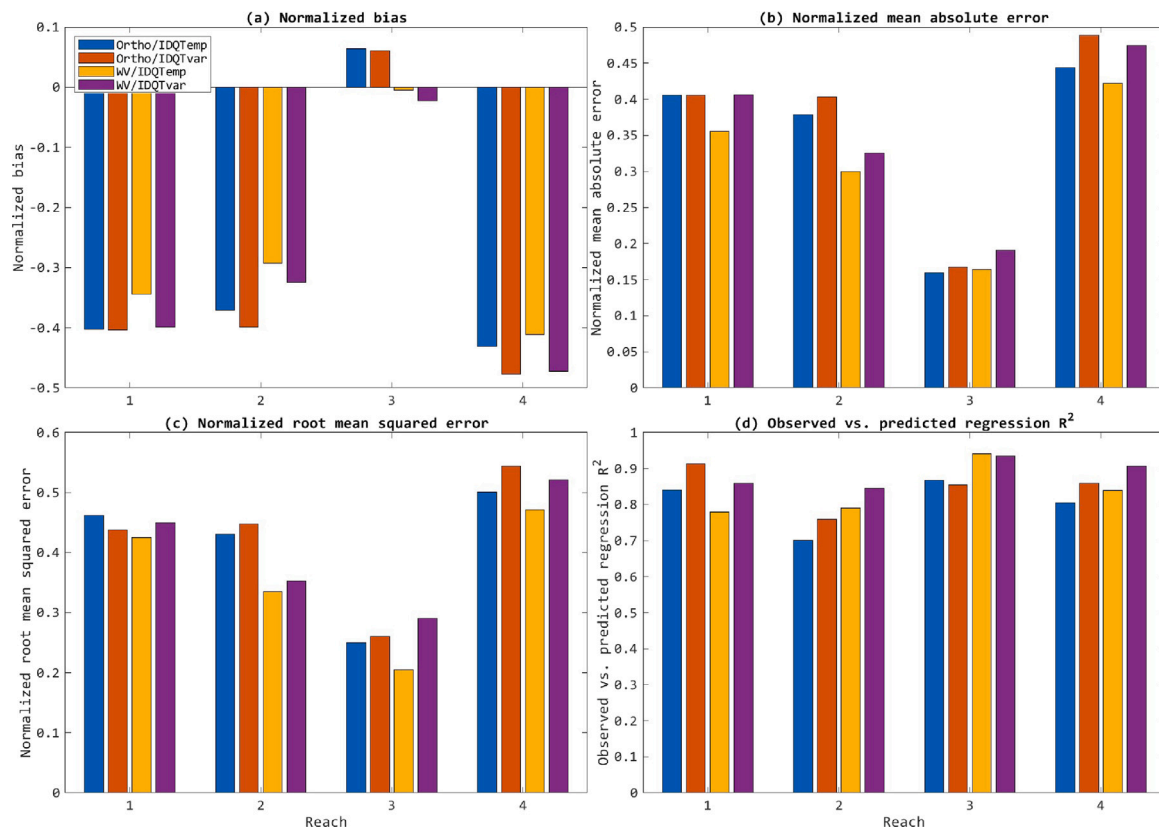


Fig. 13. Accuracy assessment summary for the image-to-depth quantile transformation (IDQT) depth retrieval method based on empirical or variance-based gamma distributions of depth, denoted as IDQTTemp and IDQTvar, for the three reaches and the full study area, which is coded as reach 4, and for the orthophoto (ortho) acquired from an aircraft and the multispectral image from the WorldView2 (WV) satellite in terms of four metrics of performance: (a) normalized bias, (b) normalized mean absolute error, (c) normalized root mean square error, and (d) observed versus predicted (OP) regression R^2 .

4. Discussion

4.1. Limitations

In this study, we performed a thorough evaluation of the feasibility of mapping the bathymetry of a proglacial river via remote sensing and compared the performance of several depth retrieval algorithms with varying input data requirements. Although our results yield some insight regarding the potential applicability of these methods to this type of fluvial environment, the investigation was also limited in several important respects. Most importantly, a time gap of up to 31 days between when we made field measurements of water depth and when the image data were acquired implies that variations in water level during this period could introduce some degree of uncertainty into our analysis. However, SWOT measurements of the elevation of Turquoise Lake (Fig. 4), along with the analysis summarized in Appendix A, suggest that depth varied by only 0.11 m during this time period. In addition, our analysis was focused on a single, possibly quite unique river; further testing would be needed to assess the degree to which our findings might be generalized to other proglacial rivers. Similarly, we collected field data from three short reaches along the Mulchatna River and produced bathymetric maps spanning a total of 3.84 km, but the depth retrieval models we developed were inherently site-specific and only pertain to the areas from which the data used to develop them were obtained. The various X vs. d relations we derived should not be extended to other rivers, or even other segments of the Mulchatna River farther downstream. Input of relatively clear water from tributaries and/or deposition of the fine-grained glacial sediment that give the river its distinctive turquoise color would further restrict the range of applicability of calibrated equations like that shown in Fig. 7a.

Additional limitations to acknowledge relate to the algorithms themselves. The original publications that introduced the various methods we employed provide more extensive discussions, but key constraints to consider include the following. For the two spectrally based techniques, OBRA and HIRBERT, the need for paired observations of depth and reflectance to train a depth retrieval model implies that a field effort must be organized and coordinated with image acquisition, which poses a number of logistical challenges and can undermine some of the primary advantages of remote sensing. In addition, ML-based methods like HIRBERT tend to be data-intensive, implying that a large number of depth measurements might be required to train and test a model. For FREEBIRD, the framework rests upon a number of hydraulic assumptions, most notably that a simple flow resistance Eq. (3) adequately relates velocity, depth, and slope, not only at the scale of a XS but also to each individual width increment across the channel. In addition, the conductance coefficient k is assumed constant. The method is also intended for single-thread channels. Although FREEBIRD does not rely upon field measurements of water depth, other types of data are required: a DSM or some other means of estimating slope and a known discharge. We used an ADCP to measure discharge in the field, but these data were not collected in accordance with standard USGS protocols (Turnipseed and Sauer, 2010). Due to limited time and adverse weather conditions, we were not able to confirm that the ADCP satisfied quality assurance checks for the temperature sensor, compass, and overall system status. In addition, we assumed that the edges of the channel beyond the last valid measurement for each ADCP transect made a negligible contribution to the total discharge. Any resulting uncertainties in the discharge data used as input to FREEBIRD might have impacted the algorithm's performance. More generally, however, discharge information could be retrieved from a gaging station where available or estimated via some kind of hydrologic model (e.g., Yang

