The Surface Water and Ocean Topography Mission (SWOT) Prior Lake Database (PLD): Lake Mask and Operational Auxiliaries

Jida Wang, Claire Pottier, Cécile Cazals, Marjorie Battude, Yongwei Sheng, Chunqiao Song, Md Safat Sikder, Xiao Yang, Linghong Ke, Manon Delhoume, et al.

Water Resources Research

RESEARCH ARTICLE
10.1029/2023WR036896

The Surface Water and Ocean Topography Mission (SWOT) Prior Lake Database (PLD): Lake Mask and Operational Auxiliaries

Jida Wang[1,2], Claire Pottier[3], Cécile Cazals[4], Marjorie Battude[4], Yongwei Sheng[5], Chunqiao Song[6,7], Md Safat Sikder[1], Xiao Yang[8], Linghong Ke[9,10], Manon Delhoume[4], Marielle Gosset[11], Rafael Reis Alencar Oliveira[12], Manuela Grippa[11], Félix Girard[11], George H. Allen[13], Xiangtao Xu[14], Xiaolin Zhu[15], Sylvain Biancamaria[16], Laurence C. Smith[17], Jean-François Crétaux[16], and Tamlin M. Pavelsky[18]

[1]Department of Geography and Geographic Information Science, University of Illinois Urbana-Champaign, Urbana, IL, USA, [2]Department of Geography and Geospatial Sciences, Kansas State University, Manhattan, KS, USA, [3]Centre National d'Études Spatiales (CNES), Toulouse, France, [4]CS Group, Toulouse, France, [5]Department of Geography, University of California, Los Angeles, CA, USA, [6]Key Laboratory of Lake and Watershed Science for Water Security | State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China, [7]University of Chinese Academy of Science Nanjing (UCASNJ), Nanjing, China, [8]Department of Earth Sciences, Southern Methodist University, Dallas, TX, USA, [9]College of Hydrology and Water Resources, Hohai University, Nanjing, China, [10]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, China, [11]Géosciences Environnement Toulouse (Université de Toulouse, CNRS, IRD, CNES), Toulouse, France, [12]Ceará State's Foundation for Meteorology and Water Resources (FUNCEME, UFC, UT, GET), Fortaleza, Brazil, [13]Department of Geosciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, [14]Department of Ecology & Evolutionary Biology, Cornell University, Ithaca, NY, USA, [15]Department of Land Surveying and Geo-informatics, The Hong Kong Polytechnic University, Hong Kong, China, [16]Université de Toulouse, Laboratoire d'Études en Géophysique et Océanographie Spatiales (LEGOS) (CNES/CNRS/IRD/UT3), Toulouse, France, [17]Department of Earth, Environmental and Planetary Sciences, Brown University, Providence, RI, USA, [18]Department of Earth, Marine and Environmental Sciences, University of North Carolina, Chapel Hill, NC, USA

Abstract Lakes are among the most prevalent and predominant water repositories on the Earth's land surface. A primary objective of the Surface Water and Ocean Topography (SWOT) satellite mission is to monitor surface water elevation, area, and storage change in lakes globally. To meet this objective, prior information on lakes, such as locations and benchmark extents, is required to organize SWOT's KaRIn observations for computing lake storage variation over time. Here, we present the SWOT mission Prior Lake Database (PLD) to fulfill this requirement. This paper emphasizes the development of the "operational PLD," which consists of (a) a high-resolution mask encompassing approximately 6 million lakes and reservoirs that meet the minimum size criterion of 1 ha, as defined in SWOT's lake observation science goals, and (b) multiple operational auxiliaries that support the lake mask in generating SWOT's standard lake vector data products. We built the prior lake mask by harmonizing the UCLA Circa-2015 Global Lake Dataset and several state-of-the-art reservoir databases. Operational auxiliaries were produced from multi-theme geospatial data to provide essential information for PLD functionality, including lake catchments and influence areas, ice phenology, relationship with SWOT prior rivers, and spatiotemporal coverage by SWOT overpasses. Globally, over three quarters of the prior lakes are smaller than 10 ha. About 97% of the lakes, constituting half of the global lake area, are fully observed at least once per orbit cycle. The PLD will be recursively improved throughout the mission lifetime and serves as a critical framework for organizing, processing, and interpreting SWOT observations over lacustrine environments with fundamental significance to lake system science.

