Remote Sensing of Environment 304 (2024) 114047

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

A trained Mask R-CNN model over PlanetScope imagery for very-high resolution surface water mapping in boreal forest-tundra

Pedro Freitas [a,b,*], Gonçalo Vieira [a,b], João Canário [b,c], Warwick F. Vincent [b,d], Pedro Pina [e,f], Carla Mora [a]

[a] Centro de Estudos Geográficos, Laboratório Associado TERRA, Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa, 1600-276 Lisbon, Portugal
[b] Centre d'études nordiques (CEN), Université Laval, Quebec City, QC G1V 0A6, Canada
[c] Centro de Química Estrutural, Institute of Molecular Sciences and Department of Chemical Engineering, Instituto Superior Técnico, Universidade de Lisboa, Portugal
[d] Département de biologie & Takuvik, Université Laval, Quebec City, QC G1V 0A6, Canada
[e] Departamento de Ciências da Terra, Universidade de Coimbra, 3030-790 Coimbra, Portugal
[f] Instituto Dom Luiz, Universidade de Coimbra, 3030-790 Coimbra, Portugal

ARTICLE INFO

Edited by Menghua Wang

Keywords:
Mask R-CNN
Deep learning
PlanetScope
Arctic and subarctic
Water mapping
Small water bodies

ABSTRACT

Small water bodies (< 0.01 km^2) showing diverse limnological properties occur in great abundance across the boreal forest and tundra landscapes of the Arctic and Subarctic. However, their classification, geographical distribution and collective importance for water, heat, nutrient, contaminant and carbon cycles are still poorly constrained. One important step for better understanding the role and evolution of small water bodies in the fast-changing northern landscapes is to develop image analysis protocols that allow their automatic remote sensing detection, delineation and inventory. In this study, we set an image analysis protocol (High Latitude Water – HLWATER V1.0) based on a trained supervised Mask R-CNN deep learning model over PlanetScope imagery for the automatic detection and delineation of small lakes and ponds that were absent in existing datasets. Most of our training dataset comprised water bodies smaller than 0.01 km^2 (97%) and spanned a wide range of environmental and hydrological settings, from the sporadic to the continuous permafrost zones of Canada. The model was tested as a fully autonomous approach for eastern Hudson Bay, Nunavik (Subarctic Canada), a region that poses challenges for water remote sensing given the abundance and variety of small water bodies. These are mainly permafrost thaw and glacial basin ponds in the boreal forest-tundra in challenging optical settings influenced by vegetation or topography shadowing, or revealing peat water logging, fen and bog pond conditions. A multi-scale validation approach was developed using water body delineations from PlanetScope imagery and ultra-high resolution orthomosaics from Unoccupied Aerial Systems. This procedure allowed a sub-pixel assessment and identified the limitations and strengths of the trained model for detecting small and large water bodies. The results varied according to different landscape units, with mean Intersection over Union (IoU) 0.5 F1 Scores of 0.53 to 0.71 and mean F1 Scores of 0.62 to 0.95. Considering 166 m^2 as the minimum pond size detection threshold, the IoU 0.5 F1 Scores were 0.7 to 0.91 and F1 Scores were 0.76 to 0.83, evaluated by comparing the model results with ultra-high resolution manual delineations. The image analysis protocol and trained model show high potential for extension to other boreal forest-tundra regions of the Arctic and Subarctic, allowing for detailed inventories of optically and morphologically diverse small water bodies over large areas of the circumpolar North.

* Corresponding author at: Instituto de Geografia e Ordenamento do Território, Rua Branca Edmée Marques, Universidade de Lisboa, 1600-276 Lisboa, Portugal.
E-mail addresses: pedro-freitas@edu.ulisboa.pt (P. Freitas), vieira@edu.ulisboa.pt (G. Vieira), joao.canario@tecnico.ulisboa.pt (J. Canário), Warwick.Vincent@bio.ulaval.ca (W.F. Vincent), ppina@dct.uc.pt (P. Pina), carlamora@edu.ulisboa.pt (C. Mora).

https://doi.org/10.1016/j.rse.2024.114047
Received 17 October 2023; Received in revised form 17 January 2024; Accepted 7 February 2024
Available online 13 February 2024

1. Introduction

1.1. Polar water bodies: importance and knowledge gaps

The boreal forest-tundra transition zone spans the sporadic and discontinuous permafrost (perennially frozen ground) regions of the Arctic and Subarctic (Obu, 2021), and its landscapes and ecosystems are especially vulnerable to climate change (Serreze et al., 2009; Biskaborn et al., 2019; Box et al., 2019). Increasing air temperatures (Rantanen et al., 2022) and precipitation (McCrystall et al., 2021) and the associated degradation of ice-rich permafrost (Leppiniemi et al., 2023), are affecting the ground stability, the structure and functioning of vegetation communities including shrubification and northward migration of the boreal forest (Myers-Smith et al., 2020), the hydrology with lake formation or drainage, the limnology such as the browning of lakes (Nitze et al., 2018; Wauthy et al., 2018; Vonk et al., 2019), the energy balance including via land surface albedo, as well as the nutrient, contaminant and carbon cycles (Tarnocai et al., 2009; Schuster et al., 2018; Hugelius et al., 2020). These impacts have global as well as local consequences (Schaefer et al., 2014; Schuur et al., 2015; Turetsky et al., 2020; Miner et al., 2021).

Polar lakes and rivers are sentinels of state shifts in climate such as snow/rainfall, the freeze-thaw regime and water balance, and ecosystem properties such as soil and vegetation (Bouchard et al., 2017; Wauthy et al., 2018). Through their outflows and inflows, as well as their physical, biological and biogeochemical properties changes over time, high latitude lakes reflect not only the biotic and abiotic processes within the water bodies but also the characteristics and evolution of the surrounding landscape, from local to regional scales (Vincent and Laybourn-Parry, 2008; Webb et al., 2022). These basins of water surrounded by land vary in size from ponds (< 0.01 km^2) to lakes (> 0.01 km^2), and are important sources of greenhouse gases, as well as centers of biological activity for aquatic food webs (Vincent and Laybourn-Parry, 2008; Vincent, 2018; Walter Anthony et al., 2018).

Polar inland water bodies are derived from a variety of origins, including wetland (e.g., peatland lakes), floodplain (e.g., oxbow lakes), volcanic (e.g., maar lakes), tectonic, meteoritic (e.g., impact crater lakes), alpine, karst, thermokarst (e.g., permafrost thaw/thermokarst lakes and ponds) and coastal uplift (e.g., meromictic lakes) systems (Pienitz et al., 2008). These waters also show diverse limnological and morphological characteristics resulting from climatic, geological, biogeographical and topographical gradients (Vincent and Laybourn-Parry, 2008).

Although successful remote sensing strategies have been developed for globally inventorying and tracking large water body surface extent changes over time (Verpoorter et al., 2014; Pekel et al., 2016; Pickens et al., 2020), existing datasets do not yet include ponds and narrow streams. This has been due to the poor detection of diverse and small water features in contrasting environmental settings, in conjunction with sensor spatial resolution, acquisition conditions (e.g., cloud cover, optical thickness, solar radiation) and data availability limitations (Adegun et al., 2023; Harlan et al., 2023; Mullen et al., 2023). These data and model omissions are evident in the high latitudes of the northern hemisphere, where ice-sheet dynamics in the Pleistocene gave rise to a postglacial landscape with widespread water bodies, including many ponds (Smith et al., 2007). This is particularly noticeable in the lowland permafrost regions, which contain the largest surface water fraction compared to any other terrestrial biome of the world (Lehner and Döll, 2004; Pekel et al., 2016; Webb et al., 2022), with an enormous abundance of small-size ponds (Vincent and Laybourn-Parry, 2008; Muster et al., 2013).

Due to enhanced remote sensing imagery and classification algorithms, an improved survey and inventory is now possible and needed. An example of the relevance of such a dataset is the need for characterizing and monitoring the small water bodies in the boreal forest-tundra ecozone, allowing to better understand their formation and roles of permafrost thaw (thermokarst) ponds in the biogeochemical carbon cycle (Negandhi et al., 2013; Heslop et al., 2020; Zandt et al., 2020). These numerous and widespread ponds result from abrupt permafrost thaw (Walter Anthony et al., 2018) and generally exhibit sizes below 0.01 km^2 and depths under 5 m, and are much more active biogeochemically than larger lakes (Abnizova et al., 2012; Bégin and Vincent, 2017; Arsenault et al., 2022). Throughout their life span, they show diverse optical and morphological dynamics and are biogeochemical hotspots for the release of carbon dioxide (CO_2), nitrous oxide (N_2O) and especially methane (CH_4), from permafrost to the atmosphere, through microbial and photochemical transformations (Breton et al., 2009; Edwards et al., 2009; Vonk et al., 2015; Zandt et al., 2020). Accounting for the role of these ponds is increasingly important due to permafrost warming, which generates more thaw ponds, potentially increasing greenhouse gas emissions (Holgerson and Raymond, 2016; Kuhn et al., 2018; Heslop et al., 2020). This concern is echoed globally, with Holgerson and Raymond (2016) demonstrating that even very small ponds can significantly contribute to CO_2 and CH_4 global emissions, due to factors such as shallow waters, frequent mixing and strong inputs of terrestrial carbon relative to water volume, underscoring the urgent need for comprehensive global pond inventories. In the Arctic and Subarctic, the formation of permafrost thaw lakes and ponds can also impact Indigenous lifestyles and settlements, causing disruptions to hunting, residential areas and infrastructure such as roads, railways and airstrips (Crate et al., 2017; Povoroznyuk et al., 2023; Tanguy et al., 2023).

Muster et al. (2017) developed the circum-Arctic Permafrost Region Pond and Lake (PeRL) database to better understand lake and pond spatial distribution and characteristics in the Arctic. Delineation of ponds and lakes ranging from 0.0001 km^2 to 1 km^2 was made using very-high resolution (< 5 m) satellite and aerial imagery. PeRL represents about 17% (1.4×10^6 km^2) of the Arctic lowlands and was a collaborative work among many scientific teams, improving hydrological dynamics predictions for the Arctic (Muster et al., 2019). However, PeRL is a compilation of feature polygons without a common methodological strategy and workflow. In addition, it covers only a small portion of the lakes and ponds that occur abundantly in Subarctic permafrost regions.

Cooley et al. (2019) used PlanetScope imagery to measure near-daily changes in lake surface area in wide regional sectors in Arctic-Boreal Canada, showing that small changes that were previously undetectable by coarser-resolution satellites are important to trace significant greenhouse gas emissions. However, due to the methodology followed based on the optimization of the Normalized Difference Water Index (NDWI; McFeeters, 1996), the model was not able to track pond (< 0.01 km^2) dynamics (Cooley et al., 2017).

Olefeldt et al. (2021) developed the Boreal-Arctic Wetland and Lake Dataset (BAWLD). The authors used random forest extrapolations to estimate fractional landcover classes within $0.5 \times 0.5^\circ$ grid cells. The model was fed by global datasets of climate, topography, lithology, permafrost conditions, vegetation, wetlands and surface water extent. For surface water extent, the authors used the Global Lakes and Wetland Dataset (GLWD) (Lehner and Döll, 2004), HydroLakes (Messager et al., 2016) and the Global Surface Water Extent (GSWE) (Pekel et al., 2016) datasets. Although these surface water extent products provide spatial consistency and generally ensure true positive quality, they have limitations for mapping small water bodies, more specifically ponds (< 0.01 km^2) (Feng et al., 2016; Muster et al., 2017; Pickens et al., 2020). A more detailed review of the available global, national and regional surface water extent products and their limitations is presented in the discussion of the current research.

More recently, Harlan et al. (2023) trained a You-Only-Look-Once (YOLO) deep learning model over PlanetScope imagery for detecting narrow beaded streams at the Arctic circumpolar scale. The river and stream catchments were further classified using Decision Trees. However, the implemented YOLO version only allowed for object detection, failing to precisely delineate the boundaries of these water bodies.

Mullen et al. (2023) focused on lake and pond mapping in Alaska using PlanetScope imagery and trained a UNet deep learning model. A total of 52,707 water bodies were mapped, 77% of which were ponds. Nevertheless, the model was not tested as a fully autonomous approach and manual cleaning had to be performed. The authors do not discuss in depth the factors impacting automatic water body delineations. The validation was done using a ground-based Global Navigation Satellite System (GNSS), but this is a time-consuming and costly method, with limitations in the number of water bodies delineated and constrained by access to their margins.

1.2. Earth observation data and machine learning advances

Given the abundance, remoteness and variety of northern water bodies, the complete inventorying and monitoring of these waters is not at all feasible through field campaigns and in-situ sampling. Remote sensing platforms are therefore needed to address the gaps. However, the fact that most of these water bodies are ponds imposes constraints for surveying their characteristics using continuous acquisitions from Earth Observation satellites (Muster et al., 2013). Beyond the Landsat series, since 1972, with 30 m spatial resolution worldwide acquisitions at a revisit time of 16 days, key advances have been provided by the Copernicus Program and Planet Labs. The former, since 2015 with the two twins Sentinel-2 satellites with worldwide acquisitions at 10/20-m resolution and a revisit time of 5 days, and since 2016, Planet Labs, with CubeSat nanosatellites such as the PlanetScope Dove, Dove-R and SuperDove, with daily acquisitions at ≈ 3 m resolution. These platforms provide data that is capable of offering major advances in the global detection and monitoring of lakes and ponds, as well as small rivers, streams, and creeks (Zeng et al., 2023).

The recent advances in Earth Observation data are allowing consistent and detailed water body delineations in space and time (Cooley et al., 2019; Qayyum et al.,2020; Mullen et al., 2023). These remote sensing advances have also been paralleled by the development of powerful non-linear (e.g., multilayer learning) algorithms for image classification (e.g., Mask R-CNN, U-Net, DeepLab, Fully Convolutional Network – FCN, Pyramid Scene Parsing Network – PSPNet) and time-series analysis through machine learning, computer vision and deep learning methods (Lecun et al., 2015; Sagan et al., 2020; Yuan et al., 2020). In deep learning most of the methods are inspired by the neural structure of the brain and its capacity to recognize features and patterns (Alzubaidi et al., 2021). In terms of new deep learning image classification techniques, some advantages compared to other traditional pixel or object-based methods are the capacity to accumulate knowledge, better generalization capabilities as well as improve the quality of the results and transfer them to analogous datasets (Huang et al., 2020; Robson et al., 2020; Alzubaidi et al., 2021; Nitze et al., 2021).

Although recent progress in deep learning has significantly impacted the remote sensing community, especially owed by the enhanced classification capabilities of satellite, airborne and Unmanned Aerial Systems (UAS) imagery (Ma et al., 2019; Yuan et al., 2020; Hao et al., 2021), this success is often constrained. This is due to the unavailability of extensive training datasets, intensive parameter optimization and high demand of costly computational resources (Ma et al., 2019; Yuan et al., 2020; Adegun et al., 2023). To mitigate some of these challenges, Vision Transformer algorithms (ViT) have been proposed. These algorithms incorporate self-attention mechanisms to facilitate the semantic association between the elements of a sequence, showing state-of-the-art performance in certain contexts (Bazi et al., 2021; Lin et al., 2021; Adegun et al., 2023; Aleissaee et al., 2023). However, ViT algorithms face scalability issues with increasing image size, leading to disproportionately high computational demands (Lin et al., 2021; Adegun et al., 2023). Furthermore, these algorithms are hard to train on small-scale datasets and, as a result, are prone to suffer from overfitting. Frequently, they demand more training data to be able to outperform other Convolution Neural Networks (CNN) and Recurrent Neural Networks (RNN) deep learning models (Lin et al., 2021).

In response to diverse challenges the development of hybrid deep learning models which seek on taking advantage of diverse architectures has been booming in the last years (Yuan et al., 2020; Aleissaee et al., 2023). A recent mediatic example of this trend is the release of the ambitious Segment Anything (SAM) model by Meta AI (Kirillov et al., 2023). This advanced foundation model relies on a variety of neural network architectures, including ViT, CNN, and Generative Adversarial Networks (GAN) for zero-shot and few-shot generalizations (Kirillov et al., 2023). SAM was trained on the SA-1B dataset, corresponding to 1 billion masks collected over 11 million images (Kirillov et al., 2023; Ji et al., 2023). Although it has proven surprising capabilities on remote sensing segmentation tasks (Osco et al., 2023), important drawbacks still remain. Some of the challenges include high user interaction requirements, fluctuating performance based on image complexity and characteristics, difficulties in segmenting objects against similar backgrounds, prominent foreground biases, and suboptimal segmentation of smaller or irregular objects, including inaccuracies in accommodating shadow effects (Ji et al., 2023).

Conversely, CNN have established dominance in diverse computer vision domains, including remote sensing (Ma et al., 2019; Yuan et al., 2020; Alem and Kumar, 2022; Aleissaee et al., 2023). One of the main advantages of using CNN is the autonomous ability to identify the relevant features without human supervision, producing meaningful outputs with reduced model implementation complexity. These characteristics have rendered them both popular and effective (Alzubaidi et al., 2021; Sarker, 2021), including for water resources applications (Sit et al., 2020). In addition, CNN networks are designed to improve generalization and avoid overfitting (Alzubaidi et al., 2021).

Notably, models like U-Net and Mask R-CNN built upon CNN architectures have been gaining widespread usage in remote sensing (Zhang et al., 2018; Huang et al., 2020; Nitze et al., 2021). In particular, Mask R-CNN, defined as a Region-based Convolutional Neural Network (R-CNN), stands out compared to similar models (Alfaro et al., 2019; Mohanty et al., 2020; Quoc et al., 2020) due to its simple structure making it computationally efficient, as well as offering outstanding performance on instance segmentation tasks, precisely delineating and classifying objects under diverse contexts and background conditions (Liu et al., 2019; He et al., 2020; Hao et al., 2021). It is an extension of the Faster R-CNN model (Ren et al., 2016) that allows not only the detection of objects through bounding boxes (branch for classification and bounding box regression), but also the segmentation and delineation of their boundaries by adding a branch (small FCN) on each Region of Interest (RoI) and a quantization-free-layer for pixel-to-pixel alignment (RoIAlign) (He et al., 2020; Xu et al., 2022). Upon its introduction, Mask R-CNN significantly outperformed existing models in COCO instance segmentation challenges, demonstrating its effectiveness in processing low-dimensional data and scalability in classifying objects of varying sizes and shapes (He et al., 2020). However, like most deep learning models, it demands a large number of training samples to optimize its learning and generalization capabilities (Alzubaidi et al., 2021; Karra et al., 2021; Sarker, 2021).

1.3. Strategy and objectives

Here, we tested the possibility of using native PlanetScope Dove (PS-D – Dove and Dove-R constellations) imagery at ≈ 3 m spatial resolution and 4-band spectral resolution (blue, green, red and near-infrared) for mapping the morphologically and optically diverse small water bodies of the boreal-forest tundra zone of Nunavik, Subarctic Canada. We present the results of a scalable supervised image analysis protocol (HLWATER V1.0), consisting of a methodological workflow based on a trained Mask R-CNN deep learning model over PS-D imagery. The objective is the automated detection and delineation of the small water bodies of the circumpolar North, contributing to significantly improve the available datasets. For training the model we constructed a very

large very high-resolution dataset of manually delineated lakes, ponds, rivers, streams, creeks and coastal sectors, located in diverse landscapes from the sporadic to continuous permafrost zones of Canada. HLWATER V1.0 is publicly available (doi:https://doi.org/10.5281/zenodo.10203553).

We evaluated the fully autonomous performance of the trained algorithm from a data user perspective instead of just referring to assessing the performance during model training. This acknowledges new advances in the development and release of deep learning models, namely that: i. Imagery consistent training samples are increasingly available and are key for achieving state-of-the-art results according to users' goals (Zhang et al., 2018; Alzubaidi et al., 2021; Sarker, 2021); ii. Intensive parametrization and coding are no longer mandatory (e.g., deep learning made simple and intermediate low-code interfaces) (Zhu et al., 2017; Alzubaidi et al., 2021); iii. Outstanding pre-trained (e.g., AlexNet, GoogleNet and ResNet) and fully trained models are increasingly available, with users needing to be able to develop robust ground truthing and testing strategies, especially for detecting small objects, for continuous improvement of the models and outputs (Liu et al., 2020; Sit et al., 2020; Alzubaidi et al., 2021).