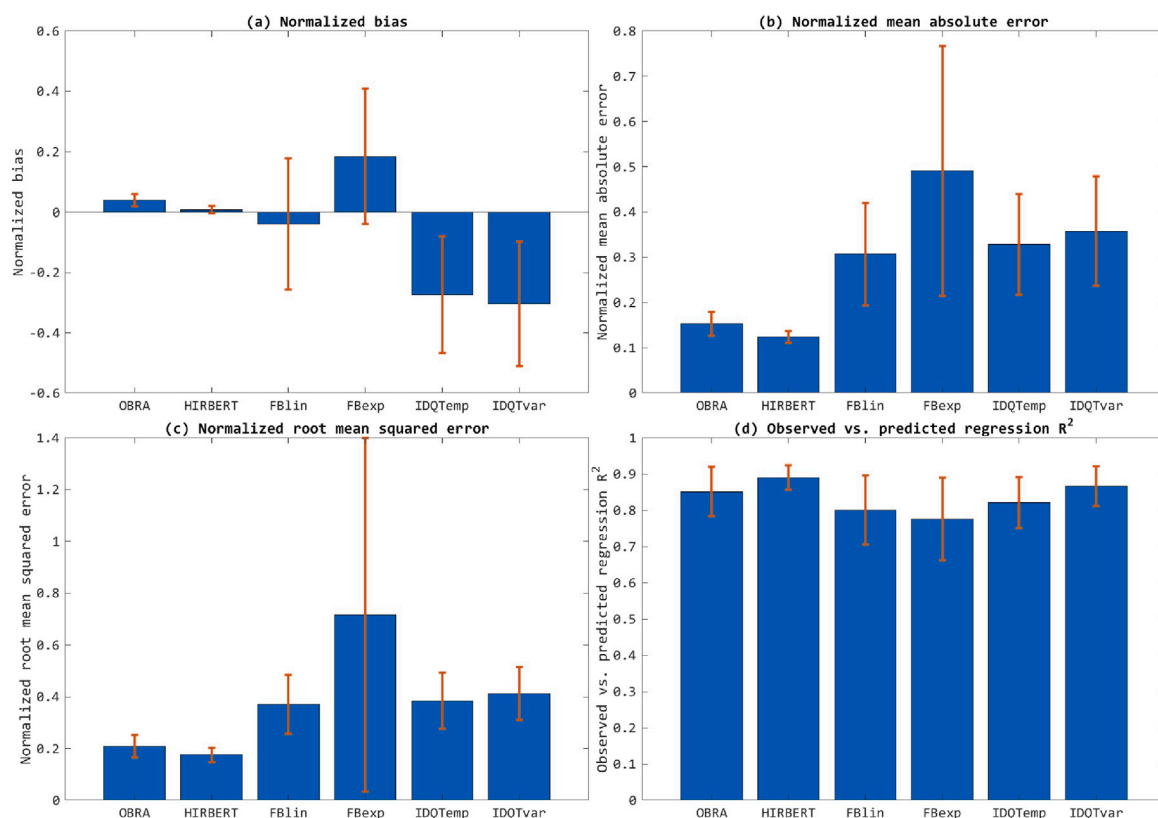


Fig. 14. Accuracy assessment summary for six different depth retrieval methods: optimal band ratio analysis (OBRA); hyperspectral imaging of river bathymetry using an ensemble of regression trees (HIRBERT); flow resistance equation-based imaging of river depths (FREEBIRD) based on linear and exponential relations between X and d , denoted as FBlin and FBexp; and image-to-depth quantile transformation (IDQT) based on empirical or variance-based gamma distributions of depth, denoted as IDQTemp and IDQTvar. The following four metrics of performance were averaged over the four sites and two sensors, with the error bars indicating the associated standard deviation: (a) normalized bias, (b) normalized mean absolute error, (c) normalized root mean square error, and (d) observed versus predicted (OP) regression R^2 .

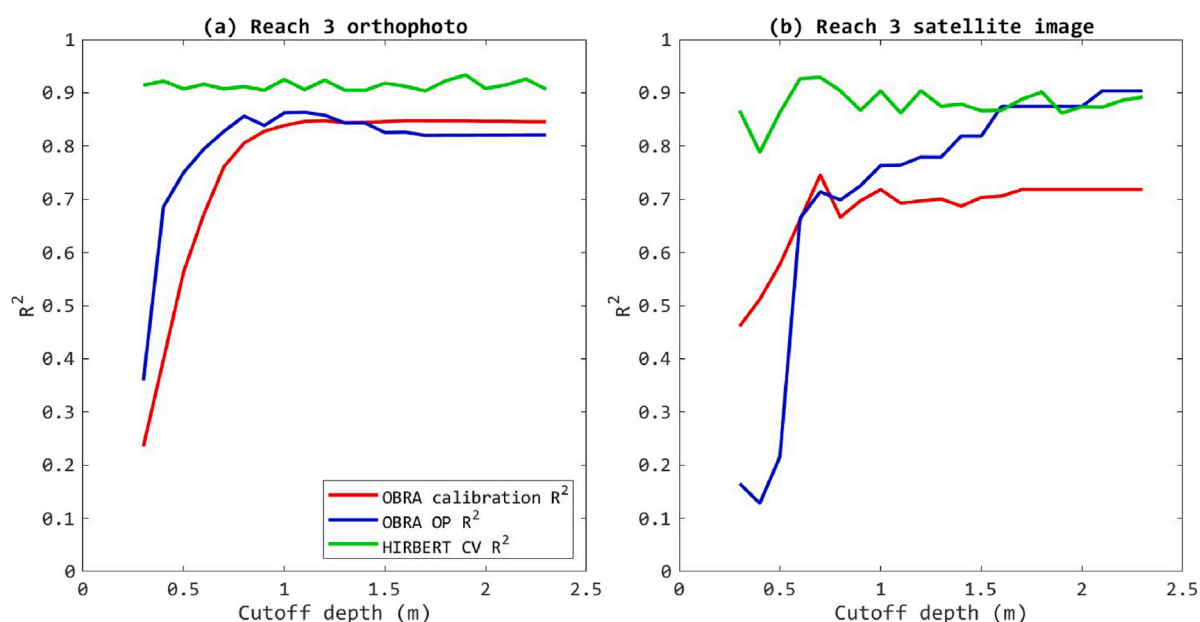


Fig. 15. Progressively truncated input depths (PTID) analysis for the (a) orthophoto and (b) satellite image of reach 3 for the optimal band ratio analysis (OBRA) and hyperspectral imaging of river bathymetry using an ensemble of regression trees (HIRBERT) methods. For OBRA, both the calibration and observed versus predicted (OP) regression R^2 values are plotted. For HIRBERT, the metric of performance is the cross-validation (CV) R^2 .