Key Points:

- Surface Water and Ocean Topography (SWOT) Prior Lake Database (PLD) provides the foundation for generating SWOT lake vector products including area, height, and storage change
- PLD inventories ~6 million lakes with a 1-ha minimum size, 76% of which are smaller than 10 ha and 97% are fully observed per orbit cycle
- PLD contains multiple operational auxiliaries to ease lake assignment, storage change computation, and lake vector data product distribution

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. Wang and C. Pottier,
jidaw@illinois.edu;
claire.pottier@cnes.fr

Citation:

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

Received 9 DEC 2023
Accepted 11 DEC 2024

Author Contributions:

Conceptualization: Jida Wang, Claire Pottier, Cécile Cazals, Marjorie Battude, Yongwei Sheng, Chunqiao Song, Sylvain Biancamaria, Laurence C. Smith, Jean-François Crétaux, Tamlin M. Pavelsky
Data curation: Jida Wang, Claire Pottier, Cécile Cazals, Marjorie Battude, Yongwei Sheng, Chunqiao Song, Md Safat Sikder, Xiao Yang, Marielle Gosset, Rafael Reis Alencar Oliveira, Manuela Grippa, Félix Girard

1. Introduction

Natural lakes and manmade reservoirs, hereafter "lakes," are among the most predominant components of land surface hydrology (Messager et al., 2016; Verpoorter et al., 2014). They collectively store nearly 90% of liquid freshwater on the Earth's surface, providing a readily accessible water resource for societal use (Abbott et al., 2019; Oki & Kanae, 2006). Lakes also represent diverse and complex aquatic ecosystems, offering unique esthetic appeals in the landscape and indispensable sources of biodiversity, food, and recreation outlets (Herdendorf, 1984). Although considered lentic systems, lakes are often dynamic, with water storage and quality reflective of basin-

Formal analysis: Jida Wang, Claire Pottier, Cécile Cazals, Marjorie Battude, Yongwei Sheng, Chunqiao Song, Md Safat Sikder, Xiao Yang, Linghong Ke, Manon Delhoume
Funding acquisition: Jida Wang, Yongwei Sheng
Investigation: Jida Wang, Claire Pottier, Cécile Cazals, Marjorie Battude, Yongwei Sheng, Chunqiao Song, Manon Delhoume, Tamlin M. Pavelsky
Methodology: Jida Wang, Claire Pottier, Cécile Cazals, Marjorie Battude, Yongwei Sheng, Chunqiao Song, Md Safat Sikder, Xiao Yang, Manon Delhoume, Tamlin M. Pavelsky
Project administration: Jida Wang, Claire Pottier, Yongwei Sheng
Resources: Jida Wang, Claire Pottier, Cécile Cazals, Marjorie Battude, Yongwei Sheng, Xiangtao Xu, Xiaolin Zhu
Software: Jida Wang, Claire Pottier, Cécile Cazals, Marjorie Battude, Yongwei Sheng, Chunqiao Song, Manon Delhoume
Supervision: Jida Wang, Claire Pottier, Yongwei Sheng, Sylvain Biancamaria, Laurence C. Smith, Jean-François Crétaux, Tamlin M. Pavelsky
Validation: Jida Wang, Yongwei Sheng, Chunqiao Song, Linghong Ke, Xiangtao Xu, Xiaolin Zhu
Visualization: Jida Wang, Claire Pottier, Cécile Cazals, Yongwei Sheng, Manon Delhoume
Writing – original draft: Jida Wang, Claire Pottier, Md Safat Sikder, Xiao Yang, George H. Allen
Writing – review & editing: Jida Wang, Claire Pottier, Cécile Cazals, Chunqiao Song, Md Safat Sikder, Manon Delhoume, Rafael Reis Alencar Oliveira, Manuela Grippa, George H. Allen, Sylvain Biancamaria, Laurence C. Smith, Jean-François Crétaux, Tamlin M. Pavelsky