For testing the performance of the trained algorithm, we developed a multi-scale validation approach using as reference not only detailed ground truthing from PS-D imagery, but also from ultra-high resolution UAS orthomosaics (0.05–0.15 m pixel size) in a variety of boreal landscapes. We aimed by this approach to assess the strengths and limitations of our trained model for detecting small water bodies, while also including larger ones such as glacial basin lakes and rivers, as well as better evaluate sub-pixel factors affecting the model output beyond PS-D native resolution (e.g., water body shape complexity, water bottom influences on reflectance, similarity with close objects). Finally, we considered model optimization strategies for fully autonomous implementations (confidence threshold trade-offs, minimum detection size assumptions), improvements compared to current global, national and regional products, and we addressed the limitations of this new trained model, as well as future research steps for ongoing improvement and application.

2. Study areas

For training, validating and testing the model, we used water bodies with high diversity of optical and morphological properties, and located in a range of environmental contexts: rock outcrops, alluvial plains, ice-wedge polygons, shrublands, wetlands, fens, bogs, peatlands and forests. These sectors are located along the boreal-forest tundra ecozone of Canada (Northwest Territories - Fort McPherson, Mackenzie Delta, Yukon and Tuktoyaktuk Peninsula, and Quebec - Eastern Hudson Bay

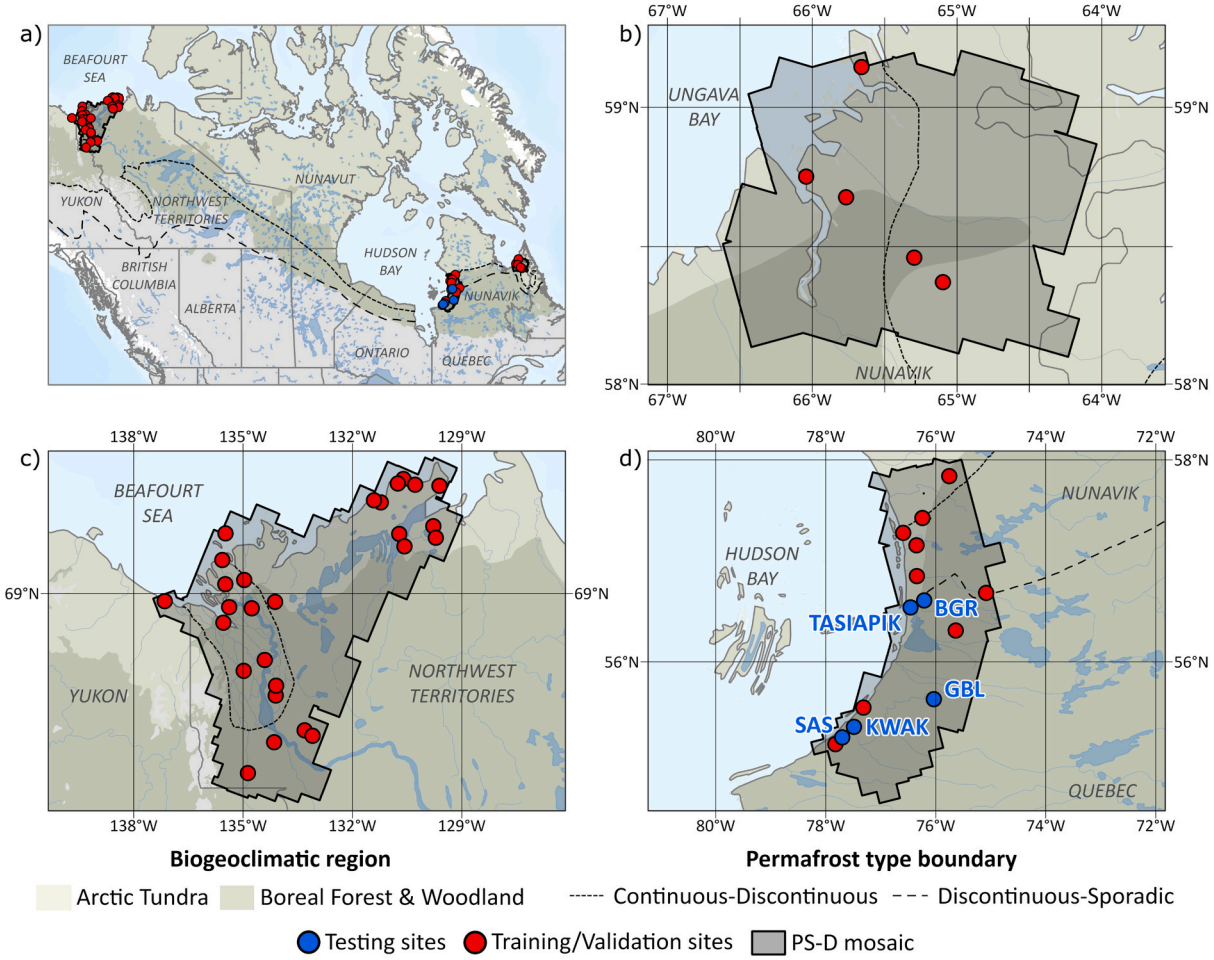

Fig. 1. Locations in Arctic and Subarctic Canada where the training/validation (red points) and testing (blue points) were performed. a) General geographical setting; b) Kangiqsualujjuaq, with lowland thermokarst close to the coast and large glacial basin lakes inland; c) Mackenzie Delta and Tuktoyaktuk Peninsula, with contrasting water bodies, those in the west more controlled by fluvial dynamics and in the east mainly related to permafrost and ice-wedge polygon degradation; d) Nunavik, Eastern Hudson Bay, where testing was performed, showing glacial basin lakes and numerous and diverse ponds. The black polygons are the full PS-D database extent (mosaic of 891 images - 118,603 km^2). The base information is from Government of Canada (CanVec, Permafrost Atlas and biogeoclimatic regions). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and Southern Ungava Bay – Kangiqsualujjuaq) (Fig. 1).

For testing the performance of the model, we selected diverse and geomorphologically complex sectors over a boreal-forest tundra transect in Nunavik, northern Quebec, that spanned the sporadic and discontinuous permafrost zones (Fig. 1). This region comprises a wide variety of water body types, environmental settings and permafrost at different states of degradation (Gorham, 1991; Rouse et al., 1997). The post-glacial evolution resulted in numerous permafrost thaw lakes and ponds, showing diverse optical properties, morphologies and hydrological connectivity (Watanabe et al., 2011; Wang et al., 2018; Freitas et al., 2019).

For a higher detail analysis and validation, we covered five areas at Very-High Resolution (VHR): i. low and high shrub dominated wetlands at Sheldrake River (BGR) and Kwakwatanikapistikw River (KWAK), respectively; ii. fen/bog palsa peatlands at Sasapimakwananisikw River (SAS); iii. Glacial valleys at Tasiapik; and iv. glacial basin lakes (GBL). BGR, KWAK, SAS and Tasiapik have also been surveyed at Ultra-High Resolution (UHR) with UAS and field work ground truthing since 2015 (Figs. 1 and 2).

The Nunavik transect at Eastern Hudson Bay lies on granite-gneiss rocks of the Precambrian Shield. The region shows one of the most rapid isostatic rebounds in the world with ≈ 13 mm·y^{-1} due to the retreat of the Laurentide Ice Sheet since ≈ 11.6 ka. The glacial retreat was followed by the marine transgression of the Tyrrell Sea at ≈ 7.9 ka (Bhiry et al., 2011). As a result, Quaternary deposits in the region are essentially glacial tills, marine clays and littoral sands that filled topographic depressions (Bouchard et al., 2014). The regression of the Tyrrell Sea led to permafrost aggradation, followed by degradation after the Little Ice Age (≈ 0.3 ka) that generated numerous thermokarst ponds. These are morphologically diverse, showing varying concentrations of organic and inorganic materials, solutes and pigments, which results in water with strikingly different colors (white, green, beige, brown, dark brown, black) across small distances (Watanabe et al., 2011; Freitas et al., 2019; Folhas et al., 2020) and with variable but potentially high greenhouse gas emissions (Breton et al., 2009; Laurion et al., 2010; Bouchard et al., 2011; Matveev et al., 2019). The region supports some of the most unique and pristine wetlands and peatlands in the world (Darnajoux et al., 2015). These are undergoing marked changes due to increasing air and soil temperatures over the last decades (Pelletier et al., 2019), which have led to thawing of ice-rich permafrost, intense terrestrialization and shrubification processes, and shifts in plant community structure (Payette et al., 2004; Ropars and Boudreau, 2012; Tremblay et al., 2012).

3. Materials and methods

3.1. General workflow

Current water body maps for the Arctic and Subarctic regions, particularly in Canada, are incomplete since they miss water bodies smaller than 0.01 km^2, as well as narrow river sectors, tributaries and streams (Muster et al., 2017; Harlan et al., 2023). To improve the existing classifications, we used deep learning over PS-D imagery. Training was done by manual water body delineation over the PS-D imagery, aided by the analysis of VHR satellite base maps, complemented by existing water body products (e.g., CanVec from Government of Canada, PeRL). Field knowledge concerning the study areas was key to produce an accurate training set.

An iterative retraining process considering the current model best state allowed to speed up deriving new training samples, better understanding model inaccuracies, as well as improving the algorithm learning ability. The training samples were used for deriving the image chips and labels that fed a supervised Mask R-CNN model. Using Residual Network (ResNet) for transfer learning, the best backbone model was selected according to the highest Average Precision (AP) score and by evaluating the training and validation losses per epoch. The performance and confidence of the algorithm were evaluated autonomously by comparing the results with manually delineated water bodies. This was done using two imagery resolutions: 1. VHR from PS-D imagery (≈ 3 m pixel size), and 2. UHR from UAS imagery (0.05–0.15 m pixel size).

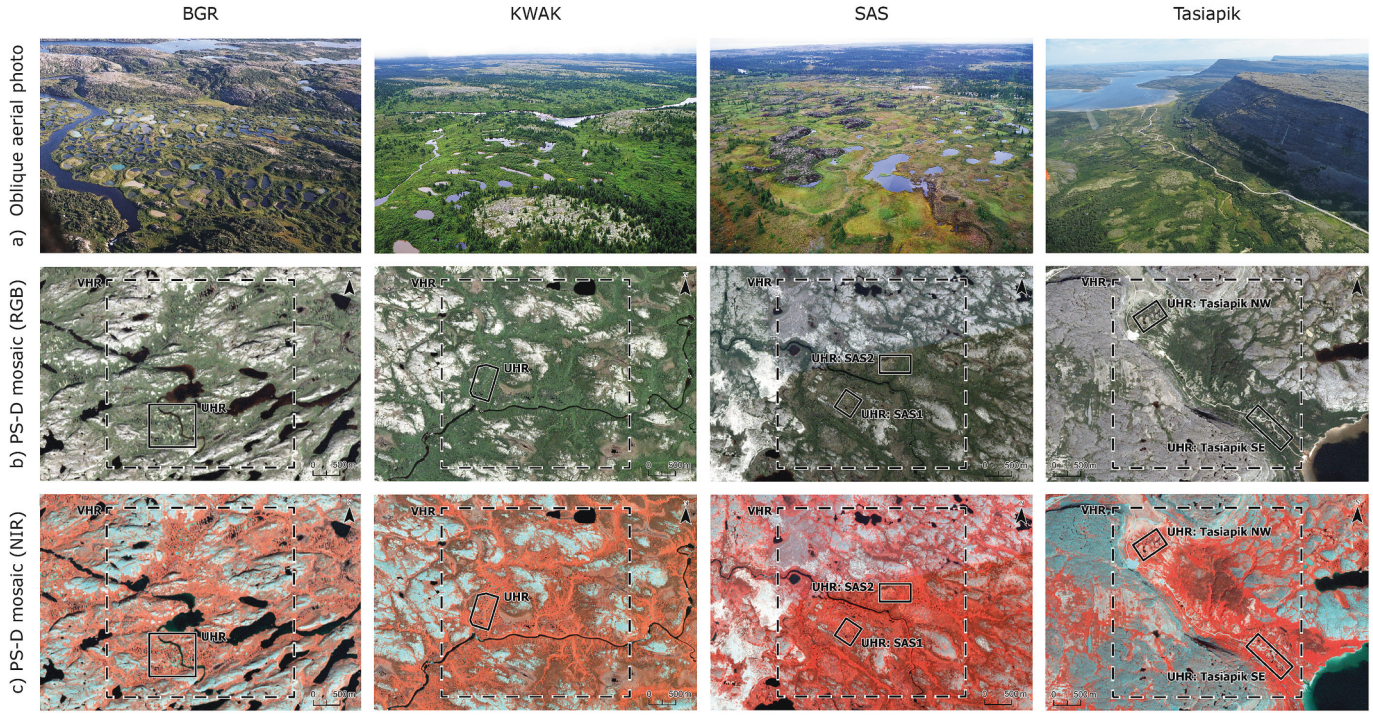

Fig. 2. Examples of the environmental settings of lakes and ponds in the boreal forest-tundra ecozone of Nunavik that were used for testing the model. Oblique aerial photos (a) and corresponding PS-D satellite images (b – true color: RGB; c – near infrared: NIR) from BGR, KWAK, SAS and Tasiapik with the delimitation of the VHR and UHR testing sites.

While the former evaluated the potential of PS-D imagery for mapping small water bodies, the latter allowed assessing the sub-pixel issues for water body delineation quality at the VHR. The model performance and confidence were evaluated according to different landscape types (Fig. 3).

3.2. PlanetScope Dove (PS-D) imagery

The PS-D constellation was formed by c. 130 sun-synchronous nanosatellites obtaining global daily 4-band multispectral imagery (blue, green, red and near-infrared) at ≈ 3 m spatial resolution and 12-bit radiometric depth, from July 2014 until April 2022 (Planet Labs, 2022). PS-D was replaced by the PlanetScope SuperDove constellation, operating since mid-March 2020, which now obtains 8-band multispectral imagery (coastal blue, blue, green I, green, yellow, red, red-edge and near-infrared) at comparable spatial and temporal resolution (Planet Labs, 2022). The low data requirements from our model along the maintenance of the blue, green, red and near-infrared bands, similar to those where our training was based, should make our model applicable to SuperDove imagery.

From a total of 891 PS-D Ortho Tile Products (Analytic MS – Surface Reflectance), radiometrically corrected, orthorectified and UTM projected (Planet Labs, 2022), 149 were used for deriving training samples of water bodies. The imagery was obtained through the Planet Labs' Education and Research Program, which allowed for free-of-charge downloads of a maximum of 10,000 km^2 per month. The study period was from July to September, but due to the dense cloud cover in some sectors, especially in eastern Hudson Bay, the acquisition of scenes for the same year was not possible. As a result, we used imagery from 2017 to 2019, which was combined in a mosaic dataset in ArcGIS Pro 2.5, without snow and cloud cover. For managing imagery overlap, we used the first sorted image, then derived mosaic radiometric seamlines and manually edited them, when necessary, to remove cloud or snow artifacts. The third order of the dodging algorithm was applied for color balance (Zhou, 2015).

3.3. Ground truthing

In remote sensing, using and reusing water body training samples is difficult due to the dynamic nature of these systems, which show water level oscillations and may shrink or expand across time (Pekel et al., 2016; Pickens et al., 2020). As a result, using water training samples without a precise sync between manual delineations and base imagery can generate bias in the classification model, resulting in inconsistent boundaries. To overcome this problem, we generated a water body dataset from scratch having as primary reference the PS-D mosaic, ensuring water body boundaries quality and consistency.

The first step for generating our dataset, was the visual inspection of the PS-D mosaic, defining sectors with distinct degrees of complexity, both in the color and morphometrics of the water bodies and also in terms of the surrounding spatial contexts involved. Selecting a complex, diverse and large dataset was crucial for improving the robustness and overall performance of the deep learning model (Qayyum et al., 2020; Adegun et al., 2023) (Fig. 4).

We considered all types of water occurring at or surrounded by land, from small ponds to large lakes, rivers, stream, creeks and also coastal sectors. Water body delineations were manual and model-assisted (e.g., derived from model retraining – section 3.4.2), a procedure conducted by a single operator, using a near-infrared false-color composite. This composite was very efficient, since water tends to fully absorb in the near-infrared part of the spectrum (Gao, 1996; Xu, 2006). The procedure was aided by systematic visual inspection of VHR satellite imagery (e.g.,

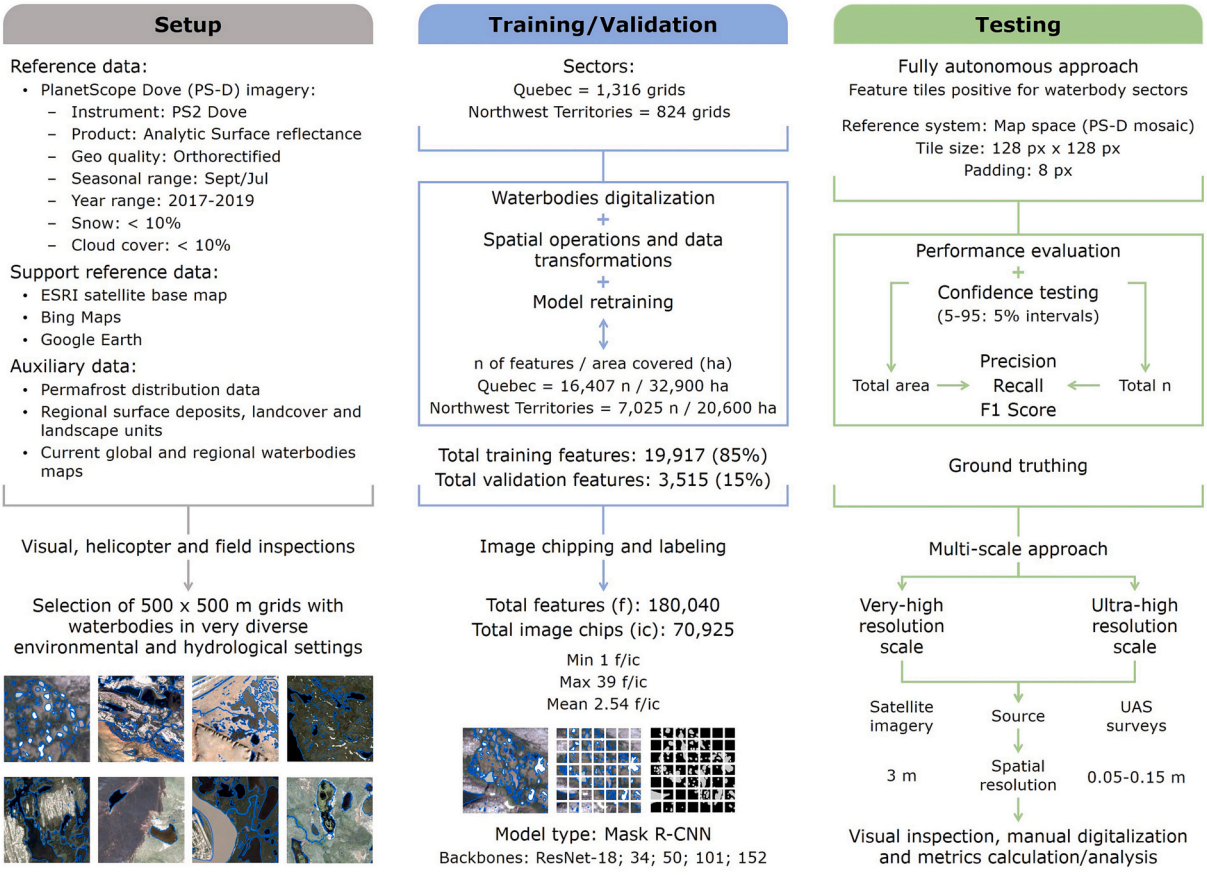

Fig. 3. Methodological workflow for the PS-D imagery model classification implementation, validation and testing procedures (HLWATER V1.0).