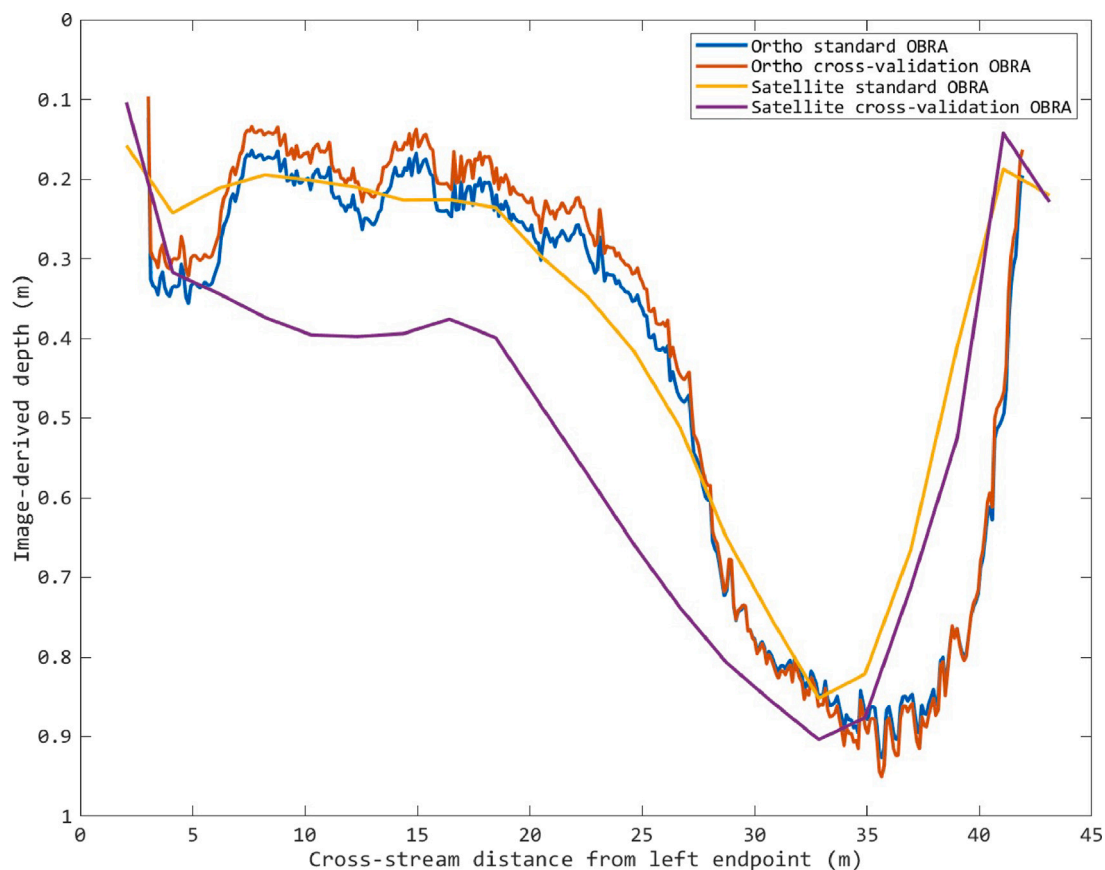


Fig. 16. Comparison of depth estimates from standard optimal band ratio analysis (OBRA) and a cross-validation-based version of the algorithm for a cross section in reach 3 derived from both the orthophoto and satellite image.

et al., 2025). For both FREEBIRD and IDQT, the selection of an appropriate X image is a subjective decision based on visual inspection of various candidates, such as log-transformed band ratios or the output from an MNF transform. The IDQT approach involves two additional, key assumptions: that X is monotonically related to d and, for the variance-based approach, that d follows a gamma distribution.

4.2. Bathymetric mapping of proglacial rivers

Although previous efforts to infer water depth from passive optical image data have primarily focused on relatively shallow, clear-flowing streams, the results of this study imply that this approach can also yield reliable bathymetric information in a much more turbid, turquoise-colored, relatively shallow proglacial river. This finding is novel because most prior applications of OBRA have relied upon an inverse relationship between depth and reflectance due to strong absorption by pure water at red and near-infrared (NIR) wavelengths, which reduces the amount of radiant energy reflected from the bottom (Legleiter et al., 2004, 2009). Under these conditions, deeper water appears darker. On the Mulchatna River, however, the spectra shown in Fig. 6 indicate that the opposite is true. In this case, greater depths are associated with higher reflectances, particularly in the green and blue bands, due to greater volume reflectance from the water column and a reduced contribution from the streambed, which is darker than the water above. In deeper areas of the channel, the larger number of sediment particles suspended within a thicker water column scatter more light back up toward the sensor. Conversely, along the shallower margins of the river fewer such scattering elements are present and more of the incident solar radiation reaches the relatively dark substrate, leading to a lower total reflectance. Although this phenomenon was captured in the radiative transfer simulations performed by Legleiter et al. (2009), this

study provides empirical evidence that in some sediment-laden rivers, a direct proportionality between depth and reflectance can also support bathymetric mapping. Moreover, our results suggest that either RGB orthophotos or eight-band multispectral images could be used for this purpose.

Although HIRBERT was originally intended for hyperspectral data sets consisting of many narrow wavelength bands, this study demonstrated that applying this ML-based workflow to more readily available types of image data, such as multispectral satellite images or even RGB orthophotos, can also yield highly accurate depth estimates. Where sufficient training data are available, HIRBERT could become a preferred technique because regression tree ensembles are capable of using information from multiple spectral bands to capture more complex, non-linear relationships between depth and reflectance across a range of turbidities. The results of this study corroborate those reported by Kwon et al. (2024), who showed that a similar random forest algorithm enabled depth retrieval in a highly turbid deltaic environment up to 30 m deep where OBRA performed poorly.