scale hydrology and/or anthropogenic activities (Fergus et al., 2017; Wurtsbaugh et al., 2017; Yang, O'Reilly, et al., 2022). Lakes also sequester a large amount of carbon from the watersheds and modulate terrestrial carbon cycling through water storage variation and lacustrine-fluvial interactions (Mendonca et al., 2017; Tranvik et al., 2009). For these reasons, lakes serve as both "sentinels" and "regulators" of climate change (Adrian et al., 2009; Schindler, 2009) and are recognized as an "Essential Climate Variable" by the Global Climate Observing System (GCOS) of the World Meteorological Organization (WMO, 2022). Monitoring temporal dynamics of global lakes, including water extent and level that are essential to deriving storage variability, has important ramifications for hydrology, ecology, the carbon cycle, and water sustainability (Yao et al., 2023).

Our capability to monitor global lake dynamics has been rapidly advancing with the expanding Earth-observing system (Crétaux et al., 2016). But, until recently, individual satellite missions for surface hydrology measured either water extent, such as through spectral radiometers and Synthetic Aperture Radar (SAR) imagers, or water surface elevation (WSE), such as through nadir-looking radar and lidar altimeters. This dilemma challenged the monitoring of water storage variation, which requires a synchronous acquisition of both variables. In addition, conventional radar altimeters usually have coarse footprint sizes (~10 km^2 or greater) and large inter-track distances (~50–100 km or wider), limiting adequate measurements to a few thousand of the world's largest lakes (Busker et al., 2019; Crétaux et al., 2011, 2016; Schwatke et al., 2015; Yao et al., 2023, 2024). With improvements of waveform processing methods, SAR-mode altimeters such as those onboard Sentinel-3A and Sentinel-3B showed potential for measuring WSEs of lakes as small as a few hectares (Boy et al., 2022). Smaller footprints (~11–70 m) were also enabled by laser altimeters such as the Ice, Cloud, and land Elevation Satellite (ICESat) and its successor ICESat-2. However, their multi-month repeat cycles and discrete nadir footprints limit the temporal density of WSE measurements for medium-sized and small lakes (Cooley et al., 2021; Luo et al., 2022; Ryan et al., 2020). Fortunately, these technical challenges have been largely overcome by the Surface Water and Ocean Topography (SWOT) satellite mission (Biancamaria et al., 2016; Fu et al., 2024), recently launched on 16 December 2022.

The main payload of SWOT is a Ka-band (8.6 mm wavelength) Radar Interferometer (KaRIn). As the first of its kind, KaRIn provides synchronous, wide-swath, and orbital surveys of both surface water extent and elevation, allowing for the derivations of river discharge and lake storage change (Biancamaria et al., 2016; Durand et al., 2010). With a near-polar orbit inclined at 78° and a wide-swath configuration (2 × 50 km), SWOT observes more than 90% of the Earth's surface area during each 21-day science or nominal orbit cycle (JPL internal document, 2018). Its lake observation requirement includes terrestrial surface water bodies larger than 250 × 250 m^2 (i.e., 6.25 ha), with an observation goal of detecting lakes as small as 100 × 100 m^2 (i.e., 1 ha) (Biancamaria et al., 2016). While this spatiotemporal coverage will reveal unprecedented details of global lake storage variability, a critical prerequisite for SWOT lake data production is the development of a Prior Lake Database (PLD).

The fundamental purpose of the SWOT PLD is to provide prior data on known lake locations, hereafter "prior lakes," making it possible to compile KaRIn observations over time and to compute lake storage variation. KaRIn observes terrestrial water features (e.g., lakes and rivers) at a High Rate (HR) mode with fine spatial resolution (~5 m × 10–70 m) (Biancamaria et al., 2016). To accommodate user needs, the HR raw data are processed by the Science Algorithm Software (SAS) to different levels of data products, which range from Level 1 single-look complex SAR images (L1B_HR_SLC) (JPL internal document, 2022a) intended only for highly specialized applications, to the standard Level 2 vector products delivering readily useable variables specific to rivers and lakes. The initial HR product suitable for general hydrological purposes is the "pixel cloud" (L2_HR_PIXC) (JPL internal document, 2022d), which consists of geolocated pixel points with measured water heights but is not organized to distinct water features. With the help of the SWOT River Database (SWORD) (Altenau et al., 2021), pixels associated with prior rivers are first identified to process the standard river vector products (JPL internal document, 2022b, 2022c, 2023). Such river pixels, except those also on SWORD-connected lakes, are eliminated from further lake processing. The remaining PIXC pixels are processed with assistance of the PLD to the standard lake vector products, which deliver the dynamics and uncertainties of WSE, area, and storage change (when applicable) for individual prior lakes per orbit pass (L2_HR_LakeSP) or cycle (L2_HR_LakeAvg) (CNES internal document, 2022a, 2022b).