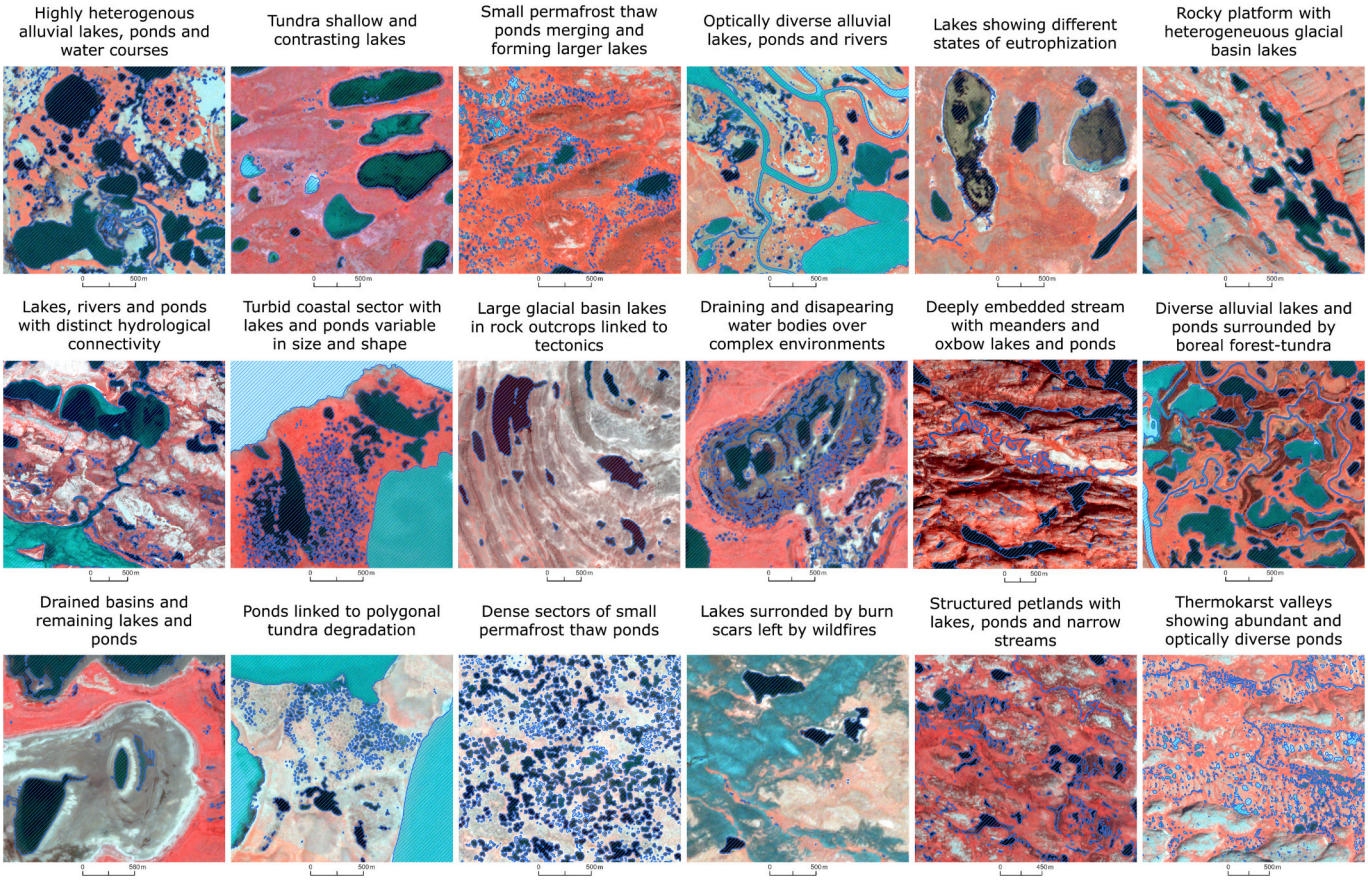

Fig. 4. Examples of the diverse water body types and environmental settings that were targeted for manual and model-assisted delineations. Delineations refer to the blue polygons. The PS-D mosaic is a false-color composite (R – near-infrared; B – blue; G – green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ESRI World Imagery, BingMaps, Google Earth) following Qayyum et al. (2020), Nitze et al. (2021) and Mullen et al. (2023). Only the water bodies that were clear on the PS-D mosaic were considered. This strategy excluded problems related to hydrology (e.g., drained basins and differences in water table and lake boundaries) and co-registration (Qayyum et al., 2020).

For training and validating the model a total of 41 sectors were selected, corresponding to 2,140 UTM projected grids of 500 × 500 m (0.25 km^2 each), which resulted in 535 km^2 that were visually inspected for water body delineations. As a result, we have manually delineated 23,432 water bodies as polygons, ranging from 1.5×10^{-5} to 12,605 km^2, with a mean area of 0.00736 km^2. From this set, 97% showed areas below 0.01 km^2. Table 1 synthesizes the metrics of the training dataset, which was randomly split in 85% (19,917 features) for training and 15% (3,515 features) for validating model performance.

The testing of model performance was conducted using two complementary resolutions: VHR PS-D imagery at ≈ 3 m pixel size, and UHR UAS orthomosaics at 0.05–0.15 m pixel size. The UHR areas were located within the VHR sectors. Each VHR sector was 12.25 km^2, resulting in a total of 61.25 km^2, and the UHR areas varied from 0.13 to 0.69 km^2 depending on the UAS surveys, resulting in a total of 1.7 km^2 (Table 2). A total of 3,805 features were used, 79% from VHR and 21% from UHR delineations. At VHR the water bodies, varied from 2.1×10^{-5} to 0.804 km^2 and at UHR from 7.5×10^{-7} to 0.0442 km^2 (Table 1).

3.4. Mask R-CNN model implementation framework

The Mask R-CNN model (https://github.com/matterport/Mask_RCNN) was implemented here using the deep learning framework of ArcGIS Pro 2.5 (https://github.com/Esri/deep-learning-frameworks/tree/master). In this framework, the Mask R-CNN model solution (https://github.com/Esri/raster-deep-learning/blob/master/python_raster_functions/Keras/MaskRCNN.py) uses TensorFlow, Keras and Scikit-Image Python libraries and following dependencies for training and validating the model during training, and subsequent object detection and instance segmentation (e.g., mask predictions) tasks. We adopted this solution since it allowed a straightforward implementation and a simple articulation between the input imagery and the model outputs, facilitating visual inspections and quality assessment throughout the process.

Table 1
Statistics of the water bodies dataset split by purpose (training/validation and testing) and resolution (VHR – ≈ 3 m pixel size; UHR – 0.05-0.15 m pixel size) with minimum (Min), maximum (Max), mean ($\bar{x}$), median (M_d) and standard deviation (σ) area in square kilometer (km^2).

Purpose	Resolution	Samples	Min (km^2)	Max (km^2)	$\bar{x}$ (km^2)	M_d (km^2)	σ (km^2)
Training/Validation	VHR	23,432	1.5×10^{-5}	12,605	0.00736	2.15×10^{-4}	0.147
Testing	VHR	3,012	2.1×10^{-5}	0.804	0.00182	1.77×10^{-4}	0.0243
Testing	UHR	793	7.5×10^{-7}	0.0442	2.84×10^{-4}	8.68×10^{-5}	0.00160

Table 2
Characteristics of the VHR and UHR imagery for testing the model performance, with comparisons of sites, dates, surveyed areas, type of imagery and respective pixel size (level of detail during ground truthing manual delineations).

Testing resolution	VHR	UHR
Sites and dates	GBL (11/09/2017); BGR (23/08/2019); KWAK (04/08/2018); SAS (02/08/2018); Tasiapik (22/08/2018)	BGR (02/09/2015, 31/08/2017, 26,27/08/2019); KWAK (28/08/2015, 01/09/2017, 14,15/08/2022); SAS1 (30/08/2015, 01/09/2017, 13,16/08/2022); SAS2 (30/08/2015, 01/09/2017); Tasiapik NW (01/09/2019); Tasiapik SE (30/08/2019)
Survey area	12.25 km^2 each site (61.25 km^2 hectares in total)	BGR = 0.69 km^2; KWAK = 0.23 km^2; SAS1 = 0.13 km^2; SAS2 = 0.19 km^2; Tasiapik NW = 0.18 km^2; Tasiapik SE = 0.28 km^2
Imagery	PlanetScope imagery aided by very- high resolution satellite base maps (e.g., ESRI World Imagery, Google Earth, Bing Maps and Google Satellite)	UAS orthomosaics (SenseFly eBee Classic and RTK Plus – S.O.D.A. and 4-band Multispectral Sequoia; Inspire 2–10-band Red Edge Dual Multispectral)
Pixel size	≈ 3 m	0.05–0.15 m

3.4.1. *Data exportation for training*

The total 23,432 water bodies were used to derive 70,925 image chips, for which we used the Export Training Data For Deep Learning Tool of ArcGIS Pro 2.5 (Maxwell et al., 2020). To enhance the capacity of the model to detect small water bodies without losing spatial context information, a tile size of 128 pixels in the X and Y axis, as well as stride size of 64 pixels in the X and Y axis were used. Within each image chip, segmentation masks for each instance of an object were obtained showing minimum, maximum and mean values of 1, 39 and 2.54 samples, respectively. After object differentiation, unique instances are needed for instance segmentation methods (Maxwell et al., 2020) (Fig. 5). This process sorted out 180,040 classified features within the range of each image chip.

3.4.2. *Training and retraining strategy*

The training considered the image chips and segmentation masks, with ResNet as the CNN architecture for transfer learning (Sarker, 2021; Alem and Kumar, 2022). We tested the shallow residual networks ResNet-18, ResNet-34 and the deeper versions ResNet-50, ResNet-101 and ResNet-152. These architectures differ in terms of the number and width of the layers, with higher networks generally showing higher ability to learn and transfer knowledge by better generalizing the results, but frequently increasing computational and training sample needs (He et al., 2016 and 2020; Sarker, 2021; Adegun et al., 2023). They were originally designed by He et al. (2016) and were set to retain their original weights and biases (e.g., model freezing) before training, allowing faster convergence (Alem and Kumar, 2022). This procedure ensured that the networks retained the knowledge acquired from pre-training on the ImageNet Dataset (Russakovsky et al., 2015), which in this case referred to a subset of around 1 million natural images labeled by humans, for enhanced image recognition purposes (He et al., 2016; Alzubaidi et al., 2021).

An iterative retraining process for deriving new ground truthing features, training, analyzing training and validation losses per epoch, as well as AP scores (e.g., common area under the curve considering the Precision averages across all Recall values between 0 and 1, defined by Padilla et al., 2020), then using the current best state model for detecting and validating water body feature tiles was conducted. When increasing from ResNet-18 to ResNet-50, the AP rose to 0.6 and the model started being used for deriving new training samples in different sectors. This allowed for increasing training (e.g., feature engineering), better understanding model failures, and refeeding it with new relevant information. During this process, the correct feature tiles were verified and merged, while the undetected water bodies were manually digitized, both refeeding the model.

The ResNet that showed the final highest AP was ResNet-50 with 0.75, achieved over 18 epochs with a batch size of 8 samples. The model demonstrated consistent learning (e.g., ability to match ground truthing and generalize to new instances), evidenced by decreasing training and validation losses per epoch at similar rates (Alem and Kumar, 2022). There was no overfitting, as indicated by the continuous decline in validation loss (Fig. 6, Alzubaidi et al., 2021). The lowest training loss was achieved at epoch 18, as a result the two small peaks at epochs 14 and 17 were likely attributed to the optimization process during those specific training iterations (Zhang et al., 2018; Alzubaidi et al., 2021; Alem and Kumar, 2022). An optimal learning rate of 6.31^{-6} was automatically extracted from the training curve for faster convergence. Additionally, we considered the early stop function, to save processing time and minimize overfitting (Zhang et al., 2018).

3.4.3. *Instance segmentation for mask predictions*

The water bodies were detected using a tile size of 120 × 120 pixels and padding size of 8 pixels. These parameters proved to be the best for detecting and preserving the boundaries of small features, as well as for decreasing processing time. The confidence threshold was set to 0 allowing for the detection of positive feature tiles for water bodies, even when the model had low confidence. This allowed for splitting the

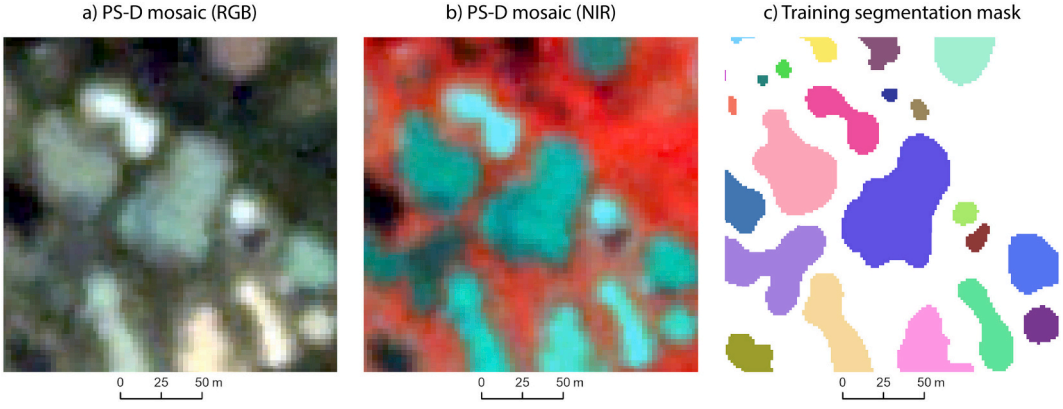

Fig. 5. Example of an image chip (a – true color composite; b – near-infrared composite) and correspondent segmentation masks (each color represents a different instance of an object) for training, referring to 23 mapped optically and morphologically diverse small water bodies in an area of 128 × 128 pixels (≈ 205 m × 205 m).

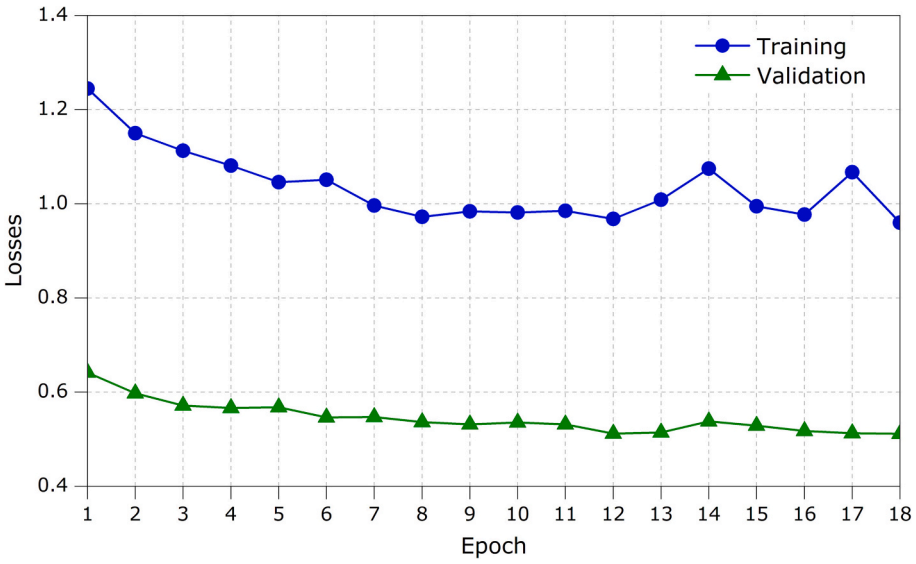

Fig. 6. Training and validation losses per epoch for the best backbone model (ResNet-50).

features into confidence classes and testing model confidence beyond general performance (Alzubaidi et al., 2021). Techniques for cleaning feature tiles, such as Non-Maximum Suppression (Hosang et al., 2017), which allows for the retrieval of overlapping feature tiles with the highest confidence scores based on a user-selected IoU, were not employed. This was because our goal was to understand the overall performance and confidence of the model without any intervention.

3.5. Model evaluation metrics

For performance evaluation, the IoU, Precision, Recall and F1 Score metrics were used (Liu et al., 2020; Maxwell et al., 2020; Alzubaidi et al., 2021). The IoU, which is an important measure of the model performance at a scale of each individual object (e.g., water body), corresponds to the ratio between the area of overlap of the detected and ground truthing polygons and their area of union (Liu et al., 2020; Padilla et al., 2020). The higher the IoU, the stronger the agreement, a value of 1 indicating a perfect automatic delineation. The IoU was also used as a reference for Precision, Recall and F1 Score metrics considering a threshold at 0.5 (Huang et al., 2020). This was used for assessing the True Positives (TP), False Positives (FP) and False Negatives (FN). Detections with IoU $\geq$ 0.5 were classified as TP and the ones with IoU $<$ 0.5 as FN (Padilla et al., 2020). An IoU of 0.5 is not only the unique mid-term condition between detecting water bodies and correctly delineating their boundaries, but it is also a commonly chosen threshold in major deep learning challenges, like PASCAL, ILSCRV and MS COCO (Liu et al., 2020). In addition, we also calculated the metrics using the total sum area of the polygons and intersections classified as TP, FP and FN. This allowed assessing the algorithm ability to detect water bodies as a function of their area, similarly to a pixel-based approach (Zhang et al., 2018; Qayyum et al., 2020; Philipp et al., 2022).

While Precision records the ability of the model to identify only relevant objects (e.g., percentage of the correct positive predictions), Recall records the ability of the model to find all relevant cases (e.g., percentage of the correct positive predictions considering all ground truthing features) (Padilla et al., 2020). The F1 Score offers an overall accuracy of the model, combining Precision and Recall, through their harmonic mean calculation (Maxwell et al., 2020; Alzubaidi et al., 2021). For these metrics, a value of 0 refers to the worst and 1 to the best possible model performances.

To assess the confidence of the algorithm for detecting feature tiles properly in different landscape units, these were analyzed for the VHR sites, considering 5% confidence interval classes (5 to 95%), having as reference the same quality evaluation metrics. The best confidence thresholds (IoU 0.5 F1 Score 1st highest peak) for each VHR were selected for a detailed sub-pixel assessment of the performance of the algorithm in the respective UHR sites.

To allow a better comparison between the detected and observed water body areas, we used Kernel Density Estimation (KDE) plots. KDE is a non-parametric method that provides a smooth curve representation of data distributions, by placing a kernel on each data point and summing the contributions of all kernels (Wand and Jones, 1994). These plots were generated using OriginPro 2023, with the bandwidth determined by the Scott method. Unlike histograms or other class-based data visualizations, KDE plots allow for continuous comparisons without discrete class gaps.

4. Results

4.1. Evaluation of model performance at Very-High Resolution (VHR)

The evaluation of the model performance considering all confidence thresholds at the VHR validation sites showed very good overall results (Table 3). Regarding the Precision, Recall and F1 Score based on total area, means and medians were higher than 0.78, except at Tasiapik with 0.58 to 0.71. This indicated that most of the water body area detected were TP, the mean being higher than 70%, except for Tasiapik with 45%. At all VHR validation sites, except Tasiapik, the FP mean area was lower than 20% and FN mean area lower than 17%.

The metrics based on total n at IoU 0.5 showed different trends (Table 3). The mean IoU 0.5 Precision was higher than 0.76 at all sites, except in KWAK, with 0.58. Considering the IoU 0.5 Recalls, GBL gave the highest mean (0.67). In BGR, KWAK and SAS the mean decreased to $\approx$ 0.6, being slightly better in BGR (0.63) and SAS (0.6) comparing to KWAK (0.55) and down to 0.5 in Tasiapik. As a result, the IoU 0.5 F1 Score means were 0.71 for GBL and BGR. For SAS, Tasiapik and KWAK these values were 0.63, 0.57 and 0.53, respectively. The mean TP n at IoU 0.5 according to all confidence thresholds was 56% for GBL, following BGR with 55%, SAS with 47%, Tasiapik with 40% and KWAK with 36% (Table 3).

Analyzing the IoU values of the detected water bodies (IoU $>$ 0), the highest mean value was in GBL with 0.7, following SAS with 0.66 and then BGR, Tasiapik and KWAK with $\approx$ 0.6 (Table 3). These indicated a moderate to high efficiency of the algorithm to delineate the ground truthing features. Comparatively, the IoU of the detected and undetected water bodies were lower due to the existence of FP and especially FN

Table 3
General performance of the model for the different VHR sites. The values are the Precision, Recall, F1 Score, TP (%), FN (%), FP (%) and IoU mean ($\bar{x}$), median (M_d) and standard deviation (σ) according to all confidence thresholds (5–95: 5% intervals) based on total area and total n at IoU 0.5. In IoU * refers to the detected water bodies (IoU > 0 as TP) and ** includes the undetected ones (IoU = 0 as FN and FP).