An additional advantage of HIRBERT is the ability to gain insight regarding the relationships between depth and reflectance that underlie the model's predictions. Legleiter et al. (2025b) describe a number of graphical tools that can facilitate this type of interpretation but in this study we focus on one that is particularly informative: Shapley summary plots like those shown in Fig. 17. As described in Sections 2.4.2 and 2.5.3, an important step in the HIRBERT workflow involves ranking predictor variables in order of importance; this information is used to sort bands from most to least important along the vertical axis of a Shapley summary plot. The Shapley values on the horizontal axis then indicate how predicted depths vary as reflectance changes in each band in the model. For each of the points on the plot, which represent individual image pixels, the Shapley value for a given band

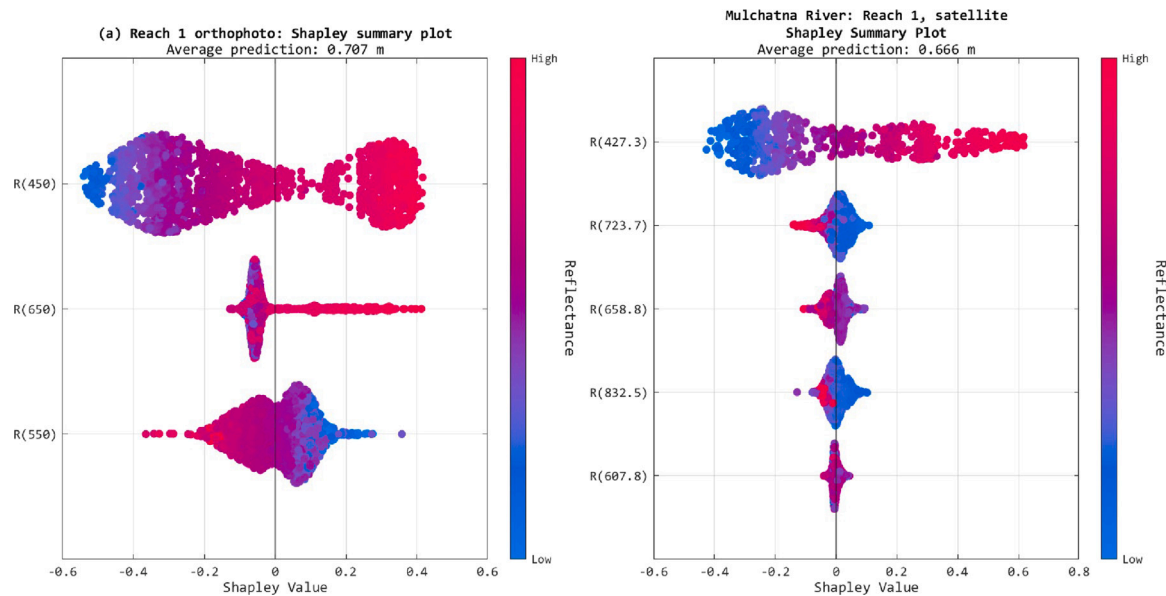


Fig. 17. Shapley summary plots for the (a) orthophoto and (b) satellite image from reach 1 yield insight on how relationships between depth and reflectance inform model predictions.

quantifies the contribution of that band to the deviation from the average depth prediction listed at the top of the plot. Coloring the points by reflectance shows whether these deviations are associated with relatively dark (blue tones in Fig. 17) or bright (red tones) pixels for that band. For example, for the HIRBERT model developed from the orthophoto, the blue band was the most important and the cluster of reddish-colored points with positive Shapley values indicates that pixels with high reflectance in the blue band tend to be associated with depths greater than the average prediction of 0.707 m. Similarly, the most important predictor in the model derived from the multispectral satellite image was also a blue wavelength (427.3 nm) and high reflectances in this band were again associated with above-average depths. A NIR band at 723.7 nm was the second most important in this model and at this wavelength negative Shapley values for pixels with relatively high reflectance led to below-average depth predictions. This finding is consistent with the preceding discussion in the context of OBRA: water absorbs strongly in the NIR, such that deeper water appears darker and shallow areas appear brighter due to greater reflectance from the streambed. The HIRBERT approach thus captured both direct and inverse relationships between depth and reflectance to enable accurate bathymetric mapping in an optically complex proglacial river environment.

4.3. Toward efficient use of remote sensing techniques

For the first two specific aims listed at the end of Section 1, results from this study suggest that reliable information on the bathymetry of a proglacial river could be derived from either high spatial resolution RGB orthophotos or multispectral satellite images consisting of eight bands but with a coarser pixel size. In deciding which of these data types to use, cost and complexity become important considerations. Acquiring the orthophotos required an aircraft with a camera and a trained pilot and involved careful flight planning and labor-intensive post-processing. In this case, the equipment, personnel, and software were all provided by the NPS, but each of these items represents a significant expense. Coordinating the flight with a ground-based survey also proved challenging, particularly given the often unfavorable weather conditions in this part of Alaska, and ultimately did not occur until a month after the field measurements were made. Although this timing issue did not adversely affect our study due to stable water levels that were modulated by proximity to a large lake, such a lag would

be problematic in a more dynamic river environment. For all of these reasons, acquiring data from a satellite can be an attractive option. Multispectral sensors like WV are operated by commercial vendors from which images can be purchased or, in some cases, obtained at no cost through agreements between vendors and end users. In addition, satellite images are often provided as analysis-ready data products that do not require extensive additional processing prior to depth retrieval. Constraints associated with satellite data include the need to avoid cloud cover, relatively large pixel sizes (2 m for WV vs. 0.1 for the orthophotos), and less reliable geo-referencing. The latter is particularly problematic and in our experience WV data often do not align well with field surveys and must be co-registered to an independent base image. In this study, we used the orthophoto to geo-reference the satellite image but in many locations publicly available base layers could be used for this purpose.

Moving on to our third aim, we found consistently strong agreement between predicted and observed depths for six different depth retrieval algorithms. Although performance varied to some degree across the four sites (three individual reaches and the full study area) and two sensors (orthophotos and multispectral satellite images) we considered, as summarized in Figs. 11–13, averaging over these two factors (site and sensor) to produce Fig. 14 led to high OP R^2 values, ranging from 0.78 for the exponential version of FREEBIRD to 0.89 for HIRBERT. By this metric alone, the choice of algorithm would seem to matter little, suggesting that methods with less onerous field data requirements, such as FREEBIRD and IDQT, might estimate water depths just as accurately as those based on paired observations of depth and reflectance, OBRA and HIRBERT.