Two primary components are required to fulfill the purpose of the PLD (Figure 1). As lakes are often dynamic over time, their water surface may split and coalesce, and new lakes may emerge whereas others disappear.

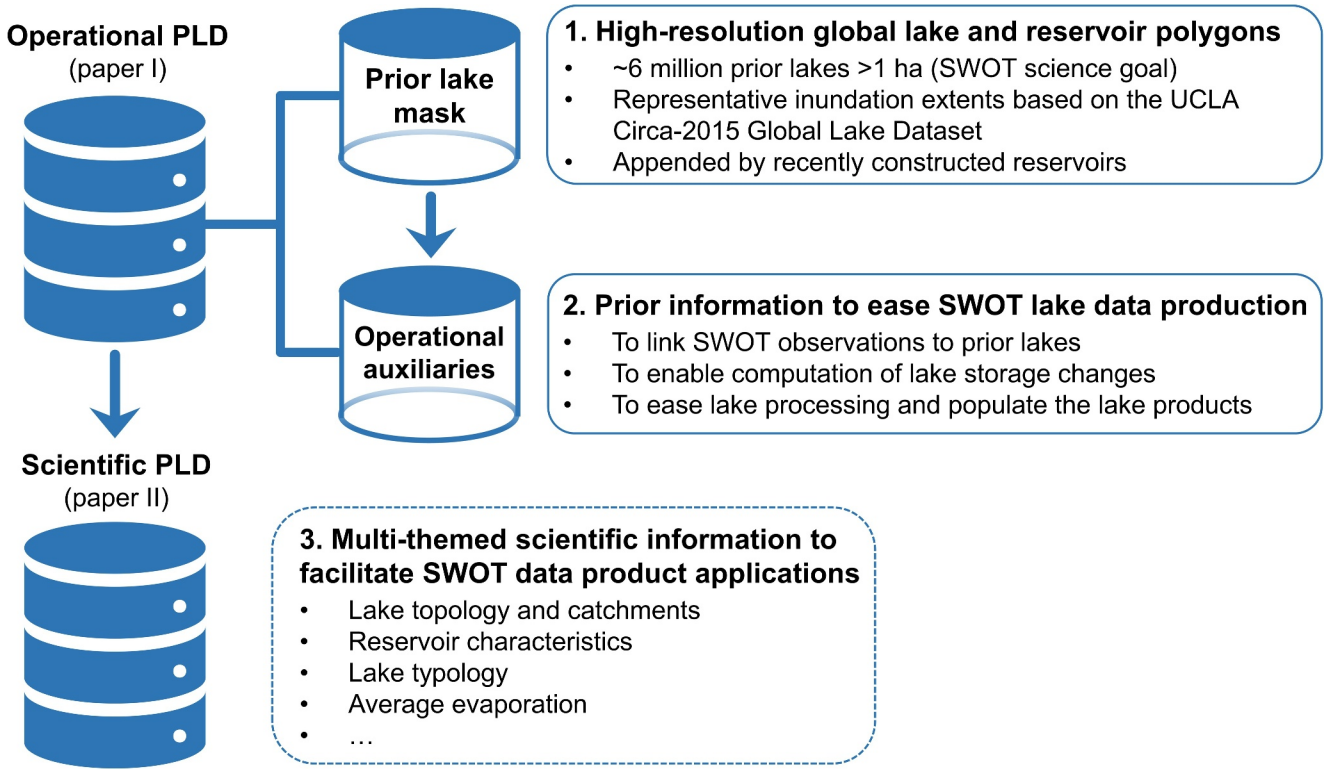

Figure 1. Conceptual structure of the SWOT Prior Lake Database (PLD).