VHR sites		Total area						Total n (IoU 0.5)						IoU * (**)
		Precision	Recall	F1 Score	TP (%)	FN (%)	FP (%)	Precision	Recall	F1 Score	TP (%)	FN (%)	FP (%)	
GBL	$\bar{x}$	0.94	0.97	0.95	91	3	6	0.78	0.67	0.71	56	27	17	0.70 * (0.48 **)
	M_d	0.94	0.97	0.95	91	3	6	0.89	0.69	0.73	58	27	8	0.69 * (0.51 **)
	σ	0.01	0.01	0.003	1	0	1	0.21	0.05	0.11	12	8	18	0.07 * (0.10 **)
BGR	$\bar{x}$	0.86	0.88	0.87	77	10	12	0.85	0.63	0.71	55	33	11	0.59 * (0.44 **)
	M_d	0.88	0.89	0.88	78	10	11	0.88	0.66	0.73	58	31	8	0.58 * (0.46 **)
	σ	0.04	0.02	0.02	3	2	4	0.13	0.08	0.05	6	11	11	0.03 * (0.05 **)
KWAK	$\bar{x}$	0.78	0.88	0.82	70	10	20	0.58	0.55	0.53	36	31	33	0.58 * (0.28 **)
	M_d	0.80	0.89	0.84	73	9	18	0.54	0.60	0.55	38	26	34	0.59 * (0.28 **)
	σ	0.09	0.03	0.05	6	4	9	0.24	0.09	0.10	9	16	21	0.04 * (0.05 **)
SAS	$\bar{x}$	0.87	0.81	0.84	72	17	11	0.76	0.60	0.63	47	35	18	0.66 * (0.35 **)
	M_d	0.89	0.83	0.85	74	16	9	0.81	0.66	0.67	50	29	14	0.67 * (0.38 **)
	σ	0.08	0.05	0.03	4	5	8	0.20	0.15	0.11	10	19	18	0.04 * (0.07 **)
Tasiapik	$\bar{x}$	0.58	0.69	0.62	45	21	34	0.76	0.50	0.57	40	43	17	0.60 * (0.31 **)
	M_d	0.60	0.71	0.63	46	19	32	0.82	0.55	0.61	44	40	11	0.59 * (0.33 **)
	σ	0.11	0.07	0.05	5	9	12	0.19	0.14	0.10	9	18	16	0.03 * (0.07 **)

(IoU = 0). This was clearer for KWAK, Tasiapik and SAS, where these values decreased to 0.28, 0.31 and 0.35, respectively (Table 3).

Precision, Recall and F1 Score at 5% confidence classes were always better for total area than for total n, except for Tasiapik, that showed generally worst performances (Fig. 7). Focusing on the IoU 0.5 Precision, the results showed a sharp increase as the confidence approached 95%, reaching 1 at all VHR sites. This was clearer for GBL, BGR and SAS, where IoU 0.5 Precision above 0.9 were reached even at 50–55% confidence. Precisions based on the total area revealed the same behavior, although with lower values, due to remaining areas classified as FP. Recall showed an opposite behavior of Precision, both for total n (IoU 0.5) and area. This revealed that the algorithm better detected water bodies with increasing confidence threshold, although detecting a smaller number of individuals. This caused decreasing F1 Score values with increasing confidence thresholds (Fig. 7).

The best confidence thresholds according to the 1st highest IoU 0.5 F1 Score peak were 50% for BGR, SAS and Tasiapik, 60% for GBL and 65% for KWAK. At these thresholds the IoU 0.5 F1 Scores were above 0.73 for GBL, BGR and SAS, decreasing to 0.66 and 0.63 for Tasiapik and KWAK, respectively (Table 4). The IoU 0.5 Precisions were above 0.81 for all sites, except for KWAK with 0.69. The lowest IoU 0.5 Recall was 0.55 for Tasiapik, following KWAK with 0.58 and the other sites showing

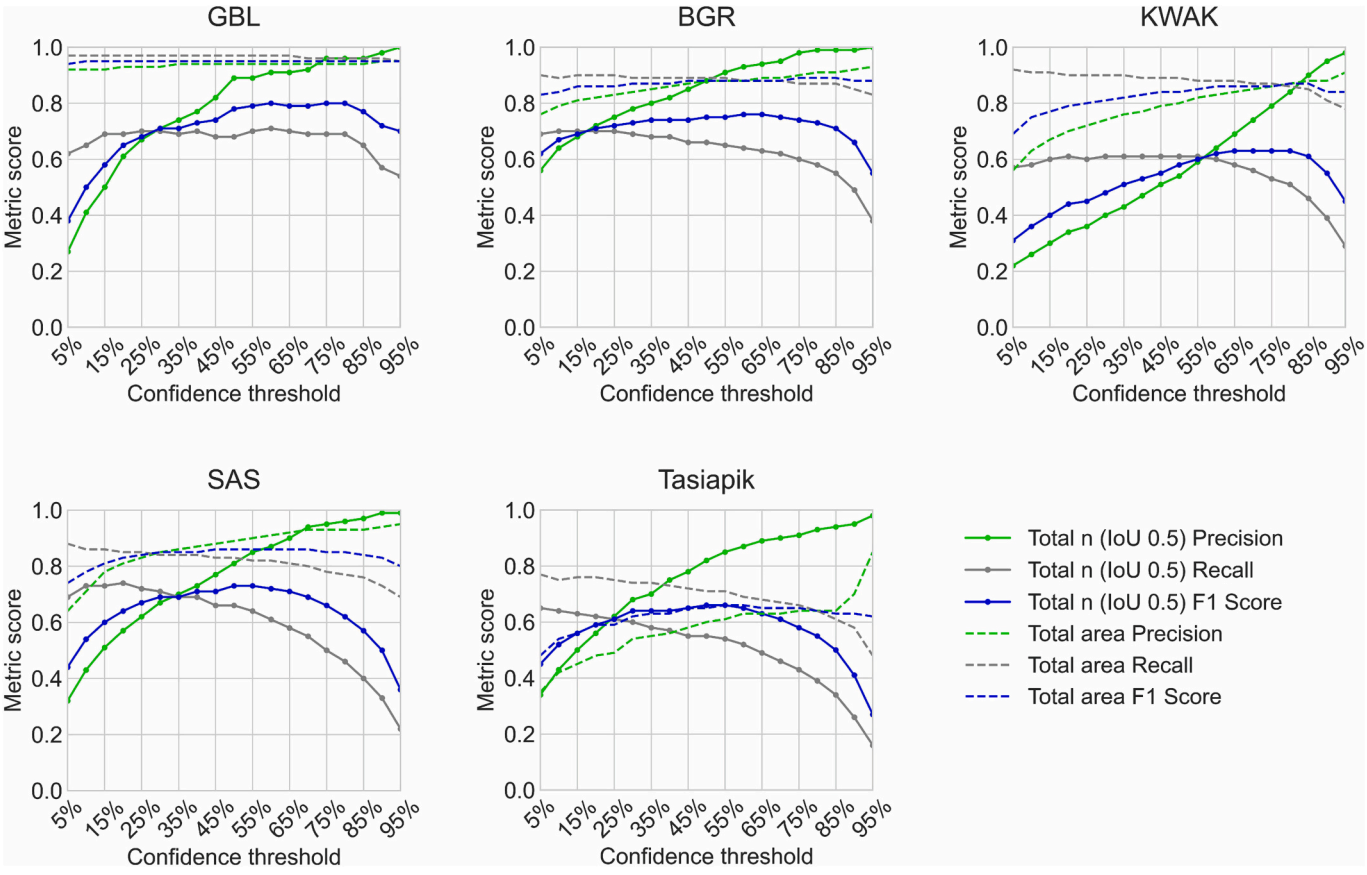

Fig. 7. Precision, Recall and F1 Score considering the total area and total n (IoU 0.5) according to 5% interval confidence classes for the VHR sites.

Table 4
Performance evaluation metrics at the best confidence threshold (50% for BGR, SAS and Tasiapik, 60% for GBL and 65% for KWAK) referring to the 1st highest IoU 0.5 F1 Score peak and TP n (%), FN n (%) and FP n (%) for the VHR sites.

VHR sites	Precision	Recall	F1 Score	TP n (%)	FN n (%)	FP n (%)
GBL	0.91	0.71	0.80	50 n (67%)	20 n (27%)	5 n (7%)
BGR	0.88	0.66	0.75	781 n (61%)	404 n (31%)	104 n (8%)
KWAK	0.69	0.58	0.63	411 n (46%)	302 n (34%)	182 n (20%)
SAS	0.81	0.66	0.73	279 n (57%)	144 n (29%)	67 n (14%)
Tasiapik	0.82	0.55	0.66	374 n (49%)	305 n (40%)	84 n (11%)

over 0.66.

At IoU 0.5, most of the detections were classified as TP, following FN and FP (Table 4). BGR, KWAK, SAS and Tasiapik revealed significantly different water bodies compared to GBL, with an increase in number and density of small lakes and ponds and, as result, more TP n and FN n (Table 4 and Fig. 9). The mean IoU of the detected water bodies (IoU > 0) for the best confidence thresholds varied from 0.58 in BGR to 0.73 in GBL. The median values were generally higher varying from 0.62 in BGR and KWAK to 0.83 in GBL (Table 5).

The IoU results, although good, revealed limitations in fully addressing the complexity of small water bodies (Table 5). For example, BGR and KWAK present numerous small water bodies, with the latter showing 707 water bodies and a median size of 134 m^2, and BGR showing 1153 water bodies and a median size of 256 m^2. Furthermore, these show distinct characteristics in shape and color, even across small distances (Fig. 9). Nevertheless, the large majority of the detected water bodies in all VHR sites had IoU $\geq$ 0.5 (> 72% n) (Fig. 8). For example, in BGR 72% of the detected water bodies had IoU $\geq$ 0.5, with 78% for Tasiapik, 80% for KWAK, 83% for GBL and 94% for SAS. Considering an IoU $\geq$ 0.7, the values were 28% for BGR and KWAK, 33% for Tasiapik, 72% for GBL and 75% for SAS.

The TP, FN and FP maps for the best IoU 0.5 F1 Score presented consistent automatic delineations even in dense and morphologically complex sectors with numerous ponds (Fig. 9). The main problems were relief cast shadows causing FP, especially in GL and BGR. In KWAK, FP were mainly related to very small shadows projected by trees due to its boreal forest setting. In SAS, some peaty waterlogged areas caused FP. Tasiapik revealed the worst FP results, with a road sector being misclassified as a narrow turbid stream and outcrops being a source of errors caused by high brightness (Fig. 9).

4.2. Evaluation of model performance using UAS data at Ultra-High Resolution (UHR)

Comparisons between the detected and observed water body areas in the UHR testing sites, showed that the detected means, medians and 1st quartiles were always higher than the observed (Fig. 10). This was because the UHR validation sites had a high density and frequency of very small ponds, which were not detected by the model. The minimum difference between interquartile ranges (Q3-Q1) of the observed and detected pond areas was 3 m^2 for KWAK (observed = 212 m^2; detected = 215 m^2), with particularly good matches between at that site and Tasiapik SE (Fig. 10).

Kernel Density Estimations (KDE) between the observed and detected pond areas revealed consistency (Fig. 11), especially at KWAK and Tasiapik SE. At BGR, SAS1 and SAS2, the strongest KDE disagreements occurred in ponds smaller than 150 m^2, with the observed densities and frequencies being higher. For instance, in BGR, the largest KDE difference occurred between 0.76 and 206 m^2 pond sizes. In that specific class, 276 ponds were observed, corresponding to a total area of 15,548 m^2. Accordingly, only 92 ponds were detected, summing up to a total area of 12,069 m^2. The same trend was apparent for SAS1 and SAS2, with challenges for detecting the smallest size ponds. The only strong KDE disagreement and mismatch occurred for Tasiapik NW.

The mean IoU for the detected water bodies was around 0.5, the highest being in KWAK (0.59), followed by SAS2 (0.58) and Tasiapik SE (0.54) (Table 6). These results are good since the UAS orthomosaics resolve water bodies at centimeter resolution, while at PS-D resolution they are generalized and unresolved in many situations. For both detected and undetected water bodies, the mean IoU decreased, varying from 0.23 in SAS2 to 0.41 in KWAK. These were affected by the large proportion of small ponds classified as FN (IoU = 0) that our algorithm was unable to detect (Table 6).

As an example, in BGR, the IoU means for all water bodies (including IoU = 0) having as reference the three main area classes classified using natural breaks, were 0.21 (M_d = 0 and σ = 0.28) for ponds ranging between 0.76 and 291 m^2 (345 n = 73%; 35,786 m^2 = 28%), increasing for 0.55 (M_d = 0.62 and σ = 0.21) for those ranging between 291 and 819 m^2 (97 n = 21%; 51,715 m^2 = 41%) and reaching 0.6 (M_d = 0.74 and σ = 0.26) for those ranging between 819 and 2,523 m^2 (30 n = 6%; 39,454 m^2 = 31%). The same occurred for SAS1, with the IoU means for all water bodies increasing from 0.15 for those showing sizes below 138 m^2 to an excellent 0.73 for those larger than 747 m^2. In fact, this trend of increasing IoU values with increasing water body sizes was common to all sites.

The IoU maps show the robustness of our new model for detecting morphologically and optically diverse small ponds, as well as streams (Fig. 12). The results are particularly accurate for BGR, KWAK, SAS1 and Tasiapik SE.

At BGR, despite many ponds of very small size (< 200 m^2), most of the larger ones were detected, the majority showing IoU $\geq$ 0.6. Only two large ponds (436 and 539 m^2) that were greenish were not detected. At the center of the UHR study area of BGR, some segmentation issues occurred in ponds showing sizes between 971 and 1,476 m^2, causing the detected frequencies and areas sum to decrease (17 observed ponds = 19,897 m^2; 7 detected ponds = 8,682 m^2) as well as IoU values (Figs. 11 and 12). These were caused by narrow water corridors connecting the ponds, which are difficult to detect at PS-D resolution. The Sheldrake River sector of BGR showed an observed water body area of 44,241 m^2 and a detected area of 40,861 m^2, 83% of which classified as TP, 12% as FN and 6% as FP, resulting in a Precision of 0.94, Recall of 0.88, F1 Score of 0.91 and IoU of 0.82 (Fig. 12).

For KWAK, the results were excellent with the detection of most ponds (77%) with high quality and individual IoU values $\geq$0.6. The undetected ponds were close to forested areas or were very small (< 150 m^2). Clustered trees caused FP linked to cast shadows. The creek in KWAK narrows from south (5 m) to north (2 m) and resulted in TP (49%) in the south and FN (21%) in the north, with Precision, Recall, F1 Score and IoU values of 0.61, 0.70, 0.65 and 0.44, respectively (Fig. 12).

Good results were also achieved for SAS1 and SAS2, with several of the very small ponds detected and with good IoU. However, detecting water bodies in these sites was challenging, given the oscillation in water table. For example, at SAS1 small ponds associated with seasonal water corridors around palsas were difficult to track. In this area, palsas

Table 5
IoU means ($\bar{x}$), medians (M_d) and standard deviations (σ) for the detected water bodies, and for detected plus undetected water bodies (*), at the VHR sites according to their best confidence threshold.

IoU/VHR sites		GBL	BGR	KWAK	SAS	Tasiapik
IoU (*)	$\bar{x}$	0.73 (0.58 *)	0.58 (0.49 *)	0.60 (0.34 *)	0.67 (0.42 *)	0.59 (0.37 *)
	M_d	0.83 (0.75 *)	0.62 (0.58 *)	0.62 (0.43 *)	0.70 (0.58 *)	0.64 (0.48 *)
	σ	0.26 (0.37 *)	0.19 (0.27 *)	0.17 (0.32 *)	0.15 (0.35 *)	0.20 (0.33 *)

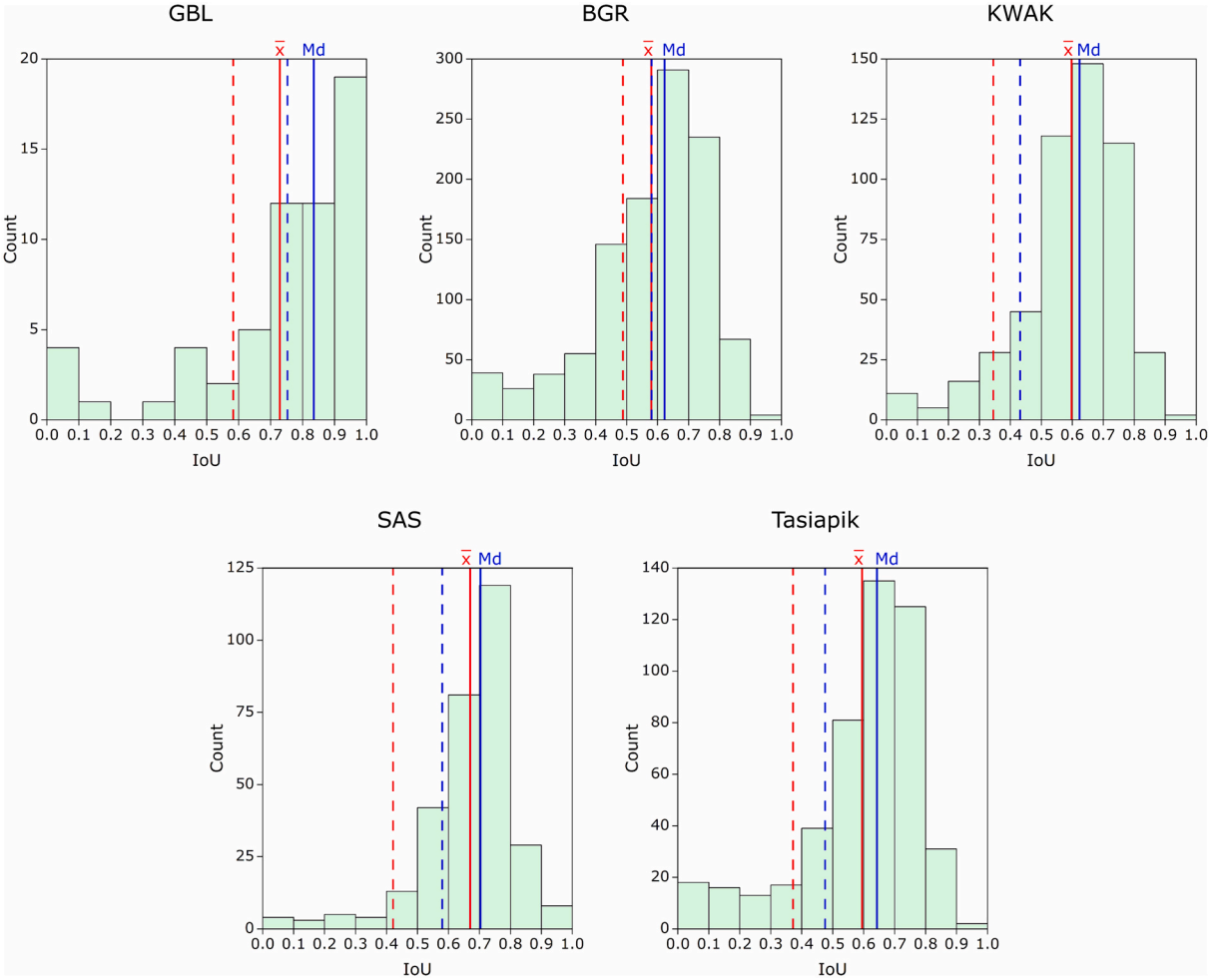

Fig. 8. Histogram plots of the IoU values of the detected water bodies (n) for the VHR sites according to their best confidence thresholds. The red vertical lines represent the means ($\bar{x}$) and the blue vertical lines represent the medians (M_d), solid for the detected water bodies, and dashed plus the undetected ones. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

were also responsible for FP, causing pond size overestimations associated with the low reflectance of organic-rich soils and shadowing (Fig. 12). In SAS2, FN occurred at waterlogged fens (12 undetected ponds with a mean size of 41 m^2), as well as at sectors close to degrading permafrost palsas (4 undetected ponds with a mean size 52 m^2). These were small and, as result, difficult to detect at PS-D resolution (Fig. 12).

The model was least accurate for the northwest sector of the Tasiapik valley (Tasiapik NW). Only the larger and well-defined ponds were accurately detected, especially towards the west with only 33% of detections (Fig. 12). This was because the site had mainly shallow and clear water ponds, with high bottom reflectance, and ponds with irregular shape frequently linked by narrow corridors. The results improved significantly in Tasiapik SE with 67% of the ponds detected. Apart from small undetected ponds (< 150 m^2), the majority showed good IoU values (mean of 0.46) and only three shallow ponds with 249 to 528 m^2 were not detected. This appeared to be due to bottom interference or even dry basins, conditions that were difficult to evaluate at PS-D resolution (Fig. 12).

The IoU 0.5 Precision was excellent for all sites, with values above 0.87 (Table 7), and only a few FP n. The IoU 0.5 Recall varied from 0.18 at Tasiapik NW to 0.54 at KWAK as a result of the small water bodies that were only detected using UHR imagery, causing FN n at all sites. At Tasiapik NW, FN were also caused by shallow depths and bottom reflectance, resulting in the highest FN frequencies (81% based on IoU 0.5 and 40% based on total area). The IoU 0.5 F1 Scores varied between 0.3 (Tasiapik NW) and 0.68 (KWAK). The F1 Scores based on total area were higher, varying from 0.64 in SAS2 to 0.80 in BGR (Table 7).