However, examining the other metrics of performance in Fig. 14 leads to an important qualifier: although the six methods had similar levels of overall agreement, errors tended to be larger for the techniques that were less reliant upon field data. Normalized MAE and RMSE values for the two variants of FREEBIRD and IDQT were approximately twice as large (30 or 40% of the reach-averaged mean depth) as those for OBRA and HIRBERT (15%–20%). The most important difference was for the mean bias, which was less than 5% for OBRA and negligible for HIRBERT but larger and more variable across sites and sensors for FREEBIRD and especially IDQT.

The original version of FREEBIRD, based on a linear relation between X and d , tended to yield relatively modest over-predictions of depth (i.e., negative values of the mean bias), whereas the new

Table 4

Advantages and limitations of depth retrieval methods. OBRA = optimal band ratio analysis, HIRBERT = hyperspectral imaging of river bathymetry using an ensemble of regression trees, ML = machine learning, FREEBIRD = flow resistance equation based imaging of river depths, IDQT = image-to-depth quantile transformation.

Method	Advantages	Limitations
OBRA	Well-established technique Clear physical basis Computationally efficient Cross-validation version for small sample sizes	Requires field measurements Only considers two spectral bands Assumes X vs. d functional form Sensitive to variations in water column optical properties, bottom composition, and surface reflectance
HIRBERT	Powerful ML-based algorithm Captures complex relationships between depth and reflectance Exploits data from all spectral bands	Large field data set needed to train Model interpretation can be difficult Computationally demanding
FREEBIRD	Depth measurements not required Not reliant on spectral information Connects remote sensing to open channel flow concepts	Discharge, water surface slope, and channel aspect ratio are key inputs Must choose X image and whether it is directly or inversely related to d Numerous hydraulic assumptions
IDQT	Depth measurements need not be geo-referenced or at same time as image Not reliant on spectral information Distribution can be parameterized based on mean and variance of depth	Could be sensitive to how depths are sampled in the field Must choose X image and whether it is directly or inversely related to d X and d must be monotonically related

formulation for an exponential X vs. d relation led to underestimates of depth (positive bias). These results suggest that the exponential version might have been overly sensitive, with small variations in depth leading to excessively large changes in depth, and imply that the linear implementation was more robust.

For IDQT, the mean bias was large, negative, and variable, both when an empirical distribution of depths was used as input and when using the mean depth and the variance of scaled depths to parameterize a gamma distribution. This result implies that depths tended to be over-predicted by this method, typically by about 30%. The cause of this persistent bias is not immediately obvious but might be related to the greater variance in X than in d , as indicated by the flatter CDF in Fig. 9c than in Fig. 9d. One reason for this disparity might be that whereas X values were extracted evenly and exhaustively from throughout the image, depths might have been sampled unevenly in the field. If the distributions of X and d were more similar in shape, IDQT might yield more reliable depth estimates; further testing on other rivers would be required to evaluate this hypothesis.

The results of this study suggest that although remote sensing methods with less demanding field data requirements could provide image-derived depth estimates that are strongly correlated with depths measured in the field, these techniques also tend to yield larger errors and are more susceptible to systematic biases. Our findings imply that in practice a tradeoff might need to be made between the level of effort invested in the field and the accuracy of the remotely sensed bathymetry. To help inform selection of an appropriate methodology, the advantages and limitations of the various depth retrieval techniques we evaluated are summarized in Table 4. An important factor to consider is whether general knowledge of spatial patterns of bathymetry is sufficient or if the application at hand requires more precise information on the absolute magnitude of the depth. If a relative depth map is adequate, methods like FREEBIRD and IDQT might be viable options whereas more field data-intensive techniques like OBRA or HIRBERT might be more appropriate if the goal is to obtain highly accurate absolute depths. The results summarized in Section 3.4 suggest that the cross-validation-based implementation of OBRA might be a reasonable compromise capable of providing reliable depth estimates based on field data from a single transect. However, this method might also be sensitive to image spatial resolution through its impact on sample size, with coarser pixels leading to a smaller number of pixel-scale mean depths for calibration. To make the techniques implemented

in this study conveniently available to end users, we have developed a free, open-source, standalone software package for this purpose: Depth Retrieval Algorithms for Mapping and Exploration of Rivers, or DReAMER for short (Fig. 18; Legleiter, 2025).

5. Conclusion

Remote sensing has emerged as an increasingly viable option for characterizing fluvial systems, with a large body of research demonstrating the ability to map river bathymetry from passive optical image data, for example. Spatially distributed depth information of this kind can help to inform a range of applications including habitat assessment, flow and sediment transport modeling, and various aspects of water resource management. In this study, we assessed the potential to infer water depths along an inaccessible proglacial river from two common types of remotely sensed data via a suite of retrieval algorithms with varying input data requirements. One of our primary objectives was to evaluate approaches that might improve efficiency by reducing the level of field effort needed to support remote sensing of river bathymetry. Based on the results of our analysis, which involved examining four sites, two sensors, and six different algorithms for a total of 48 distinct scenarios, we arrived at the following conclusions:

1. Remote sensing of water depth along this particular proglacial river was not only feasible but highly accurate, with strong agreement between predicted and observed depths (R^2 up to 0.94). In contrast to previous studies in clear-flowing streams, spectrally based bathymetric mapping along this proglacial river was enabled by a direct relationship between depth and reflectance, with greater scattering by fine-grained suspended sediment where the water column was thicker.
2. Reliable depth information was derived from both RGB orthophotos acquired from an aircraft and an eight-band multi-spectral image obtained from a satellite.
3. Algorithms that rely upon paired observations of depth and reflectance for model calibration and training led to the most accurate depth estimates, with mean absolute errors and root mean squared errors on the order of 15%–20% of the reach-averaged mean depth and biases less than about 5%.
4. Although alternative hydraulic and statistical methods that do not require any field measurements of depth in one case or only