Without defining lakes a priori, it would be difficult to sort out how water features observed in different periods are spatially related to each other, which would then pose a challenge for effectively comparing lake changes. So, the first component of the PLD is a comprehensive prior mask that inventories global lakes larger than SWOT's observation goal (1 ha). Albeit temporally static, this lake mask offers a standardized spatial reference, based on which observed water features can be assigned, aggregated, or partitioned to the corresponding prior lakes. This ensures water dynamics, especially storage change, are characterized and delivered consistently at the scale of each known lake. On the other hand, the lake mask also identifies observed water features that cannot be assigned to any prior lake. These unassigned features will be used to recursively improve the prior lake mask as SWOT data accumulate and to investigate changes in wetlands, newly emerged lakes, and similar phenomena. To make the prior lake mask functional, we need the second component of the PLD, namely "operational auxiliaries," which supplement the prior lakes with other necessary attributes, geometries, and logical information. This additional prior information works synergistically to ease the linkage of SWOT observations to the prior lakes, the calculation of lake storage change, and the population of SWOT lake vector data products.

An accurate prior lake mask is essential to the function of the PLD. We consider the defining criteria of a prior lake mask to be "exclusive" (excluding non-lake features), "representative" (with lake polygons depicting representative rather than extreme inundation conditions), and "exhaustive" (including lakes $\geq$1 ha as thoroughly as possible). At the current stage of the SWOT mission, we prefer representative water extent because it resembles how a lake typically appears when observed by SWOT, which eases the spatial linkage of SWOT observations to the prior lake. Despite the recent proliferation of global lake data sets, none of them alone can meet all three criteria. Two fine-resolution and publicly accessible global lake masks are HydroLAKES (v1.0) (Messager et al., 2016), which inventories 1.4 million lakes larger than 10 ha, and GLAKES, which comprises 3.4 million polygons depicting the maximum lake water extents large than 3 ha (Pi et al., 2022). The primary data source of HydroLAKES for the landmass below 60° N is the Shuttle Radar Topography Mission Water Body Data (SWBD) (Farr et al., 2007), where lake extents were based on water occurrence during February 2000. This timing concurred with the dry winter season across a large proportion of the northern hemisphere, meaning the areas of many lakes in HydroLAKES are likely skewed toward their seasonal minimums. In addition, SWBD was acquired over 20 years

ago, predating prominent recent lake changes such as the shrinkage of many saline lakes (Wang et al., 2018; Wurtsbaugh et al., 2017), the expansion of glacial lakes (Nie et al., 2017; Shugar et al., 2020; Song et al., 2017), and a boom in new reservoir construction (Wang et al., 2022a; Wu et al., 2023; Yao et al., 2023). In comparison, GLAKES used the Landsat-derived Global Surface Water Occurrence (GSWO) data set (Pekel et al., 2016) to extract all-time water area maximum from 1984 to 2019, where non-lake features (e.g., rivers, estuaries, and floodwaters) were removed by a deep-learning algorithm (Pi et al., 2022). While GLAKES is more up-to-date and of high overall quality, its lake polygons correspond to the maximum extreme. Based on our visual inspection, these maximum extents occasionally encompass inundated riparian zones and wetlands (e.g., 35.325°N, 110.415°E; 48.454°N, 44.943°E), floodplains (e.g., 19.423°S, 22.687°E; 0.366°N, 62.468°W; 11.905°N, 104.844°E), and paddy fields (e.g., 30.272°N, 92.420°W; 14.389°N, 100.095°E). Critically, neither data set achieves a minimum lake size of 1 ha, meaning that lakes potentially visible to SWOT are not fully inventoried.

We describe the development of the SWOT PLD in a pair of companion papers (Figure 1). The first paper (this article) emphasizes the prior lake mask and its operational auxiliaries, which constitute the "operational PLD," addressing the above-described fundamental purpose for assisting SWOT lake data production. The second paper (in preparation) will focus on the development of a "scientific PLD," consisting of multi-themed scientific attributes, features, and metadata to facilitate a wide range of limnological applications using SWOT lake data products. We next describe the input data sources (Section 2) and methods (Section 3) used to construct the operational PLD. This is followed by results (Section 4) comparing the prior lake mask with other lake data sets, the theoretical SWOT coverage of global lakes during a nominal orbit cycle, and the functionality of the operational auxiliaries. With a primary focus on data development instead of algorithms (SASs), this paper does not elaborate how lake storage change is computed. However, we do describe the purpose of each prior attribute including those for computing lake storage change (Section 3) and illustrate how the PLD works to ease SWOT lake data production (Section 4). We conclude the paper by discussing plans for future PLD improvements and versioning (Section 5).