5. Discussion

5.1. Model confidence trade-off

The trained model provided different quality and confidence performances depending on the landscape type. In general, as the confidence of the water detection tiles increased, the model became more conservative, removing FP, but also detecting less TP. At the confidence threshold of 95% the IoU 0.5 Recall was 0.54 for GBL, 0.39 for BGR, 0.29 for KWAK, 0.22 for SAS and 0.16 for Tasiapik. The same metric based on total TP, FP and FN areas was 0.95 for GBL, 0.83 for BGR, 0.78 for KWAK, 0.69 for SAS and 0.48 for Tasiapik. These results highlight the challenges for detecting the numerous optically and morphologically diverse small water bodies at PS-D resolution, but also show that the large majority of the total water body area is still correctly detected, even under the most conservative of scenarios.

In the current research we adopted the water body tile detection confidence thresholds for the different VHR sites according to the 1st highest IoU 0.5 F1 Score peak. This model trade-off allowed better understanding the main factors controlling FN and FP, as well as maximizing TP (Zhang et al., 2018). Understanding the best confidence threshold during model detections is a good practice and depends on the purposes of the research and the characteristics of the study area. However, following this principle may require using complementary

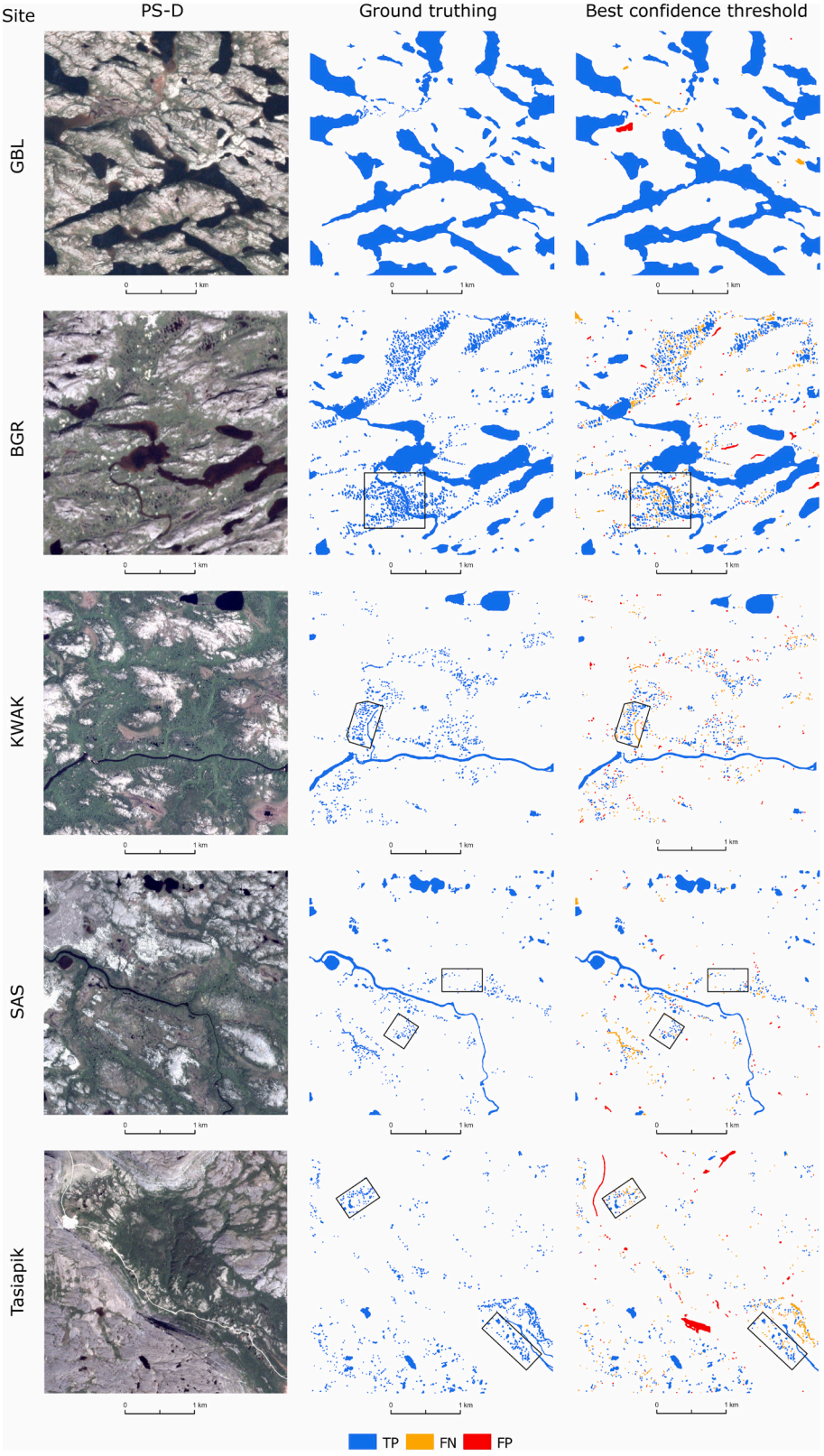

Fig. 9. PS-D, water bodies ground truthing and results of the model considering fully autonomous water bodies delineations according to the best confidence threshold (65% for KWAK valleys, 60% for GBL and 50% for BGR, SAS and Tasiapik valleys). The TP, FN and FP polygons and intersections were classified according to the IoU 0.5. The black bounding boxes represent the UHR survey sites.

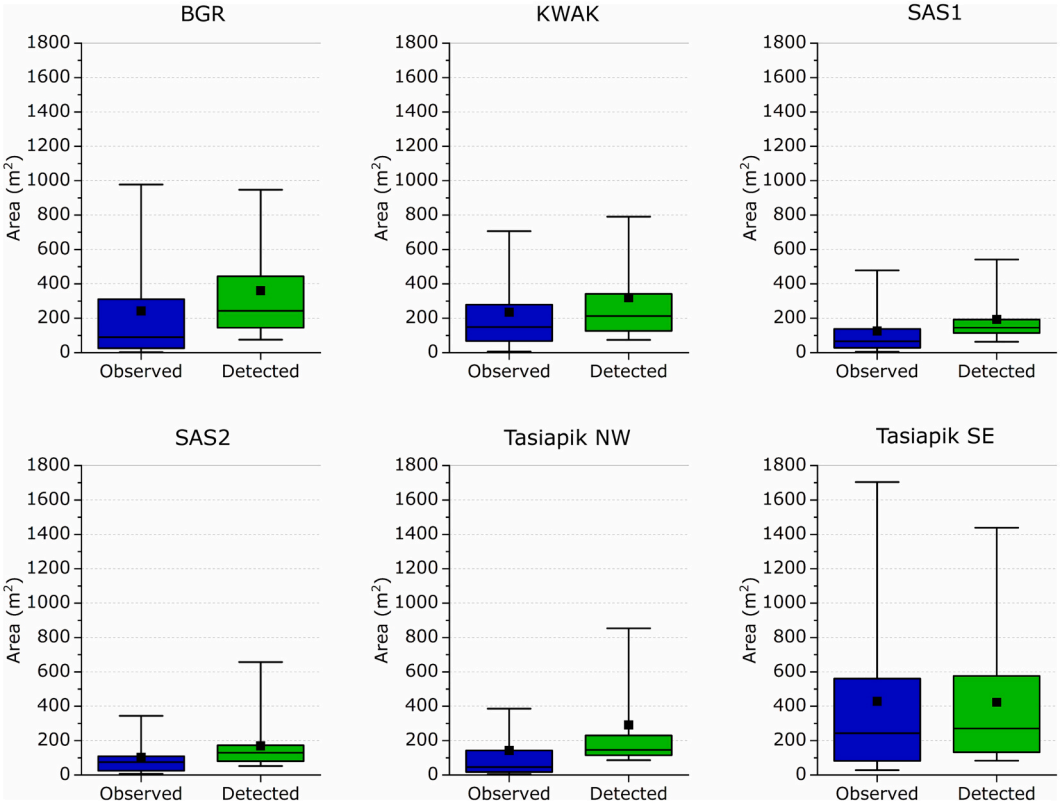

Fig. 10. Box plots for the observed and detected ponds in all UHR sites. The river and creek sectors were excluded (main distribution outliers). The minimum and maximum values are presented here at 5% and 95% percentiles, respectively.

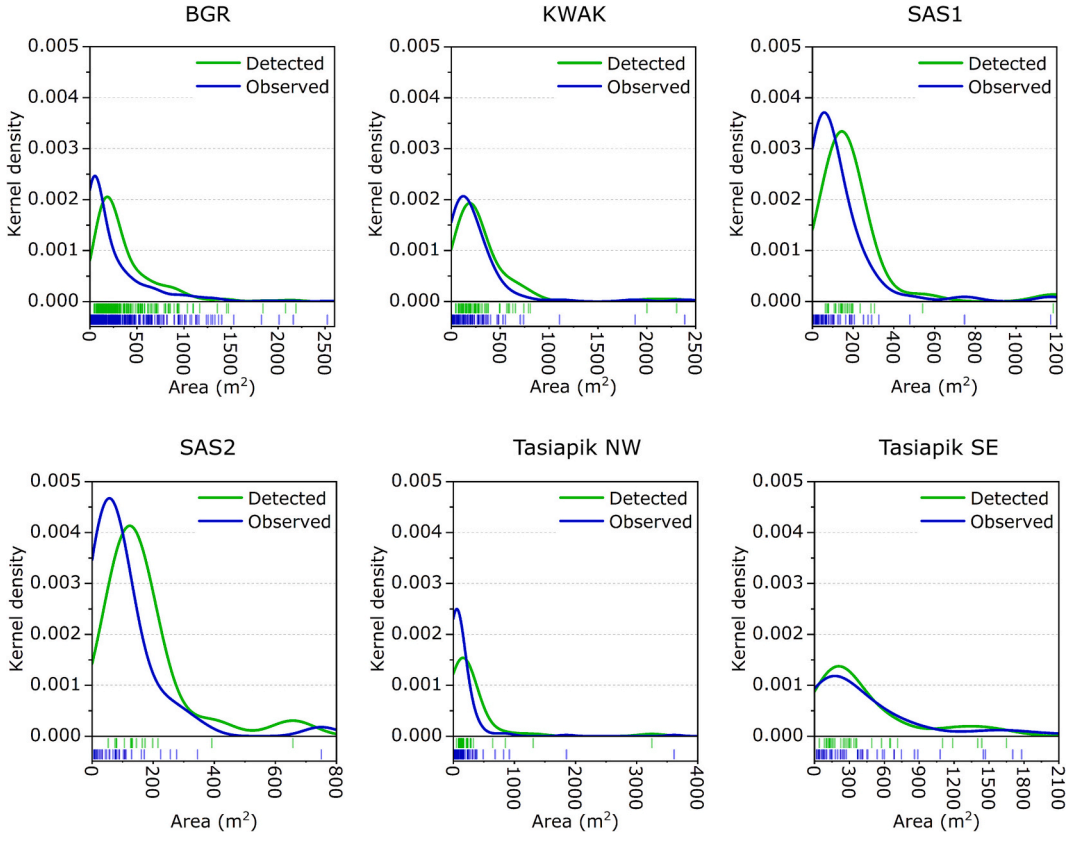

Fig. 11. Kernel Density Estimation (KDE) plots for the observed and detected ponds in all UHR validation sites. The river and creek sectors are excluded (main distribution outliers).

Table 6
IoU means ($\bar{x}$), medians (M_d) and standard deviations (σ) for the detected water bodies, and for detected plus undetected water bodies (*), at the UHR sites.

IoU/UHR sites		BGR	KWAK	SAS1	SAS2	Tasiapik NW	Tasiapik SE
IoU (*)	$\bar{x}$	0.56 (0.29 *)	0.59 (0.41 *)	0.48 (0.26 *)	0.58 (0.23 *)	0.48 (0.15 *)	0.54 (0.31 *)
	M_d	0.61 (0.08 *)	0.64 (0.51 *)	0.53 (0.20 *)	0.6 (0 *)	0.54 (0 *)	0.62 (0.18 *)
	σ	0.19 (0.31 *)	0.19 (0.32 *)	0.21 (0.29 *)	0.14 (0.3 *)	0.21 (0.25 *)	0.22 (0.32 *)

techniques or methods, such as testing remote sensing adapted pre-trained models (Yuan et al., 2020; Alem and Kumar, 2022), feeding the model with more data dimensions (e.g., digital surface/terrain models, space-time PS-D or other passive and active satellite platforms data cubes) (Sarker, 2021), providing more training samples, enhancing model robustness with data augmentation techniques (e.g., flipping, color space, rotation, cropping, translation or noise injection) (Huang et al., 2020), using feature tile cleaning techniques (e.g., Non-Maximum Suppression), adding TP feature tiles with less confidence (Zhang et al., 2018), finding techniques for removing FP (e.g., decision trees, morphometric analysis, shadow modeling) or manual inspection (Huang et al., 2020; Mullen et al., 2023). These post-detection procedures would likely improve the final quality of the water body dataset, although some of them are computationally costly and involve intense parametrization, or are laborious and highly time-consuming (Alzubaidi et al., 2021; Sarker, 2021; Adegun et al., 2023), especially when considering the diversity and abundance of water bodies in the Arctic and Subarctic (Muster et al., 2019).

Our results and the high quality performance evaluation metrics based on total area showed that FN were not problematic, and were mostly related to issues in detecting narrow water connections among ponds or in fully resolving morphological complexity of water bodies at PS-D native spatial resolution. Furthermore, most FP were linked to relief cast shadows that are easy to mitigate in a second classification step by using digital elevation models (Feyisa et al., 2014; Fisher et al., 2016; Freitas et al., 2022). While clusters of trees have also caused FP, most of these were small in terms of individual sizes and can be excluded by introducing a minimum waterbody size threshold, as described below.

5.2. Minimum pond size threshold

Our test water bodies datasets showed median sizes of 177 m^2 at VHR and 87 m^2 at UHR. This implied that 50% of the performance evaluation metrics were influenced by the ability of the model to automatically delineate such small and numerous ponds (VHR = 1,506 water bodies <177 m^2; UHR = 397 water bodies <87 m^2). This impacted the model performance metrics in general. For instance, at UHR this was clearer at BGR and subtler at Tasiapik SE, but was the main factor decreasing the quality of the metrics at all UHR sites, due to the large proportion of FN (e.g., very small water bodies) comparing to TP (e.g., impacting Recall).

The observed ponds at the UHR sites revealed a Pareto-like distribution, with substantial declining frequencies from the smallest to the largest pond areas (Skewness = 1.8, Kurtosis = 3.1). The detected ponds distribution was not able to fully represent that behavior, especially considering the smallest area classes (Skewness = 1.5, Kurtosis = 1.7) (Fig. 13). As a result, the observed counts and relative frequencies were much higher in the <50 m^2 (37%) and 50–100 m^2 (16%) classes, compared to the detected ponds (< 50 m^2 = 1%; 50–100 m^2 = 13%). However, both observed and detected histogram frequencies and KDE curves were similar for ponds above 100–150 m^2 (78 observations and 83 detections). Accordingly, the data suggests that a KDE peak of 166 m^2 may be used as the minimum size detection threshold of our model using PS-D imagery (Fig. 13).

The definition of the above-mentioned minimum size detection threshold led us to recalculate the performance quality metrics at all UHR sites (Table 8). The most significant improvement was in Recall, showing that the model became highly efficient for delineating and detecting ponds above that threshold. This was particularly evident at BGR, where the IoU 0.5 Recall increased from 0.38 to 0.74, at KWAK, where it increased from 0.54 to 0.79 and at SAS2, from 0.34 to 0.83. Better Precisions were also noticed, especially at IoU 0.5, resulting in major TP reduction, except in BGR, although still showing 0.98. In general, the IoU 0.5 F1 Score increased from 0.3-0.68 to 0.7-0.91 and from 0.64-0.8 to 0.76-0.83 considering the metrics based on the total area (Table 8). Watercourses, in particular narrow stream and creek detections for the UHR sites demonstrated that the model was efficient for water features wider than 4–5 m, while missing detections below that width.

Our proposed minimum size threshold is similar to the one of 100 m^2 demonstrated by Mullen et al. (2023). In that case, the authors used GNSS-based ground validation for 53 water bodies, showing that a trained U-Net model over PlanetScope imagery performed better above that threshold. Compared to GNSS-based ground validation, which is both time-consuming and costly, our UAS surveys covered much more small water bodies (793). This also demonstrated that numerous ponds are smaller than the 100 (53%) and 166 m^2 (64%) minimum proposed size thresholds.

5.3. Improvements compared to other available products

Remote sensing products for water bodies and landcover products have been showing improvements since the beginning of this century at all levels, from regional to global. This has resulted from greatly improved data availability and quality (e.g., sensor spatial, spectral, temporal and radiometric resolutions), and increased robustness of classification techniques, including algorithms and workflows, accompanied and facilitated by greater processing power.

At the global level, Lehner and Döll (2004) assembled the Global Lakes and Wetlands Database (GLWD) with a mapping scale of 1:1 to 3 million. The authors compiled global and regional datasets digitizing historical maps and converting large lakes (≥ 50 km^2), reservoirs (storage capacity ≥ 0.5 km^3), small water bodies (surface area ≥ 0.1 km^2) and wetland point locations to polygonal features.

Verpoorter et al. (2014) significantly improved the delineation of water bodies worldwide, with the production of the Globally WAter BOdies database (GLOWABO) using remote-sensing and image classification. Having as reference a global static cloud-free mosaic of Landsat imagery aided by digital elevation data (GeoCover™ circa 2000 image dataset), the authors developed the GeoCover™ Water bodies Extraction Method (GWEM), automatically mapping approximately 117 million lakes worldwide (Verpoorter et al., 2012 and 2014).

Since GLOWABO was launched, various authors have developed other image classification algorithms for deriving and improving global water body products. Feng et al. (2016) developed the Global Land Cover Facility inland surface water dataset (GIW) and Pekel et al. (2016) the Global Surface Water Explorer (GSWE). These remote sensing-based products have in common the use of Landsat imagery at native (30 m) or pansharpened (≈15 m) resolution, sometimes aided by other datasets (e.g., digital elevation models). In 2016, Messager et al. (2016) produced the HydroLAKES database, following the GIS-based compilation approach of Lehner and Döll (2004). The authors compiled and standardized near-global and regional available datasets mapping 1,427,688 lakes with area above 10 ha.

More recently, although not focusing on surface water extent mapping and classification, Karra et al. (2021) developed the ESRI

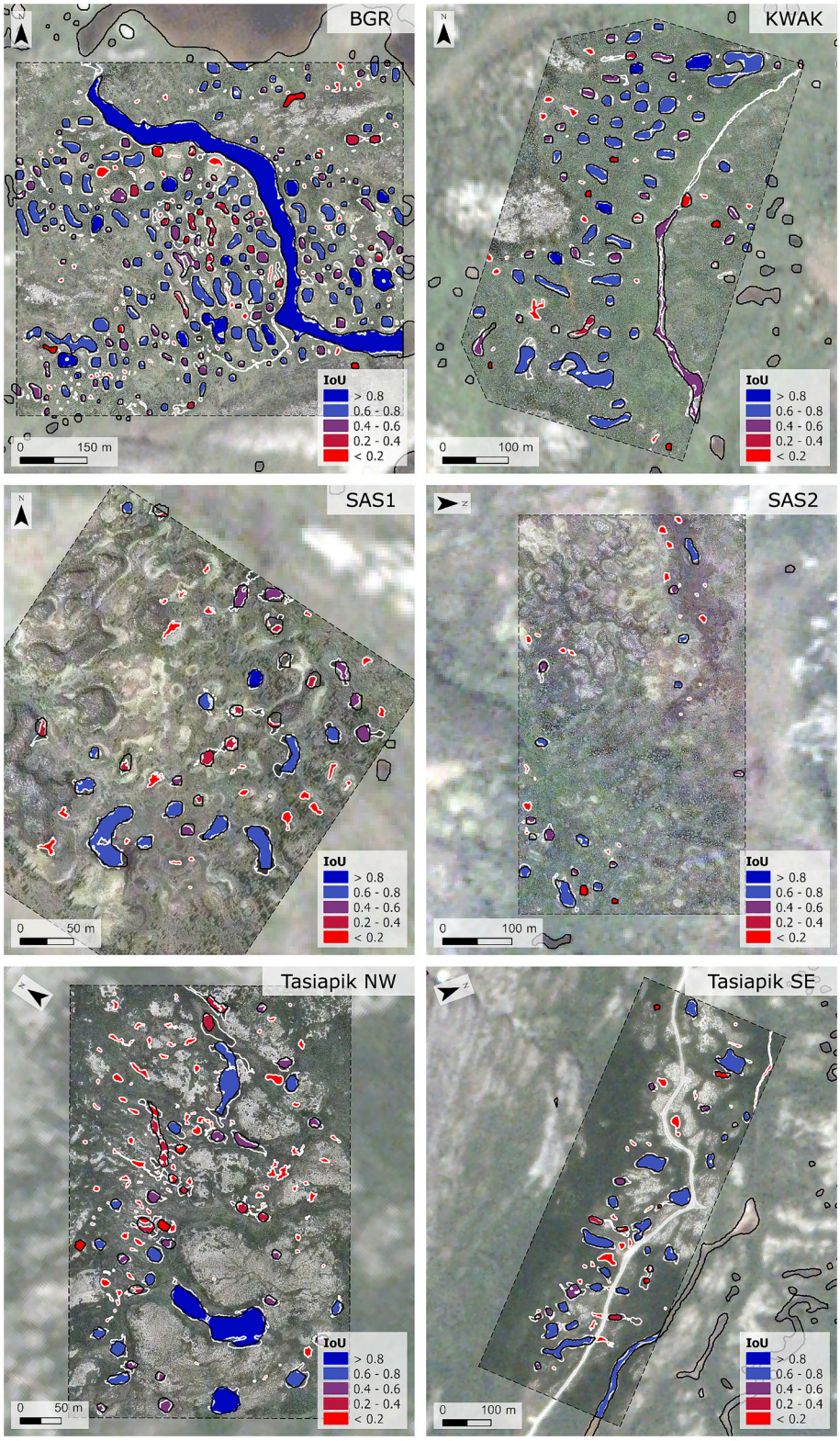

Fig. 12. IoU maps for BGR, KWAK, SAS1, SAS2, Tasiapik NW and Tasiapik SE showing the ultra-high resolution manual delineations (ground truthing features – white polygons) and automatic delineations from our trained Mask R-CNN algorithm (black polygons). The UAS (dashed sector) and PS-D imagery (remaining area) have added transparency (40%) to highlight delineations and IoU classes.