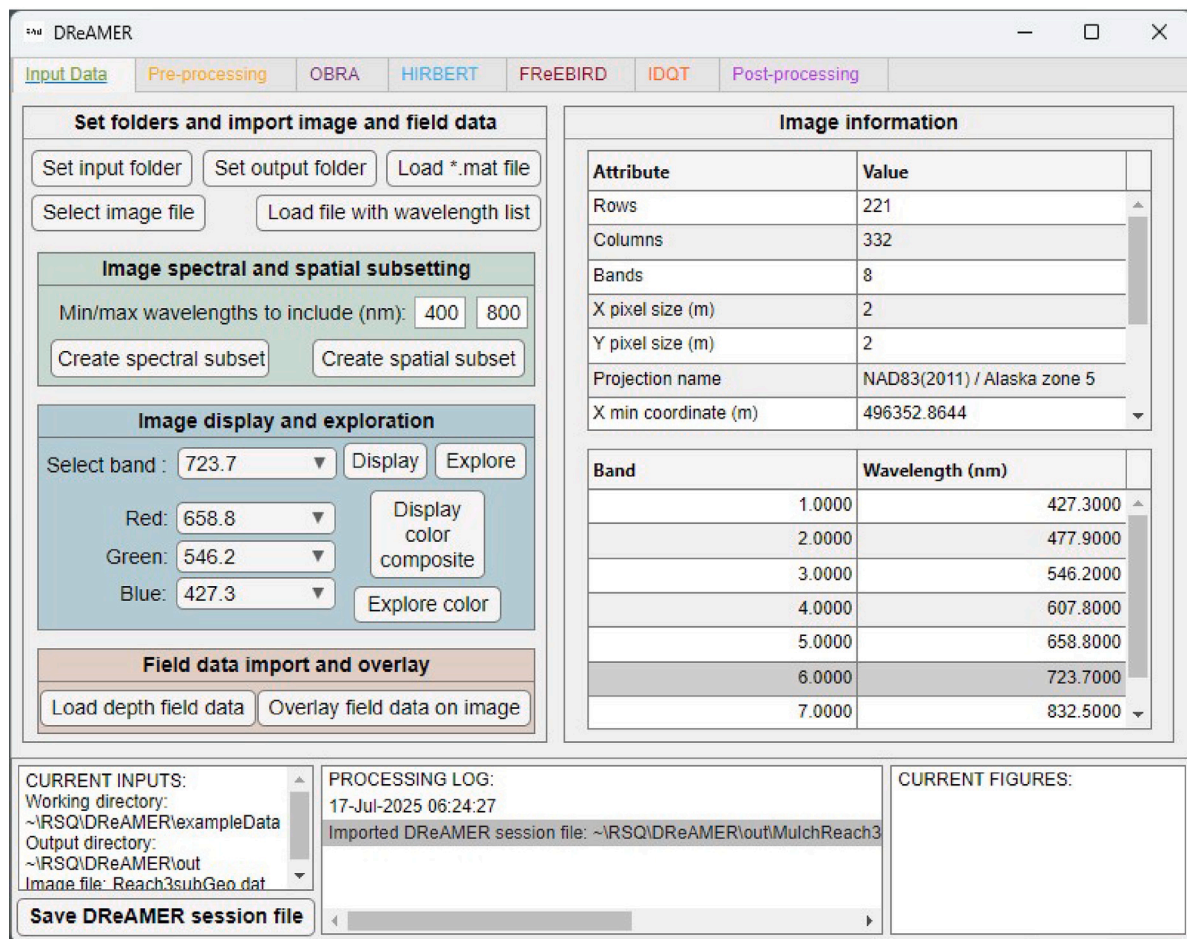


Fig. 18. User interface for the Depth Retrieval Algorithms for Mapping and Exploration of Rivers (DReAMER) software package developed to make the techniques employed in this study available to end users (Legleiter, 2025).

a limited, non-georeferenced sample in the other can yield strong correlations between predicted and observed depths, these techniques are more susceptible to systematic biases toward under- or over-estimation of depth.

5. A progressively truncated input depths analysis indicated that both the camera mounted on the aircraft and the imaging system deployed on the satellite were capable of detecting the full range of depths observed within our study area.
6. A cross-validation version of optimal band ratio analysis intended for small samples from a single transect provided depth estimates that were just as accurate as those based on a much larger field data set, but this approach is better-suited to higher spatial resolution image data.
7. In selecting a method, compromises might need to be made between the level of field effort required and the accuracy of the resulting image-derived bathymetry. If general knowledge of the spatial pattern of depths within a reach is sufficient, less data-intensive techniques might be appropriate, but if accurate absolute depths are desired additional field data could be necessary.
8. A standalone software package for implementing a variety of depth retrieval algorithms is freely available.

CRediT authorship contribution statement

Carl J. Legleiter: Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Christina M. Leonard:** Writing – review & editing,

Validation, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Paul A. Burger:** Writing – review & editing, Validation, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Addison G. Pletcher:** Writing – review & editing, Resources, Data curation. **Paul J. Kinzel:** Writing – review & editing, Validation, Resources, Investigation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Given Carl Legleiter's role as a member of the editorial board for Geomorphology, he had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Chad Hults of the National Park Service Alaska Region facilitated collection of remotely sensed data and subsequent image processing. William Edwards of the National Park Service assisted with field data collection. Merritt Harlan provided advice on obtaining and processing SWOT data. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Calculation of water level changes

To assess how the lake level fluctuations described in Section 2.3.3 translated into variations in flow depth within our study area, we assumed that the Mulchatna River exits the lake as an infinitely long, mild channel. Based on this assumption, we then calculated the normal flow depth and discharge by simultaneously solving the normal flow and energy equations (Henderson, 1966). For convenience, we further assumed that the channel was rectangular and calculated the energy per unit width as

$$E = y_n + \frac{q^2}{2gy_n^2}, \quad (\text{A.1})$$

where E is the vertical distance between the lake water surface and the channel bed, y_n is the normal depth, g is gravitational acceleration, and q is the unit discharge. Unit discharge was defined via Manning's equation as

$$q = \frac{1}{n} y_n^{5/3} S^{1/3}, \quad (\text{A.2})$$

where n is Manning's roughness, taken to be 0.03, and S is the channel slope (0.0035 for reach 1) (Henderson, 1966). If we assume the vertical distance between the channel bed and lake level was 0.7 m on July 7, resulting in a similar discharge as that measured on July 14, we calculated that a 0.12 m drop in water level would decrease the normal depth within the river by 0.11 m, corresponding to a 0.2 m³/s reduction in the unit discharge.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.geomorph.2025.110140>.