2. Input Data Sources

We leveraged multiple data sources to compose the operational PLD. These input data and their contributions are summarized in Table 1. The primary data source is the UCLA Circa-2015 Global Lake Dataset (Sheng et al., 2016), which provides most of the polygons in the high-resolution prior lake mask. A collection of other input data, covering the themes of lake name, reservoir identity, prior river locations, hydrography, and SWOT orbits, were used to populate the prior attribute information and derive other operational auxiliaries. Details of the input data are described below.

2.1. UCLA Circa-2015 Global Lake Dataset

The foundation of the PLD, that is, the prior lake mask, mainly comes from the UCLA Circa-2015 Global Lake Dataset, or hereafter "Circa-2015" lake data set. The Circa-2015 lake data set is a global extension from the continental lake map for Oceania produced by Sheng et al. (2016): it was produced by the same research team, following data sources and mapping methods consistent with those described in Sheng et al. (2016). The Circa-2015 lake data set inventories representative inundation extents of 9.0 million open-water lakes and reservoirs larger than 0.4 ha (i.e., four 30-m-resolution Landsat pixels) globally. These lakes were mapped from a selection of high-quality Operational Land Imager (OLI) images acquired during the initial ∼2.5 years of the Landsat-8 mission operation (May 2013 to August 2015). A summary of the mapping procedure, including image selection, water extraction algorithm, quality assurance and quality control (QA/QC), and multi-scene composition, is provided in the rest of this section, but we suggest consulting Sheng et al. (2016) for a full description of the methods.

Compared to other global lake data, a unique merit of the Circa-2015 data set is the emphasis on representative lake extents, echoing one of the three criteria expected for the SWOT prior lake mask (Section 1). Specifically, the images selected for mapping were acquired during the region-specific "lake stable season" to minimize the misrepresentation of lake size due to intra-annual inundation extremes. The lake stable season was defined as the optimal lake mapping month that is ice-free and after the rainy season, when inflows equal outflows and the lake reaches a relatively stable water budget within the annual cycle. To implement this idea, an image selection tool "LakeTime" (Lyons & Sheng, 2018) was developed using multi-decadal climate data and a simple water

Table 1
Data Sources Used to Develop the Operational PLD

Data source	Contribution
UCLA Circa-2015 Global Lake Dataset (Sheng et al., 2016)	Provides the primary data source of the prior lake mask
Georeferenced global Dams And Reservoirs (GeoDAR) data set v1.1 (Wang et al., 2022a)	Supplements the UCLA Circa-2015 Global Lake Dataset with additional lakes and reservoirs
Global Reservoir Inventory of the post-2000 impoundment (GREI-p2k) (Fan et al., 2024)	
Other regional and miscellaneous water body data (see Section 3.2)	
Global Reservoir and Dam database (GRanD) v1.3 (Lehner et al., 2011)	Provides the identities of large reservoirs
HydroBASINS v1.c (Lehner & Grill, 2013)	Populates basin IDs and produces the *basin* table
SWORD v17 (Altenau et al., 2021)	Identifies spatial relationships between prior lakes and prior rivers, including river-connected lakes and lake-river proximity
SWOT orbit swaths (https://www.aviso.altimetry.fr/en/missions/current-missions/swot/orbit.html)	Populates attributes related to SWOT data coverage of prior lakes
Global Lakes and Wetlands Database (GLWD) (Lehner & Döll, 2004)	Populates lake names
HydroLAKES v1.0 (Messager et al., 2016)	
Natural Earth Data (scale 1:30,000,000) (https://www.naturalearthdata.com)	
OpenStreetMap (OSM; https://www.openstreetmap.org)	
The IGN Carthage database (BD CARTHAGE®) (https://services.sandre.eaufrance.fr/telechargement/geo/ETH/BDCarthage/FXX/2017)	
Vector Map Level 0 (VMap0) (https://mdl.library.utoronto.ca/collections/geospatial-data/vector-map-level-0-vmap0)	