LandCover global product at 10 m spatial resolution. The authors used 5 billion manually labeled Sentinel-2 pixels and U-Net as a deep learning method for mapping nine landcover classes, including water, allowing for time-series retrievals. Zanaga et al. (2021) produced the ESA WorldCover product at 10 m, with eleven landcover classes, including permanent water bodies. The authors used Sentinel-2 and Sentinel-1 imagery, following the image classification workflow of the dynamic yearly Copernicus Global Land Service Land Cover (CGLS-LC) map at 100 m resolution (Buchhorn et al., 2020). Although these datasets are not fully adapted for water bodies, improvements in this class were

Table 7
Performance evaluation metrics for the UHR validation sites at IoU 0.5 considering all water bodies (ponds and river/creek sectors). The * values are for results based on the total area.

UHR sites	Precision	Recall	F1 Score	TP (%)	FP (%)	FN (%)
BGR	0.98 (0.89 *)	0.38 (0.73 *)	0.54 (0.80 *)	62 (67 *)	0.7 (8 *)	37 (25 *)
KWAK	0.91 (0.73 *)	0.54 (0.77 *)	0.68 (0.75 *)	51 (60 *)	5 (22 *)	44 (18 *)
SAS1	1 (0.75 *)	0.31 (0.63 *)	0.47 (0.68 *)	31 (52 *)	0 (18 *)	69 (30 *)
SAS2	0.87 (0.72 *)	0.34 (0.57 *)	0.49 (0.64 *)	63 (47 *)	5 (18 *)	33 (36 *)
Tasiapik NW	0.92 (0.84 *)	0.18 (0.56 *)	0.3 (0.67 *)	18 (51 *)	2 (9 *)	81 (40 *)
Tasiapik SE	0.96 (0.9 *)	0.41 (0.63 *)	0.58 (0.74 *)	41 (59 *)	2 (7 *)	58 (34 *)

expected as a function of the application of state-of-the-art image classification algorithms over Sentinel-1 and 2 imagery.

At the national level, the CanVec Series by Natural Resources Canada (Government of Canada, 2019), consists of a vector delineation of all types of hydrographic features. The dataset was released in 2017 and was lastly updated in February 2022. The product is available from 1:15 million to 1:50,000 mapping scales. In addition, Sui et al. (2022) developed the water body dataset for the North American high latitudes (WBD-NAHL), developing an image classification method applied over Sentinel-2 imagery.

At the regional level, the Government of Quebec provides some landcover products (e.g., ecological mapping of the vegetation, landscapes units and surface deposits), through the open forest web mapping services (Leboeuf et al., 2018). However, in these products, the water bodies derived from the CanVec Series and many small lakes were excluded as a function of cartographical generalization procedures and minimum size units mapping protocols.

For the Arctic and Subarctic, Muster et al. (2017) developed the PeRL database. This dataset includes only one map for Eastern Hudson Bay, namely for the KWAK site, consisting of precise manually delineated ponds (1959 and 2006) for an area of only 0.18 km^2 (Bouchard et al., 2014; Muster et al., 2017). For Eastern Hudson Bay specifically, Wang et al. (2018) developed a thaw pond database for part of the Sheldrake River basin and associated valleys, but it is not publicly available.

The evaluation of the datasets above described using the manually digitized ground truthing at our VHR sites, showed much weaker water body detections when compared to our model (Fig. 14). The evaluated products showed a low to high capacity for detecting large glacial basin lakes (e.g., GBL), but severe limitations for detecting smaller water

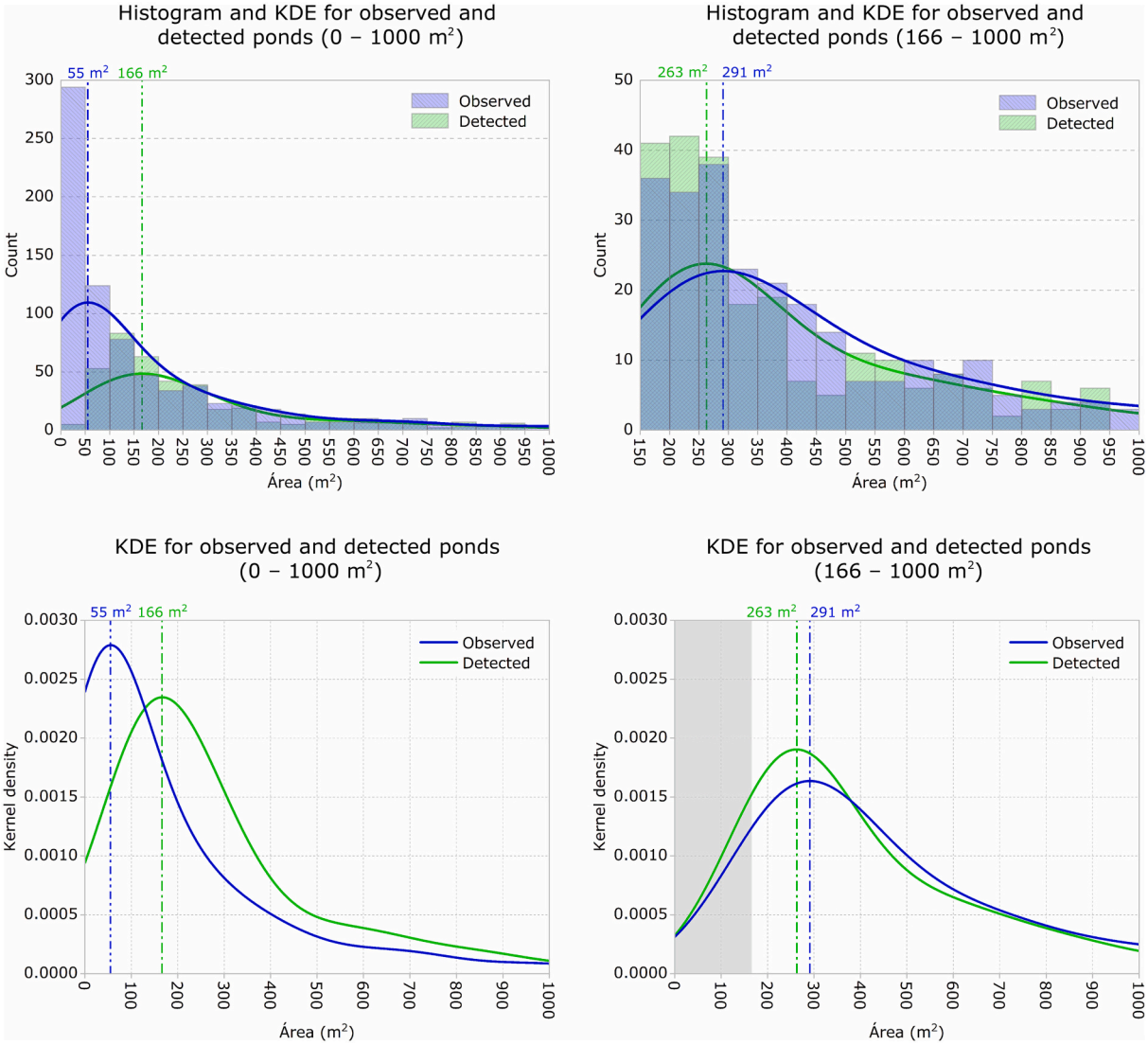

Fig. 13. Histograms and KDE plots for the detected and observed ponds using the UHR validation for water bodies from 0 to 1000 m^2 and 166 to 1000 m^2.

Table 8
Performance evaluation metrics at IoU 0.5 for the UHR validation scales considering all water bodies (no size threshold) and considering those >166 m^2. At this size threshold all metrics showed significant improvements. The * values are for the metrics results based on the total area.

UHR sites	Precision		Recall		F1 Score	
	All	> 166 m^2	All	> 166 m^2	All	> 166 m^2
BGR	0.98 (0.89 *)	0.98 (0.90 *)	0.38 (0.73 *)	0.74 (0.77 *)	0.54 (0.8 *)	0.84 (0.83 *)
KWAK	0.91 (0.73 *)	1 (0.76 *)	0.54 (0.77 *)	0.79 (0.81 *)	0.68 (0.75 *)	0.88 (0.78 *)
SAS1	1 (0.75 *)	1 (0.82 *)	0.31 (0.63 *)	0.67 (0.75 *)	0.47 (0.68 *)	0.80 (0.78 *)
SAS2	0.87 (0.72 *)	1 (0.83 *)	0.34 (0.57 *)	0.83 (0.73 *)	0.49 (0.64 *)	0.91 (0.78 *)
Tasiapik NW	0.92 (0.84 *)	1 (0.89 *)	0.18 (0.56 *)	0.54 (0.67 *)	0.3 (0.67 *)	0.7 (0.76 *)
Tasiapik SE	0.96 (0.9 *)	1 (0.91 *)	0.41 (0.63 *)	0.59 (0.67 *)	0.58 (0.74 *)	0.75 (0.77 *)

bodies (e.g., ponds) like those in BGR, KWAK, SAS and Tasiapik. On the other hand, our approach was able to enhance glacial basin lakes detections, as well as to improve the detection of ponds significantly in all VHR sites with mean 0.5 IoU F1 Scores ranging from 0.53 to 0.71 according to all confidence thresholds (Fig. 14).

Two main factors contribute to the small IoU 0.5 F1 Scores for the evaluated products: the lack of detection of water bodies, especially ponds, and also failure to detect water body connectivity (via rivers and other conduits) (Fig. 15). These results reinforce the importance of using satellite or airborne imagery with better spatial resolution and improving training labelling for small feature detections, as we present here.

5.4. Factors affecting the delineation of small water bodies

The water bodies of Eastern Hudson Bay are highly variable in terms of color and shape, and this is particularly the case for permafrost thaw lakes and ponds (Bhiry et al., 2011; Watanabe et al., 2011; Freitas et al., 2019). The performance of our trained Mask R-CNN model showed the high potential of PS-D imagery for the automatic detection and delineation of such small and diverse ponds and its applicability to these environments. However, in some cases, the model was unable to accurately detect or delineate ponds. Even though the use of UHR UAS orthomosaics is not always feasible or needed, in the tested terrain and pond settings, they allowed for the accurate analysis into these problems and on how they can impact satellite imagery driven algorithms for pond detection, providing more information and robustness compared to the traditional GNSS-based surveys or manual delineations from satellite imagery. As a result, the key challenges identified were (Fig. 16):

- Difficulty in detecting surface macrophytes and/or algae in lithalsa-formed green ponds due to their resemblance to the surrounding environment. These unique and rare aquatic ecosystems pose challenges in training deep learning models due to misrepresentation and often resulted in lower confidence detections (Fig. 16, a).
- Shadowing effects from tree clusters leading to FP and uncertainties near forested areas, impacting accurate pond identification (Fig. 16, b). This was particularly evident in KWAK. Vegetation cast shadows are also known to impact water-leaving spectral signals, affecting monitoring efforts, especially over turbid waters (Freitas et al., 2022).
- Shadowing from palsas and/or soil and water spectral similarity, as occurred in the SAS1 peatland, causing pond size overestimation (Fig. 16, c). Topographical shadows are a very common factor generating water body FP, which typically are easily mitigated by using digital elevation models (Feyisa et al., 2014; Fisher et al., 2016; Pekel et al., 2016). Recent products like the ArcticDEM, offering a high-quality digital elevation model for all land north of 60° at 2 m pixel size, hold promise (Porter et al., 2023), but were not available for our testing sites.
- Segmentation issues arising from narrow water corridors and complex pond shapes, particularly in dense areas like BGR. Whenever the water corridors linking ponds were too narrow, our trained model resulted in broken or generalized detections (Fig. 16, d). Some of these issues could be potentially resolved with higher-resolution satellite or airborne imagery.

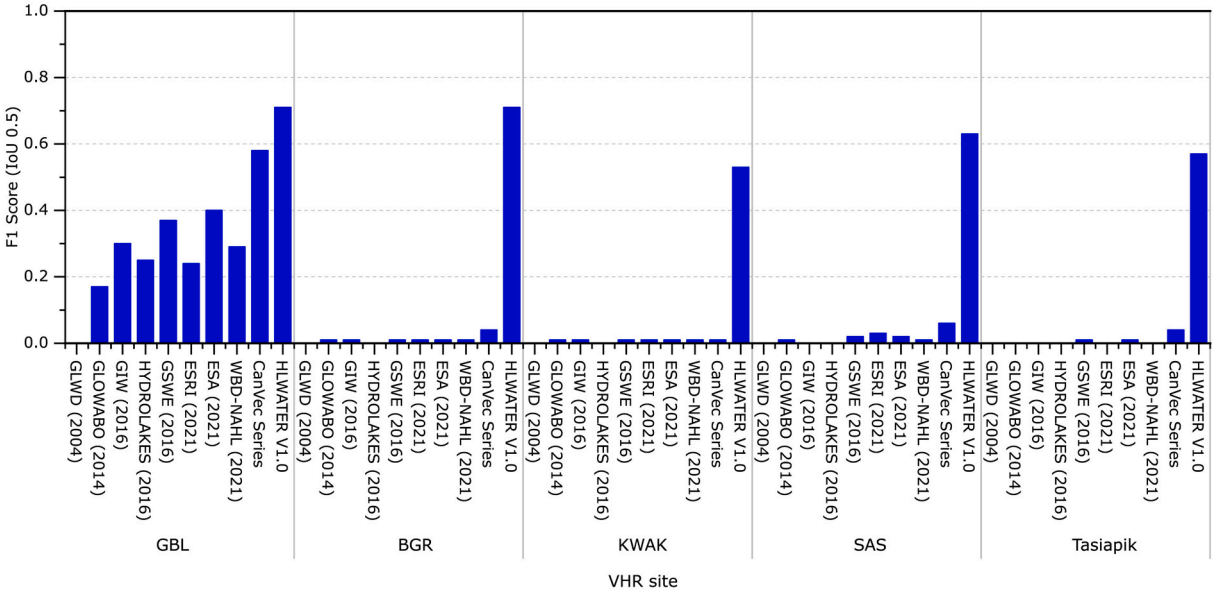

Fig. 14. IoU 0.5 F1 Scores for the VHR sites from global, national and regional products for water body and landcover analysis in comparison to the model developed here (HLWATER V1.0). For HLWATER V1.0 the IoU 0.5 F1 Scores are the means according to all confidence thresholds (5 to 95% at 5% intervals) for each VHR site.

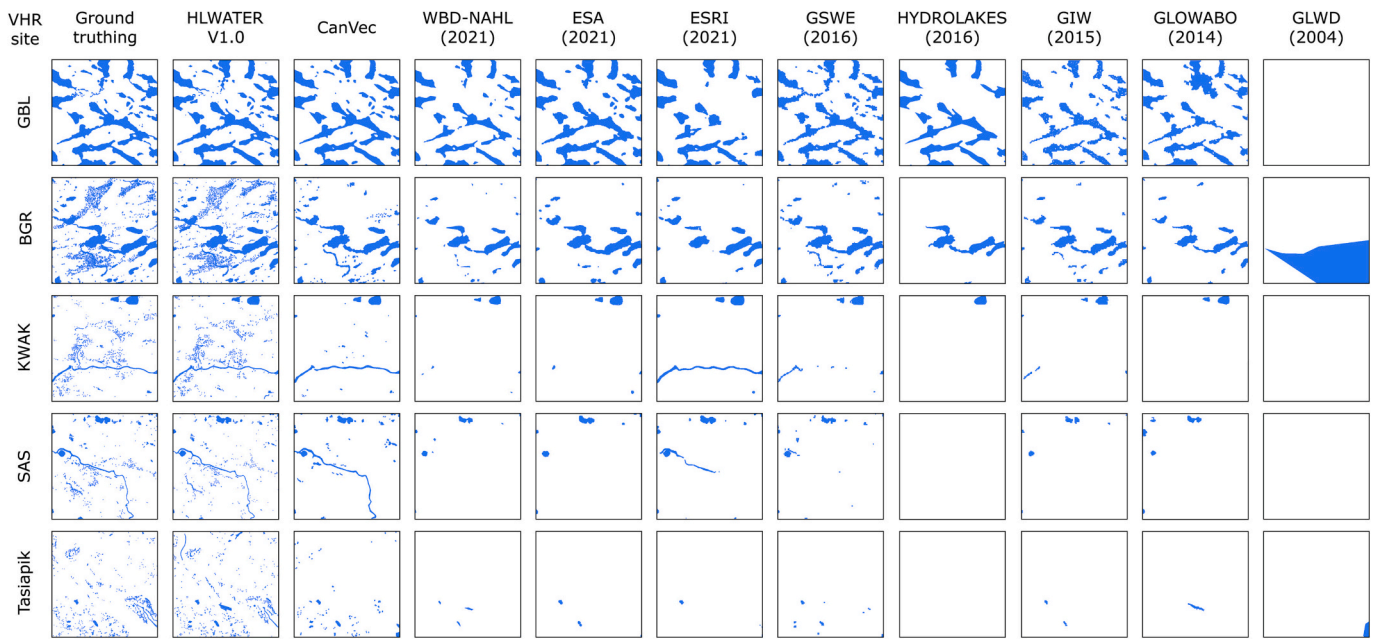

Fig. 15. Water bodies from publicly available databases and comparison with HLWATER V1.0. CanVec scale is 1:50,000. ESA WorldCover product version is 100. GLW is the union of Level 1 and 2. The grids are 3.5 × 3.5 km.