Data availability

The data used in this study is available through a USGS ScienceBase data release cited in the paper (<https://doi.org/10.5066/P14F2LE9>).

References

- Bjerklie, D.M., Birkett, C.M., Jones, J.W., Carabajal, C., Rover, J.A., Fulton, J.W., Garambois, P.-A.A., 2018. Satellite remote sensing estimation of river discharge: Application to the Yukon River Alaska. *J. Hydrol.* 561, 1000–1018. <https://doi.org/10.1016/j.jhydrol.2018.04.005>.
- Bjerklie, D.M., Dingman, S.L., Bolster, C.H., 2005. Comparison of constitutive flow resistance equations based on the Manning and Chezy equations applied to natural rivers. *Water Resour. Res.* 41 (W11502), <https://doi.org/10.1029/2004WR003776>.
- Blaskey, D., Koch, J.C., Gooseff, M.N., Newman, A.J., Cheng, Y., O'Donnell, J.A., Musselman, K.N., 2023. Increasing Alaskan river discharge during the cold season is driven by recent warming. *Environ. Res. Lett.* 18 (2), 024042. <https://doi.org/10.1088/1748-9326/acb661>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/a:1010933404324>.
- Carbonneau, P., Fonstad, M.A., Marcus, W.A., Dugdale, S.J., 2012. Making riverscapes real. *Geomorphology* 137 (1), 74–86. <https://doi.org/10.1016/j.geomorph.2010.09.030>.
- Conaway, J., Eggleston, J., Legleiter, C., Jones, J., Kinzel, P., Fulton, J., 2019. Remote sensing of river flow in Alaska—New technology to improve safety and expand coverage of USGS streamgaging. *U.S. Geol. Surv. Fact Sheet* 2019 (3024), 4. <https://doi.org/10.3133/fs20193024>.
- Dilbone, E., Legleiter, C.J., Alexander, J.S., McElroy, B., 2018. Spectrally based bathymetric mapping of a dynamic, sand-bedded channel: Niobrara River, Nebraska, USA. *River Res. Appl.* 34 (5), 430–441. <https://doi.org/10.1002/rra.3270>.
- Green, A.A., Berman, M., Switzer, P., Craig, M.D., 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geosci. Remote Sens.* 26 (1), 65–74. <https://doi.org/10.1109/36.3001>.
- 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 Sens. Environ.* 285, 113378. <https://doi.org/10.1016/j.rse.2022.113378>.
- He, J., Zhang, S., Cui, X., Feng, W., 2024. Remote sensing for shallow bathymetry: A systematic review. *Earth-Sci. Rev.* 258, 104957. <https://doi.org/10.1016/j.earscirev.2024.104957>.
- Henderson, F., 1966. *Open Channel Flow*. John Wiley & Sons, New York.
- Jiang, H., Rutherford, I., 2024. Potential to use free satellite imagery to retrieve the past bathymetry of large rivers. *Remote. Sens. Appl.: Soc. Environ.* 33, 101133. <https://doi.org/10.1016/j.rsase.2023.101133>.
- Kääb, A., Altena, B., Mascaro, J., 2019. River-ice and water velocities using the Planet optical cubesat constellation. *Hydrol. Earth Syst. Sci.* 23 (10), 4233–4247. <https://doi.org/10.5194/hess-2019-62>.
- Kwon, S., Passalacqua, P., Soloy, A., Jensen, D., Simard, M., 2024. Depth mapping in turbid and deep waters using AVIRIS-NG imagery: A study in Wax Lake Delta, Louisiana, USA. *Water Resour. Res.* 60 (11), <https://doi.org/10.1029/2023WR036875>, e2023WR036875.
- Lanik, A., Rogers, J., Karpilo, R., 2021. Lake Clark National Park and Preserve: Geologic Resources Inventory Report. Technical Report NPS/NRSS/GRD/NRR—2021/2331, National Park Service, Fort Collins, Colorado, <https://doi.org/10.36967/nrr-2288490>.
- Legleiter, C., 2013. Mapping river depth from publicly available aerial images. *River Res. Appl.* 29 (6), 760–780. <https://doi.org/10.1002/rra.2560>.
- Legleiter, C., 2015. Calibrating remotely sensed river bathymetry in the absence of field measurements: Flow resistance equation-based imaging of river depths (FREEBIRD). *Water Resour. Res.* 51 (4), 2865–2884. <https://doi.org/10.1002/2014wr016624>.
- Legleiter, C., 2016. Inferring river bathymetry via image-to-depth quantile transformation (IDQT). *Water Resour. Res.* 52 (5), 3722–3741. <https://doi.org/10.1002/2016wr018730>.
- Legleiter, C., 2025. DREAMER (ver. 1.0.0, August, 2025): U.S. Geological Survey software release. <https://doi.org/10.5066/P1FZWQTX>.
- Legleiter, C., Harrison, L.R., 2019. Remote sensing of river bathymetry: Evaluating a range of sensors, platforms, and algorithms on the Upper Sacramento River, California, USA. *Water Resour. Res.* 55 (3), 2142–2169. <https://doi.org/10.1029/2018WR023586>.
- Legleiter, C.J., Harrison, L.R., 2024. Evaluating the potential for efficient, UAS-based reach-scale mapping of river channel bathymetry from multispectral images. *Front. Remote. Sens.* 5, 1305991. <https://doi.org/10.3389/frsen.2024.1305991>.
- Legleiter, C.J., Kinzel, P.J., Overstreet, B.T., Harrison, L.R., 2025b. Hyperspectral imaging of river bathymetry using an ensemble of regression trees. *Earth Surf. Process. Landf.* 50 (12), e70155. <https://doi.org/10.1002/esp.70155>.
- Legleiter, C., Leonard, C., Burger, P., Pletcher, A., Kinzel, P., 2025a. Field measurements of water depth and remotely sensed data from the Mulchatna River, Alaska, collected July 12–August 14, 2024: U.S. Geological Survey data release. <https://doi.org/10.5066/P14F2LE9>.
- Legleiter, C., Overstreet, B.T., 2012. Mapping gravel bed river bathymetry from space. *J. Geophys. Res. - Earth Surf.* 117, F04024. <https://doi.org/10.1029/2012jf002539>.
- Legleiter, C., Overstreet, B., Kinzel, P., 2018. Sampling strategies to improve passive optical remote sensing of river bathymetry. *Remote. Sens.* 10 (6), 935. <https://doi.org/10.3390/rs10060935>.
- Legleiter, C.J., Overstreet, B.T., Kinzel, P.J., 2024. Integrating depth measurements from gaging stations with image archives for spectrally based remote sensing of river bathymetry. *Water Resour. Res.* 60 (7), <https://doi.org/10.1029/2024WR037295>, e2024WR037295.
- Legleiter, C., Roberts, D.A., Lawrence, R.L., 2009. Spectrally based remote sensing of river bathymetry. *Earth Surf. Process. Landf.* 34 (8), 1039–1059. <https://doi.org/10.1002/esp.1787>.
- Legleiter, C., Roberts, D.A., Marcus, W.A., Fonstad, M.A., 2004. Passive optical remote sensing of river channel morphology and in-stream habitat: Physical basis and feasibility. *Remote Sens. Environ.* 93 (4), 493–510. <https://doi.org/10.1016/j.rse.2004.07.019>.
- Leonard, C., Legleiter, C., Overstreet, B., 2017. Effects of lateral confinement in natural and leveed reaches of a gravel-bed river: Snake River, Wyoming, USA. *Earth Surf. Process. Landf.* 42 (13), 2119–2138. <https://doi.org/10.1002/esp.4157>.
- Lim, J.S., 1990. *Two-Dimensional Signal and Image Processing*. Prentice Hall, Englewood Cliffs, NJ.
- Marcus, W.A., Fonstad, M.A., 2010. Remote sensing of rivers: The emergence of a subdiscipline in the river sciences. *Earth Surf. Process. Landf.* 35 (15), 1867–1872. <https://doi.org/10.1002/esp.2094>.
- Michalek, A.T., Villarini, G., Prein, A.F., Done, J.M., Johnson, D.R., Wang, C., 2025. Precipitation- and temperature-driven future changes to flooding in Alaska. *Geophys. Res. Lett.* 52 (3), <https://doi.org/10.1029/2024GL112004>.
- Mobley, C.D., 1994. *Light and Water: Radiative Transfer in Natural Waters*. Academic Press, San Diego.
- Mueller, D., 2020. QRev, U.S. Geological Survey software release. <https://doi.org/10.5066/P9OZ8QDL>.
- Natural Earth, 2025. Natural Earth. URL <https://www.naturalearthdata.com/>.
- NPS, 2022. Digital surficial geologic-GIS map of Lake Clark National Park and Preserve and vicinity, Alaska (NPS, GRD, GRI, LACL, LACL surficial digital map) adapted from a U.S. Geological Survey open-file report map by Wilson, Bundtzen, Harvey and Spencer (2009). URL <https://irma.nps.gov/DataStore/Reference/Profile/2266406>.

- NPS, 2025. What we do (U.S. National Park Service). URL <https://www.nps.gov/aboutus/index.htm>.
- O'Donnell, J.A., Carey, M.P., Koch, J.C., Baughman, C., Hill, K., Zimmerman, C.E., Sullivan, P.F., Dial, R., Lyons, T., Cooper, D.J., Poulin, B.A., 2024. Metal mobilization from thawing permafrost to aquatic ecosystems is driving rusting of Arctic streams. *Commun. Earth Environ.* 5 (1), 268. <http://dx.doi.org/10.1038/s43247-024-01446-z>.
- Piégay, H., Arnaud, F., Belletti, B., Bertrand, M., Bizzi, S., Carbonneau, P., Dufour, S., Liébault, F., Ruiz-Villanueva, V., Slater, L., 2020. Remotely sensed rivers in the Anthropocene: State of the art and prospects. *Earth Surf. Process. Landf.* 45 (1), 157–188. <http://dx.doi.org/10.1002/esp.4787>.
- Pineiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M., Pi, G., Perelman, S., Guerschman, J.P., 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecol. Model.* 216 (3–4), 316–322. <http://dx.doi.org/10.1016/j.ecolmodel.2008.05.006>.
- Rowland, J.C., Schwenk, J.P., Shelef, E., Muss, J., Ahrens, D., Stauffer, S., Pilliouras, A., Crosby, B., Chadwick, A., Douglas, M.M., Kemeny, P.C., Lamb, M.P., Li, G.K., Vulis, L., 2023. Scale-dependent influence of permafrost on riverbank erosion rates. *J. Geophys. Res.: Earth Surf.* 128 (7), <http://dx.doi.org/10.1029/2023JF007101>, e2023JF007101.
- Segura, C., Pitlick, J.C.J.F., 2015. Coupling fluvial-hydraulic models to predict gravel transport in spatially variable flows. *J. Geophys. Res.: Earth Surf.* 120 (5), 834–855. <http://dx.doi.org/10.1002/2014jf003302>.
- SWOT, 2023. SWOT level 2 water mask pixel cloud data product, version 1.1. URL https://podaac.jpl.nasa.gov/dataset/SWOT_L2_HR_PIXC_1.1.
- Tamminga, A., Eaton, B., Mossop, B., 2025. Evaluating satellite-based depth mapping for large river monitoring: A case study of Peace River, British Columbia, Canada. *Geomorphica* 2 (1), <http://dx.doi.org/10.59236/geomorphica.v2i1.27>.
- Turnipseed, D., Sauer, V., 2010. Discharge measurements at gaging stations. U.S. Geol. Surv. Tech. Methods Book 3 (Chap. A8), 87, URL <https://pubs.usgs.gov/tm/tm3-a8/>.
- Vowels, L.K., Armstrong, W.H., Overeem, I., McGrath, D., Rick, B., Dye, A., Martin, D., 2025. Investigating changes in proglacial stream suspended sediment concentration and their drivers using large scale remote sensing. *Geomorphology* 475, 109664. <http://dx.doi.org/10.1016/j.geomorph.2025.109664>.
- Wohl, E., Scamardo, J.E., 2022. Aufeis as a major forcing mechanism for channel avulsion and implications of warming climate. *Geophys. Res. Lett.* 49 (20), <http://dx.doi.org/10.1029/2022GL100246>.
- Yang, Y., Feng, D., Beck, H.E., Hu, W., Abbas, A., Sengupta, A., Delle Monache, L., Hartman, R., Lin, P., Shen, C., Pan, M., 2025. Global daily discharge estimation based on grid long short-term memory (LSTM) model and river routing. *Water Resour. Res.* 61 (6), <http://dx.doi.org/10.1029/2024WR039764>.