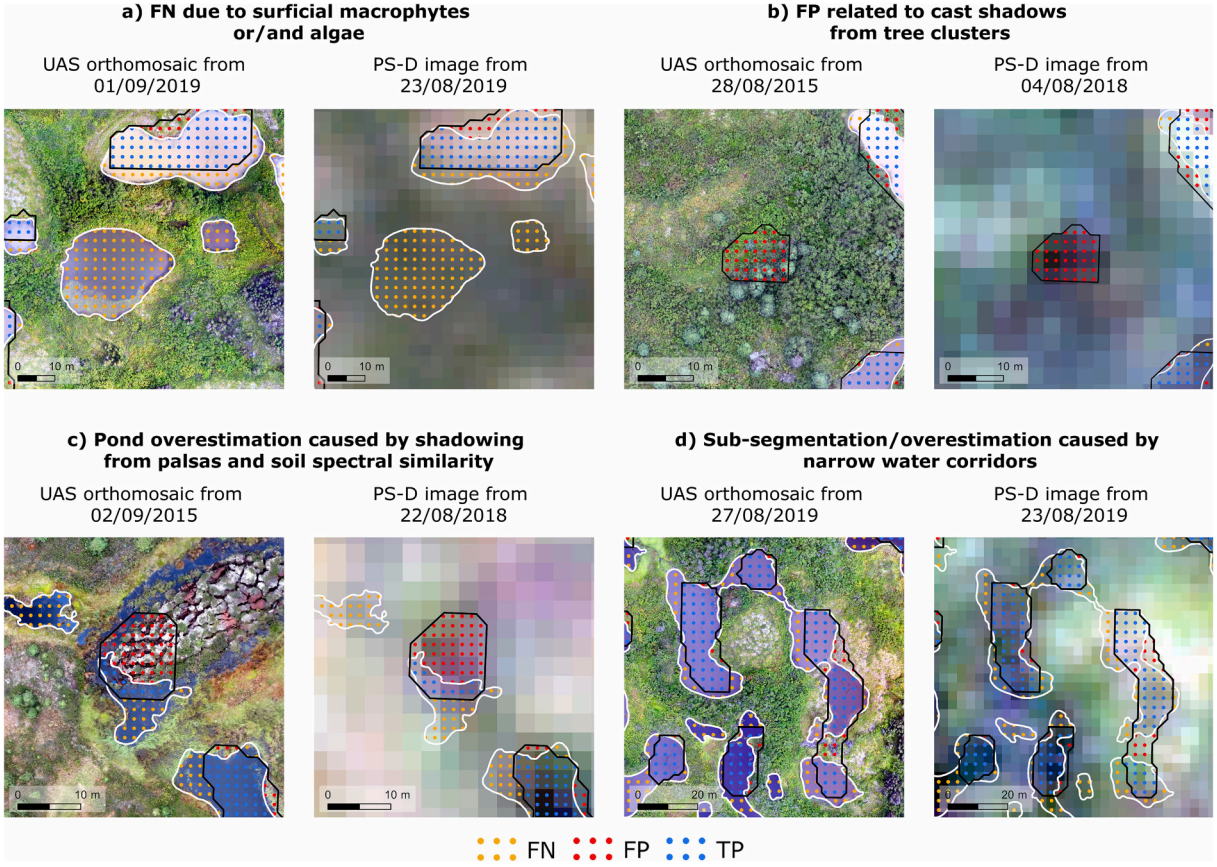

Fig. 16. Typical cases of FN and FP problems on small water bodies automatic delineations. Black and white polygons represent model and UAS delineations, respectively.

- Challenges in assessing pond depth and optical properties in certain landscapes, such as the glacial valley of Tasiapik. The comparison between Tasiapik NW and Tasiapik SE, along with our field knowledge and data, indicates that water properties and local conditions (e.g., lithology, vegetation) impact remote sensing observations (Zeng et al., 2017), affecting the accuracy of pond detection. In Tasiapik NW, our UAS data from 2019 and field sampling from 2022 show that: i. lake bottoms are easily visible when analyzing the UAS

orthomosaics; ii. the large majority of the ponds are oligotrophic, with clear water columns; iii. Ponds show very shallow depths (mean depth of 0.7 m over 10 sampled ponds). Gagnon et al. (2019) further showed that Tasiapik NW has among the lowest soil carbon stocks in the Arctic, which is consistent with the presence of low organic carbon oligotrophic ponds. At the opposite end of the spectrum, Tasiapik SE showed better model performances, with its optically deep ponds containing brownish waters and slightly deeper water columns (mean depth of 1 m on a total of 5 sampled ponds). According to the Quebec open forest web mapping services the surface deposits in Tasiapik SE are essentially composed by clay and silt which are more likely to impact water quality, as opposed to the ones of Tasiapik NW, which are composed of sand and gravel materials (Leboeuf et al., 2018).

6. Conclusions

Small pond detection models and derived datasets are essential for assessing the biogeochemical significance and variability of small water bodies. These datasets can, for instance, feed and support synergistic remote sensing observations along with in-situ measurements on water properties. Our research draws attention to the high potential of PlanetScope imagery for detecting very small ponds covering a variety of characteristics and spatial contexts, focusing on the boreal-forest tundra transition zone. Freitas et al. (2019) showed that Sentinel-2 imagery has good performance for characterizing the spectral properties of thaw ponds with >350 m^2, if their boundaries are previously delineated using VHR imagery. This supports that PlanetScope (radiometrically standardized for direct Sentinel-2 comparison), along with Sentinel-2 are highly relevant satellite platforms for monitoring the morphological and optical properties of Artic and Subarctic ponds, with relation to their limnological evolution and dynamics.

In this research, training a Mask R-CNN model over PlanetScope imagery, with detailed ground truthing features as reference, resulted in high quality and precise water body automatic delineations, even comparing with delineations from ultra-high resolution UAS orthomosaics. This comparison allowed better evaluation of model performance, as well as understanding sub-pixel issues affecting small and narrow water body delineation at PS-D native spatial resolution. It showed that better model performances are obtained when using 166 m^2 as the minimum size detection threshold. Setting this threshold allowed excellent IoU 0.5 F1 Scores ranging from 0.70 and 0.91 depending on the UHR validation site.

The performance of the model was not uniform, but at the same time showed flexibility for detecting small water bodies in different geographical contexts. It further showed potential for transfer to other regions of the Arctic and Subarctic as soon as PlanetScope datasets are available. Comparing the results of this new model with currently available global, national and regional products showed major improvements, allowing to detect and precisely delineate numerous small water bodies which were omitted in past surveys, including permafrost thaw ponds that seem to have a very important role on the carbon cycle (Kuhn et al., 2018; Heslop et al., 2020; Beckebanze et al., 2022). Implementing this or similar models over wide regional sectors will allow for greatly improved mapping and monitoring Arctic and Subarctic ponds, and improved assessment of their biogeochemical roles in the global environment.

Data and model access

The PlanetScope images can be downloaded from https://www.planet.com/products/platform/. The HLWATER V1.0 is made freely available in Zenodo (doi:https://doi.org/10.5281/zenodo.10203553). The training dataset will be provided by P. Freitas upon request.

Funding sources

This research was funded by the Portuguese Foundation for Science and Technology (FCT, Portuguese Polar Program – PROPOLAR) under the projects THAWPOND, by the Centre of Geographical Studies (IGOT, Universidade de Lisboa – FCT I.P. UIDB/00295/2020 and UIDP/00295/2020) and is also part of the project PERMAMERC – Mercury Biogeochemistry, Fate, and Impact in Permafrost Thaw Ecosystems (PTDC/CTA-AMB/4744/2020). Additional support was provided from ArcticNet (NCE), Sentinel North (CFREF) and Center for Northern Studies (CEN) Université Laval. Pedro Freitas is funded by FCT under a PhD grant (SFRH/BD/145278/2019).

CRediT authorship contribution statement

Pedro Freitas: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Gonçalo Vieira:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **João Canário:** Project administration, Supervision, Writing – review & editing. **Warwick F. Vincent:** Funding acquisition, Project administration, Supervision, Writing – review & editing. **Pedro Pina:** Formal analysis, Writing – review & editing. **Carla Mora:** Funding acquisition, Project administration, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Pedro Freitas reports financial support was provided by the Portuguese Foundation for Science and Technology.

Data availability

Data will be made available on request.

Acknowledgements

This research is framed under the College on Polar and Extreme Environments (POLAR2E) of the University of Lisbon. We would like to thank the Planet's Education and Research Program, through which we were able to obtain the Planet CubeSat Images for this research. We thank the three anonymous reviewers and the Editor, for the insightful discussion and suggestions that very much contributed to consolidate the final version of the manuscript. The Indigenous communities of Kuujjuarapik/Whapmagoostui and Umiujaq are warmly thanked for allowing us to conduct research on their Lands. In addition, we thank Dr. Ingmar Nitze, Dr. Bennet Juhls and Dr. Birgit Heim from the Alfred Wegener Institute for their suggestions, Pedro Cruz from ESRI Portugal, Prof. Jorge Rocha from GeomodLab (CEG/IGOT – ULisboa) and the CEN Aquatic Geochemistry Group at Université Laval led by Raoul-Marie Couture, for offering a broader perspective on the role of thaw lakes and ponds in the Arctic and Subarctic.

References

Abnizova, A., Siemens, J., Langer, M., Boike, J., 2012. Small ponds with major impact: the relevance of ponds and lakes in permafrost landscapes to carbon dioxide emissions. Glob. Biogeochem. Cycles 26 (2). https://doi.org/10.1029/2011GB004237.

Adegun, A.A., Viriri, S., Tapamo, J.R., 2023. Review of deep learning methods for remote sensing satellite images classification: experimental survey and comparative analysis. J. Big Data 10 (93). https://doi.org/10.1186/s40537-023-00772-x.

Aleissaee, A.A., Kumar, A., Anwer, R.M., Khan, S., Cholakkal, H., Xia, G.-S., Khan, F.S., 2023. Transformers in remote sensing: a survey. Remote Sens. 15 (1860) https://doi.org/10.3390/rs15071860.

Alem, A., Kumar, S., 2022. Deep learning models performance evaluations for remote sensed image classification. IEEE Access 10. https://doi.org/10.1109/ACCESS.2022.3215264.
Alfaro, E., Fonseca, X.B., Albornoz, E.M., Martínez, C.E., Ramrez, S.C., 2019. A brief analysis of U-net and mask R-CNN for skin lesion segmentation. In: 2019 IEEE Int. Work Conf. Bioinspired Intell, pp. 123–126. https://doi.org/10.1109/IWOBI47054.2019.9114436.
Alzubaidi, L., Zhang, J., Humaidi, A.J., Al Dujaili, A., Duan, Y., Al Shamma, O., Santamaría, J., Fadhel, M.A., Al Amidie, M., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8 (53). https://doi.org/10.1186/s40537-021-00444-8.
Arsenault, J., Talbot, J., Brown, L.E., Holden, J., Martinez-cruz, K., Sepulveda-jauregui, A., Swindles, G.T., Wauthy, M., Lapierre, J.-F., 2022. Biogeochemical distinctiveness of peatland ponds, thermokarst waterbodies, and lakes. Geophys. Res. Lett. 49 (11) https://doi.org/10.1029/2021GL097492.
Bazi, Y., Bashmal, L., Rahhal, M.M., Al, Dayil, Al, R., Ajlan, N.Al., 2021. Vision transformers for remote sensing image classification. Remote Sens. 13 (516) https://doi.org/10.3390/rs13030516.
Beckebanze, L., Rehder, Z., Holl, D., Wille, C., Mirbach, C., Kutzbach, L., 2022. Ignoring carbon emissions from thermokarst ponds results in overestimation of tundra net carbon uptake. Biogeosciences 19 (4), 1225–1244. https://doi.org/10.5194/bg-19-1225-2022.
Bégin, P.N., Vincent, W.F., 2017. Permafrost thaw lakes and ponds as habitats for abundant rotifer populations. Arct. Sci. 3 (2), 354–377. https://doi.org/10.1139/as-2016-0017.
Bhiry, N., Delwaide, A., Allard, M., Bégin, Y., Filion, L., Lavoie, M., Nozais, C., Payette, S., Pienitz, R., Saulnier-Talbot, É., Vincent, W.F., 2011. Environmental change in the great Whale River region, Hudson Bay: five decades of multidisciplinary research by Centre d'études nordiques (CEN). Écoscience 18 (3), 182–203. https://doi.org/10.2980/18-3-3469.
Biskaborn, B.K., Smith, S.L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D.A., Schoeneich, P., Romanovsky, V.E., Lewkowicz, A.G., Abramov, A., Allard, M., Boike, J., Cable, W.L., Christiansen, H.H., Delaloye, R., Diekmann, B., Drozdov, D., Etzelmüller, B., Grosse, G., Lantuit, H., 2019. Permafrost is warming at a global scale. Nat. Commun. 10 (264) https://doi.org/10.1038/s41467-018-08240-4.
Bouchard, F., Francus, P., Pienitz, R., Laurion, I., 2011. Sedimentology and geochemistry of thermokarst ponds in discontinuous permafrost, subarctic Quebec, Canada. J. Geophys. Res. Biogeosciences 116 (G2). https://doi.org/10.1029/2011JG001675.
Bouchard, F., Francus, P., Pienitz, R., Laurion, I., Feyte, S., 2014. Subarctic thermokarst ponds: investigating recent landscape evolution and sediment dynamics in thawed permafrost of northern Québec (Canada). Arct. Antarct. Alp. Res. 46 (1), 251–271. https://doi.org/10.1657/1938-4246-46.1.251.
Bouchard, F., MacDonald, L.A., Turner, K.W., Thienpont, J.R., Medeiros, A.S., Biskaborn, B.K., Korosi, J., Hall, R.I., Pienitz, R., Wolfe, B.B., 2017. Paleolimnology of thermokarst lakes: a window into permafrost landscape evolution. Arct. Sci. 3 (2), 91–117. https://doi.org/10.1139/as-2016-0022.
Box, J.E., Colgan, W.T., Christensen, T.R., Schmidt, N.M., Lund, M., Parmentier, F.W., Brown, R., Bhatt, U.S., Euskirchen, E.S., Romanovsky, V.E., 2019. Key indicators of Arctic climate change: 1971–2017. Environ. Res. Lett. 14 (045010) https://doi.org/10.1088/1748-9326/aafc1b.
Breton, J., Vallières, C., Laurion, I., 2009. Limnological properties of permafrost thaw ponds in northeastern Canada. Can. J. Fish. Aquat. Sci. 66 (10), 1635–1648. https://doi.org/10.1139/F09-108.
Buchhorn, M., Lesiv, M., Tsendbazar, N., Herold, M., Bertels, L., Smets, B., 2020. Copernicus global land cover layers — collection 2. Remote Sens. 12 (6) https://doi.org/10.3390/rs12061044.
Cooley, S.W., Smith, L.C., Stepan, L., Mascaro, J., 2017. Tracking dynamic northern surface water changes with high-frequency planet CubeSat imagery. Remote Sens. 9 (12) https://doi.org/10.3390/rs9121306.
Cooley, S.W., Smith, L.C., Ryan, J.C., Pitcher, L.H., Pavelsky, T.M., 2019. Arctic-boreal Lake dynamics revealed using CubeSat imagery. Geophys. Res. Lett. 46 (4), 2111–2120. https://doi.org/10.1029/2018GL081584.
Crate, S., Ulrich, M., Habeck, J.O., Desyatkin, A.R., Desyatkin, R.V., Fedorov, A.N., Hiyama, T., Iijima, Y., Ksenofontov, S., Mészáros, C., Takakura, H., 2017. Permafrost livelihoods: a transdisciplinary review and analysis of thermokarst-based systems of indigenous land use. Anthropocene 18, 89–104. https://doi.org/10.1016/j.ancene.2017.06.001.
Darnajoux, R., Lutzoni, F., Miadlikowska, J., Bellenger, J., 2015. Determination of elemental baseline using peltigeralean lichens from northeastern Canada (Québec): initial data collection for long term monitoring of the impact of global climate change on boreal and subarctic area in Canada. Sci. Total Environ. 533 https://doi.org/10.1016/j.scitotenv.2015.06.030.
Edwards, M., Walter, K., Grosse, G., Plug, L., Slater, L., Valdes, P., 2009. Arctic thermokarst lakes and the carbon cycle. PAGES News 17 (1), 16–18. https://doi.org/10.22498/pages.17.1.16.
Feng, M., Sexton, J.O., Channan, S., Townshend, J.R., 2016. A global, high-resolution (30-m) inland water body dataset for 2000: first results of a topographic–spectral classification algorithm. Int. J. Digit. Earth 9 (2), 113–133. https://doi.org/10.1080/17538947.2015.1026420.
Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R., 2014. Automated water extraction index: a new technique for surface water mapping using Landsat imagery. Remote Sens. Environ. 140, 23–35. https://doi.org/10.1016/j.rse.2013.08.029.
Fisher, A., Flood, N., Danaher, T., 2016. Comparing Landsat water index methods for automated water classification in eastern Australia. Remote Sens. Environ. 175, 167–182. https://doi.org/10.1016/j.rse.2015.12.055.
Folhas, D., Duarte, A.C., Pilote, M., Vincent, W.F., Freitas, P., Vieira, G., Silva, A.M.S., Duarte, R.M.B.O., Canário, J., 2020. Structural characterization of dissolved organic matter in permafrost peatland lakes. Water 12 (11). https://doi.org/10.3390/w12113059.
Freitas, P., Vieira, G., Canário, J., Folhas, D., Vincent, W.F., 2019. Identification of a threshold minimum area for reflectance retrieval from thermokarst lakes and ponds using full-pixel data from Sentinel-2. Remote Sens. 11 (6) https://doi.org/10.3390/rs11060657.
Freitas, P., Vieira, G., Mora, C., Canário, J., Vincent, W.F., 2022. Vegetation shadow casts impact remotely sensed reflectance from permafrost thaw ponds in the subarctic forest-tundra zone. Environ. Earth Sci. 81 (522) https://doi.org/10.1007/s12665-022-10640-1.
Gagnon, M., Domine, F., Boudreau, S., 2019. The carbon sink due to shrub growth on Arctic tundra: a case study in a carbon-poor soil in eastern Canada. Environ. Res. Commun. 1 (9) https://doi.org/10.1088/2515-7620/ab3cdd.
Gao, B., 1996. NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58 (3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3.
Gorham, E., 1991. Northern peatlands: role in the carbon cycle and probable responses to climatic warming. Ecol. Appl. 1 (2), 182–195. https://doi.org/10.2307/1941811.
Government of Canada, 2019. Lakes, Rivers and Glaciers in Canada - CanVec Series - Hydrographic Features. https://open.canada.ca/data/en/dataset/9d96e8c9-22fe-4ad2-b5e8-94a6991b744b (accessed 17 May 2023).
Hao, Z., Lin, L., Post, C.J., Mikhailova, E.A., Li, M., Chen, Y., Yu, K., Liu, J., 2021. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (mask R-CNN). ISPRS J. Photogramm. Remote Sens. 178, 112–123. https://doi.org/10.1016/j.isprsjprs.2021.06.003.
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, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conf. Comput. Vis. Pattern Recognit, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
He, K., Gkioxari, G., Dollár, P., Girshick, R., 2020. Mask R-CNN. IEEE Trans. Pattern Anal. Mach. Intell. 42 (2), 386–397. https://doi.org/10.1109/TPAMI.2018.2844175.
Heslop, J.K., Walter Anthony, K.M., Winkel, M., Sepulveda-Jauregui, A., Martinez-Cruz, K., Bondurant, A., Grosse, G., Liebner, S., 2020. A synthesis of methane dynamics in thermokarst lake environments. Earth Sci. Rev. 210 (103365) https://doi.org/10.1016/j.earscirev.2020.103365.
Holgerson, M.A., Raymond, P.A., 2016. Large contribution to inland water CO2 and CH4 emissions from very small ponds. Nat. Geosci. 9 (222–226) https://doi.org/10.1038/NGEO2654.
Hosang, J., Benenson, R., Schiele, Bernt, 2017. Learning non-maximum suppression. ArXiv. https://doi.org/10.48550/arXiv.1705.02950.
Huang, L., Luo, J., Lin, Z., Niu, F., Liu, L., 2020. Using deep learning to map retrogressive thaw slumps in the Beiluhe region (Tibetan Plateau) from CubeSat images. Remote Sens. Environ. 237 (111534) https://doi.org/10.1016/j.rse.2019.111534.
Hugelius, G., Loisel, J., Chadburn, S., Jackson, R.B., Jones, M.C., MacDonald, G.M., Marushchak, M.E., Olefeldt, D., Packalen, M., Siewert, M.B., Treat, C.C., Turetsky, M. R., Voigt, C., Yu, Z., 2020. Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw. PNAS Latest Artic. 117 (34), 20438–20446. https://doi.org/10.1073/pnas.1916387117.
Ji, W., Li, J., Bi, Q., Liu, T., 2023. Segment anything is not always perfect: an investigation of SAM on different real-world applications. ArXiv. https://doi.org/10.48550/arXiv.2304.05750.
Karra, K., Kontgis, C., Statman-weil, Z., Mazzariello, J.C., Mathis, M., Brumby, S.P., 2021. Global land use / land cover with Sentinel 2 and deep learning. In: 2021 IEEE Int. Geosci. Remote Sens. Symp. IGARSS, pp. 4704–4707. https://doi.org/10.1109/IGARSS47720.2021.9553499.
Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., Girshickder, R., 2023. Segment anything. ArXiv. https://doi.org/10.48550/arXiv.2304.02643.
Kuhn, M., Lundin, E.J., Giesler, R., Johansson, M., Karlsson, J., 2018. Emissions from thaw ponds largely offset the carbon sink of northern permafrost wetlands. Sci. Rep. 8 (9535), 1–7. https://doi.org/10.1038/s41598-018-27770-x.
Laurion, I., Vincent, W.F., MacIntyre, S., Retamal, L., Dupont, C., Francus, P., Pienitz, R., 2010. Variability in greenhouse gas emissions from permafrost thaw ponds. Limnol. Oceanogr. 55 (1), 115–133. https://doi.org/10.4319/lo.2010.55.1.0115.
Leboeuf, A., Morneau, C., Robitaille, A., Dufour, E., Grondin, P., 2018. Ecological Mapping of the Vegetation of Northern Québec - Mapping Standard. Ministère des Forêts, de la Faune et des Parcs.
Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.
Lehner, B., Döll, P., 2004. Development and validation of a global database of lakes, reservoirs and wetlands. J. Hydrol. 296 (1–4) https://doi.org/10.1016/j.jhydrol.2004.03.028.
Leppiniemi, O., Karjalainen, O., Aalto, J., Luoto, M., Hjort, J., 2023. Environmental spaces for palsas and peat plateaus are disappearing at a circumpolar scale. Cryosph 17, 3157–3176. https://doi.org/10.5194/tc-17-3157-2023.
Lin, T., Wang, Y., Liu, X., Qiu, X., 2021. A survey of transformers. ArXiv. https://doi.org/10.48550/arXiv.2106.04554.
Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M., 2020. Deep learning for generic object detection: a survey. Int. J. Comput. Vis. 128, 261–318. https://doi.org/10.1007/s11263-019-01247-4.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

Matveev, A., Laurion, I., Vincent, W.F., 2019. Winter accumulation of methane and its variable timing of release from thermokarst lakes in subarctic peatlands. J. Geophys. Res. Biogeosci. 124 (11), 3521–3535. https://doi.org/10.1029/2019JG005078.

Maxwell, A.E., Pourmohammadi, P., Poyner, J.D., 2020. Mapping the topographic features of mining-Related Valley fills using mask R-CNN deep learning and digital elevation data. Remote Sens. 12 (3) https://doi.org/10.3390/rs12030547.

McCrystall, M.R., Stroeve, J., Serreze, M., Forbes, B.C., Screen, J.A., 2021. New climate models reveal faster and larger increases in Arctic precipitation than previously projected. Nat. Commun. 12 (6765) https://doi.org/10.1038/s41467-021-27031-y.

McFeeters, S.K., 1996. The use of the normalized difference water index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17 (7) https://doi.org/10.1080/01431169608948714.

Messager, M.L., Lehner, B., Grill, G., Nedeva, I., Schmitt, O., 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nat. Commun. 7 (13603) https://doi.org/10.1038/ncomms13603.

Miner, K.R., Andrilli, J.D., Mackelprang, R., Edwards, A., Malaska, M.J., Waldrop, M.P., Miller, C.E., 2021. Emergent biogeochemical risks from Arctic permafrost degradation. Nat. Clim. Chang. 11, 809–819. https://doi.org/10.1038/s41558-021-01162-y.

Mohanty, S.P., Czakon, J., Kaczmarek, K.A., Pyskir, A., Tarasiewicz, P., Kunwar, S., Rohrbach, J., Luo, D., Prasad, M., Salathe, M., Schilling, M., 2020. Deep learning for understanding satellite imagery: an experimental survey. Front. Artif. Intell. 3 https://doi.org/10.3389/frai.2020.534696.

Mullen, A.L., Watts, J.D., Rogers, B.M., Carroll, M.L., Elder, C.D., Noomah, J., Williams, Z., Caraballo-vega, J.A., Bredder, A., Rickenbaugh, E., Levenson, E., Cooley, S.W., Hung, J.K.Y., Fiske, G., Potter, S., Yang, Y., Miller, C.E., Natali, S.M., Douglas, T.A., Kyzivat, E.D., 2023. Using high-resolution satellite imagery and deep learning to track dynamic seasonality in small water bodies. Geophys. Res. Lett. 50 (7) https://doi.org/10.1029/2022GL102327.

Muster, S., Heim, B., Abnizova, A., Boike, J., 2013. Water body distributions across scales: a remote sensing based comparison of three Arctic tundra wetlands. Remote Sens. 5 (4), 1498–1523. https://doi.org/10.3390/rs5041498.

Muster, S., Roth, K., Langer, M., Lange, S., Cresto Aleina, F., Bartsch, A., Morgenstern, A., Grosse, G., Jones, B., Sannel, A.B.K., Sjöberg, Y., Günther, F., Andresen, C., Veremeeva, A., Lindgren, R.P., Bouchard, F., Lara, J.M., Fortier, D., Charbonneau, S., Boike, J., 2017. PeRL: a circum-Arctic permafrost region pond and Lake database. In: Earth System Science Data (Vol. 9, Issue 1). https://doi.org/10.5194/essd-9-317-2017.

Muster, S., Riley, W.J., Roth, K., Langer, M., Aleina, F.C., Koven, C.D., Lange, S., Bartsch, A., Grosse, G., Wilson, C.J., Jones, B.M., Boike, J., 2019. Size distributions of arctic waterbodies reveal consistent relations in their statistical moments in space and time. Front. Earth Sci. 7, 1–15. https://doi.org/10.3389/feart.2019.00005.

Myers-Smith, I.H., Kerby, J.T., Phoenix, G.K., Bjerke, J.W., Epstein, H.E., Assmann, J.J., John, C., Andreu-Hayles, L., Angers-Blondin, S., Beck, P.S.A., Berner, L.T., Bhatt, U. S., Bjorkman, A.D., Blok, D., Bryn, A., Christiansen, C.T., Cornelissen, J.H.C., Cunliffe, A.M., Elmendorf, S.C., Wipf, S., 2020. Complexity revealed in the greening of the Arctic. Nat. Clim. Chang. 10, 106–117. https://doi.org/10.1038/s41558-019-0688-1.

Negandhi, K., Laurion, I., Whiticar, M.J., Galand, P.E., Xu, X., Lovejoy, C., 2013. Small thaw ponds: an unaccounted source of methane in the Canadian high Arctic. PLoS One 8 (11). https://doi.org/10.1371/journal.pone.0078204.

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

Nitze, I., Heidler, K., Barth, S., Grosse, G., 2021. Developing and testing a deep learning approach for mapping retrogressive thaw slumps. Remote Sens. 13 (21) https://doi.org/10.3390/rs13214294.

Obu, J., 2021. How much of the Earth's surface is underlain by permafrost? Case Rep. Med. 126 (5) https://doi.org/10.1029/2021JF006123.

Olefeldt, D., Hovemyr, M., Kuhn, M.A., Bastviken, D., Bohn, T.J., Connolly, J., Crill, P., Euskirchen, E.S., Finkelstein, S.A., Genet, H., Grosse, G., Harris, L.I., Heffernan, L., Helbig, M., Hugelius, G., Hutchins, R., Juutinen, S., Lara, M.J., Malhotra, A., Watts, J.D., 2021. The Boreal-Arctic wetland and lake dataset (BAWLD). Earth Syst. Sci. Data. https://doi.org/10.5194/essd-2021-140.

Osco, L.P., Wub, Q., Lemos, E.L., Gonçalves, W.N., Ramos, A.P.M., Li, J., Junior, J.M., 2023. The segment anything model (SAM) for remote sensing applications: from zero to one shot. Int. J. Appl. Earth Obs. Geoinf. 124 https://doi.org/10.1016/j.jag.2023.103540.

Padilla, R., Netto, S.L., Silva, E.A.B., 2020. A survey on performance metrics for object-detection algorithms. In: 2020 Int. Conf. Syst. Signals Image Process, pp. 237–242. https://doi.org/10.1109/IWSSIP48289.2020.9145130.

Payette, S., Delwaide, A., Caccianiga, M., Beauchemin, M., 2004. Accelerated thawing of subarctic peatland permafrost over the last 50 years. Geophys. Res. Lett. 31 (18) https://doi.org/10.1029/2004GL020358.

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

Pelletier, M., Allard, M., Levesque, E., 2019. Ecosystem changes across a gradient of permafrost degradation in subarctic Québec (Tasiapik Valley, Nunavik, Canada). Arct. Sci. 5 (1) https://doi.org/10.1139/as-2016-0049.

Philipp, M., Dietz, A., Ullmann, T., Kuenzer, C., 2022. Automated extraction of annual Erosion rates for Arctic permafrost coasts using Sentinel-1, deep learning, and change vector analysis. Remote Sens. 14 (15) https://doi.org/10.3390/rs14153656.

Pickens, A.H., Hansen, M.C., Hancher, M., Stehman, S.V., Tyukavina, A., Potapov, P., Marroquin, B., Sherani, Z., 2020. Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series. Remote Sens. Environ. 243 (111792) https://doi.org/10.1016/j.rse.2020.111792.

Pienitz, R., Doran, P.T., Lamoureux, S.F., 2008. Origin and geomorphology of lakes in the polar regions. In: Vincent, W.F., Layborn-Parry, J. (Eds.), Polar Lakes and Rivers. Oxford University Press, U.K, pp. 25–42. https://doi.org/10.1093/acprof:oso/9780199213887.003.0002.

Planet Labs, 2022. Planet Imagery Product Specifications.

Porter, C., Howat, I., Noh, M.J., Husby, E., Khuvis, S., Danish, E., Tomko, K., Gardiner, J., Negrete, A., Yadav, B., Klassen, J., Kelleher, C., Cloutier, M., Bakker, J., Enos, J., Arnold, G., Bauer, G., Morin, P., 2023. ArcticDEM – Mosaics, Version 4.1. Harvard Dataverse. https://doi.org/10.7910/DVN/3VDC4W.

Povoroznyuk, O., Vincent, W.F., Schweitzer, P., Laptander, R., Bennett, M., Calmels, F., Sergeev, D., Arp, C., Forbes, B.C., Roy-Léveillée, P., Walker, D.A., 2023. Arctic roads and railways: social and environmental consequences of transport infrastructure in the circumpolar north. Arct. Sci. 9, 297–330. https://doi.org/10.1139/as-2021-0033.

Qayyum, N., Ghuffar, S., Ahmad, H.M., Yousaf, A., Shahid, I., 2020. Glacial Lakes mapping using multi satellite PlanetScope imagery and deep learning. Int. J. Geo-Inform. 9 (10) https://doi.org/10.3390/ijgi9100560.

Quoc, T.T.P., Linh, T.T., Minh, T.N.T., 2020. Comparing U-Net convolutional network with mask R-CNN in agricultural area segmentation on satellite images. In: 2020 7th NAFOSTED Conf. Inf. Comput. Sci, pp. 124–129. https://doi.org/10.1109/NICS51282.2020.9335856.

Rantanen, M., Karpechko, A.Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., Laaksonen, A., 2022. The Arctic has warmed nearly four times faster than the globe since 1979. Nat. Commun. Earth Environ. 3 (168) https://doi.org/10.1038/s43247-022-00498-3.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: towards real-time object detection with region proposal networks. ArXiv. https://doi.org/10.48550/arXiv.1506.01497.

Robson, B.A., Bolch, T., MacDonell, S., Hölbling, D., Rastner, P., Schaffer, N., 2020. Automated detection of rock glaciers using deep learning and object-based image analysis. Remote Sens. Environ. 250 (112033) https://doi.org/10.1016/j.rse.2020.112033.

Ropars, P., Boudreau, S., 2012. Shrub expansion at the forest tundra ecotone: spatial heterogeneity linked to local topography. Environ. Res. Lett. 7 (1) https://doi.org/10.1088/1748-9326/7/1/015501.

Rouse, W.R., Douglas, M.S.V., Hecky, R.E., Hershey, A.E., Kling, G.W., Lesack, L., Marsh, P., Mcdonald, M., Nicholson, B.J., Roulet, N.T., Smol, J.P., 1997. Effects of climate change on the freshwaters of arctic and subarctic North America. Hydrol. Process. 11 (8), 873–902. https://doi.org/10.1002/(SICI)1099-1085(19970630)11:8<873::AID-HYP510>3.0.CO;2-6.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y.

Sagan, V., Peterson, K.T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B.A., Maalouf, S., Adams, C., 2020. Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. Earth Sci. Rev. 205 (103187), 1–31. https://doi.org/10.1016/j.earscirev.2020.103187.

Sarker, I.H., 2021. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput. Sci. 2 (420) https://doi.org/10.1007/s42979-021-00815-1.

Schaefer, K., Lantuit, H., Romanovsky, V.E., Schuur, E.A.G., Witt, R., 2014. The impact of the permafrost carbon feedback on global climate. Environ. Res. Lett. 9 (8) https://doi.org/10.1088/1748-9326/9/8/085003.

Schuster, P.F., Schaefer, K.M., Aiken, G.R., Antweiler, R.C., Dewild, J.F., Gryziec, J.D., Gusmeroli, A., Hugelius, G., Elchin, J., Krabbenhoft, D.P., Liu, L., Herman-Mercer, N., Mu, C., Roth, D.A., Schaefer, T., Striegl, R.G., Wickland, K.P., Zhang, T., 2018. Permafrost stores a globally significant amount of mercury. Geophys. Res. Lett. 45 (3), 1463–1471. https://doi.org/10.1002/2017GL075571.

Schuur, E.A.G., McGuire, A.D., Schädel, C., Grosse, G., Harden, J.W., Hayes, D.J., Hugelius, G., Koven, C.D., Kuhry, P., Lawrence, D.M., Natali, S.M., Olefeldt, D., Romanovsky, V.E., Schaefer, K., Turetsky, M.R., Treat, C.C., Vonk, J.E., 2015. Climate change and the permafrost carbon feedback. Nature 520, 171–179. https://doi.org/10.1038/nature14338.

Serreze, M.C., Barrett, A.P., Stroeve, J.C., Kindig, D.N., Holland, M.M., 2009. The emergence of surface-based Arctic amplification. Cryosphere 3 (1), 11–19. https://doi.org/10.5194/tc-3-11-2009.

Sit, M., Demiray, B.Z., Xiang, Z., Ewing, G.J., Sermet, Y., Demir, I., 2020. A comprehensive review of deep learning applications in hydrology and water resources. Water Sci. Technol. 82 (12) https://doi.org/10.2166/wst.2020.369.

Smith, L.C., Sheng, Y., Macdonald, G.M., 2007. A first pan-arctic assessment of the influence of glaciation, permafrost, topography and peatlands on northern hemisphere Lake distribution. Permafr. Periglac. Process. 18 (2), 201–208. https://doi.org/10.1002/ppp.

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

Tanguy, R., Whalen, D., Prates, G., Pina, P., Freitas, P., Bergstedt, H., Vieira, G., 2023. Permafrost degradation in the ice-wedge tundra terrace of Paulatuk peninsula (Darnley Bay, Canada). Geomorphology 435 (108754). https://doi.org/10.1016/j.geomorph.2023.108754.

Tarnocai, C., Canadell, J.G., Schuur, E.A.G., Kuhry, P., Mazhitova, G., Zimov, S., 2009. Soil organic carbon pools in the northern circumpolar permafrost region. Glob. Biogeochem. Cycles 23 (2). https://doi.org/10.1029/2008GB003327.

Tremblay, B., Lévesque, E., Boudreau, S., 2012. Recent expansion of erect shrubs in the low arctic: evidence from eastern Nunavik. Environ. Res. Lett. 7 (3) https://doi.org/10.1088/1748-9326/7/3/035501.

Turetsky, M.R., Abbott, B.W., Jones, M.C., Anthony, K.W., Olefeldt, D., Schuur, E.A.G., Grosse, G., Kuhry, P., Hugelius, G., Koven, C., Lawrence, D.M., Gibson, C., Sannel, A. B.K., McGuire, A.D., 2020. Carbon release through abrupt permafrost thaw. Nat. Geosci. 13, 138–143. https://doi.org/10.1038/s41561-019-0526-0.

Verpoorter, C., Kutser, T., Tranvik, L., 2012. Automated mapping of water bodies using Landsat multispectral data. Limnol. Oceanogr. Methods 10 (12), 1037–1050. https://doi.org/10.4319/lom.2012.10.1037.

Verpoorter, C., Kutser, T., Seekell, D.A., Tranvik, L.J., 2014. A global inventory of lakes based on high-resolution satellite imagery. Geophys. Res. Lett. 41 (18), 6396–6402. https://doi.org/10.1002/2014GL060641.

Vincent, W.F., 2018. Lakes: A Very Short Introduction. Oxford University Press, UK. https://doi.org/10.1093/actrade/9780198766735.001.0001.

Vincent, W.F., Laybourn-Parry, J., 2008. Polar Lakes and Rivers. Oxford University Press, U.K.. https://doi.org/10.1093/acprof:oso/9780199213887.001.0001

Vonk, J.E., Tank, S.E., Bowden, W.B., Laurion, I., Vincent, W.F., Alekseychik, P., Amyot, M., Billet, M.F., Canário, J., Cory, R.M., Deshpande, B.N., Helbig, M., Jammet, M., Karlsson, J., Larouche, J., Macmillan, G., Rautio, M., Walter Anthony, K.M., Wickland, K.P., 2015. Reviews and syntheses: effects of permafrost thaw on Arctic aquatic ecosystems. Biogeosciences 12 (23), 7129–7167. https://doi.org/10.5194/bg-12-7129-2015.

Vonk, J.E., Tank, S.E., Walvoord, M.A., 2019. Integrating hydrology and biogeochemistry across frozen landscapes. Nat. Commun. 10 (1), 3–6. https://doi.org/10.1038/s41467-019-13361-5.

Walter Anthony, K., Schneider von Deimling, T., Nitze, I., Frolking, S., Emond, A., Daanen, R., Anthony, P., Lindgren, P., Jones, B., Grosse, G., 2018. 21st-century modeled permafrost carbon emissions accelerated by abrupt thaw beneath lakes. Nat. Commun. 9 (3262) https://doi.org/10.1038/s41467-018-05738-9.

Wand, B.P., Jones, M.C., 1994. Kernel Smoothing. https://doi.org/10.1201/b14876.

Wang, L., Jolivel, M., Marzahn, P., Bernier, M., Ludwig, R., 2018. Thermokarst pond dynamics in subarctic environment monitoring with radar remote sensing. Permafr. Periglac. Process. 29 (4), 231–245. https://doi.org/10.1002/ppp.1986.

Watanabe, S., Laurion, I., Chokmani, K., Pienitz, R., Vincent, W.F., 2011. Optical diversity of thaw ponds in discontinuous permafrost: a model system for water color analysis. J. Geophys. Res. Biogeosci. 116 (G2) https://doi.org/10.1029/2010JG001380.

Wauthy, M., Rautio, M., Christoffersen, K.S., Forsström, L., Laurion, I., Mariash, H.L., Peura, S., Vincent, W.F., 2018. Increasing dominance of terrigenous organic matter in circumpolar freshwaters due to permafrost thaw. Limnol. Oceanogr. Lett. 3 (3), 186–198. https://doi.org/10.1002/lol2.10063.

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

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27 (14) https://doi.org/10.1080/01431160600589179.

Xu, X., Zhao, M., Shi, P., Ren, R., He, X., Wei, X., Yang, H., 2022. Crack detection and comparison study based on faster R-CNN and mask R-CNN. Sensors 22 (3). https://doi.org/10.3390/s22031215.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: achievements and challenges. Remote Sens. Environ. 241 (111716) https://doi.org/10.1016/j.rse.2020.111716.

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N., Ramoino, F., Arino, O., 2021. ESA WorldCover 10 m 2020 v100. Zenodo. https://doi.org/10.5281/zenodo.5571936.

Zandt, M.H., Liebner, S., Welte, C.U., 2020. Roles of Thermokarst Lakes in a warming world. Trends Microbiol. 28 (9), 769–779. https://doi.org/10.1016/j.tim.2020.04.002.

Zeng, C., Richardson, M., King, D.J., 2017. The impacts of environmental variables on water reflectance measured using a lightweight unmanned aerial vehicle (UAV)-based spectrometer system. ISPRS J. Photogramm. Remote Sens. 130, 217–230. https://doi.org/10.1016/j.isprsjprs.2017.06.004.

Zeng, F., Song, C., Cao, Z., Xue, K., Lu, S., Chen, T., Liu, K., 2023. Monitoring inland water via sentinel satellite constellation: a review and perspective. ISPRS J. Photogramm. Remote Sens. 204, 340–361. https://doi.org/10.1016/j.isprsjprs.2023.09.011.

Zhang, W., Witharana, C., Liljedahl, A.K., Kanevskiy, M., 2018. Deep convolutional neural networks for automated characterization of Arctic ice-wedge polygons in very high spatial resolution aerial imagery. Remote Sens. 10 (1487) https://doi.org/10.3390/rs10091487.

Zhou, X., 2015. Multiple auto-adapting color balancing for large number of images. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XL-7/W3, 735–742. https://doi.org/10.5194/isprsarchives-XL-7-W3-735-2015.

Zhu, X.X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing. IEEE Geosci. Remote Sens. Mag. 5 (4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307.