balance model to determine the optimal mapping month independently for each Landsat tile. Due to variable climate conditions, the determined optimal mapping month varies across the globe, but the months of July to October dominate the tiles. For most tiles, OLI image scenes within ∼45 days of the middle of the optimal month were selected from the initial 2.5 years of the mission operation. The selected scenes also only included low cloud contamination, usually <10%–30% depending on image availability. These criteria ensured that all tiles have at least one high-quality OLI scene available, although the number of available scenes generally increases with aridity. This image selection process rendered a total of ∼60,000 OLI scenes globally, with an average of about six scenes per tile.

For each selected scene, open water was segmented from land using a hierarchical and self-adaptive algorithm to ensure lakes across different landscapes were mapped as accurately and thoroughly as possible. Together with a minimum mapping unit of 0.4 ha, the algorithm aimed to address the second criterion "exhaustive" for the prior lake mask. Since lakes are diverse aquatic systems, multiple factors such as water turbidity, mineral and chlorophyll contents, ice and snow, and mountain shadows can all complicate their spectral characteristics. To tackle this challenge, the adaptive mapping algorithm was automated to simulate how a human operator segments lakes from the background landscapes (Li & Sheng, 2012). In brief, each Landsat scene was first transformed to a normalized difference water index (NDWI) image (McFeeters, 1996) to enhance water appearance and suppress other land covers. Then, the algorithm performs a two-step, "global-to-local" segmentation. In the global segmentation, a loose preliminary NDWI threshold was used to initiate the extraction of potential lake extents from the entire scene. In the local segmentation, each extracted lake was re-examined as an object, and the boundary was fine-tuned by an updated NDWI threshold, determined only using the spectral histogram based on the vicinity of the lake. The local segmentation was implemented iteratively until the result converged to a stable water extent. Through this design, the final threshold and lake extent were tailored optimally to the unique spectral condition for each lake. Validated against shoreline perimeters acquired in situ around a dozen of Alaskan thermokarst lakes, the mapped lake water areas using this self-adaptive algorithm showed a relative error of 2%–8% for lakes of ∼1 ha and below 1%–2% for lakes larger than 10 ha (Lyons et al., 2013).

To further increase the accuracy, the mapping for Circa-2015 lakes used a previously produced Circa-2000 lake map, which was generated from Landsat 7 ETM+ scenes collected between 1999 and 2003 (Sheng et al., 2016),

as a historical reference to help determine the initial NDWI thresholds for the scene-level global segmentation. In addition, the mapping algorithm also included a mechanism that applied digital elevation models and the solar angles at the image acquisition time to automatically reduce commission errors (false positives) caused by mountain shadows.

Following the automated mapping, a rigorous QA/QC process was performed with reference to the source OLI images to further optimize the product quality. On the one hand, the QA/QC process served as a visual validation of the automated mapping result. While the self-adaptive NDWI thresholding proved to be highly effective in characterizing water extents and shoreline complexity under various spectral conditions, the iterative nature of the segmentation algorithm tended to identify only pixels having a larger water fraction as water (Lyons et al., 2013). This limitation resulted in conservative water areas for small lakes that are more subject to littoral spectral mixture. The QA/QC did not intend to correct this systematic underestimation, but its impact is acknowledged and assessed in Section 4.3. On the other hand, the QA/QC was necessary to correct the remaining random errors, such as false positives caused by cloud shadows and false negatives due to high algae or sediment concentrations. A semi-automated editing tool (Wang et al., 2014) was used to increase the accuracy and efficiency of such corrections. Free-flowing river segments were also removed using both automatically generated shape statistics and manual editing. Floodwater and coastal aquacultural impoundments were removed manually if visually detected. As such, the resultant mapping contained only water bodies deemed as lakes and reservoirs, and thus satisfies the third criterion "exclusive" for the SWOT prior lake mask.

The quality-controlled vector lake extents from multi-temporal scenes were then composited across the Landsat tiles. The composition dissolved partial lake extents due to scene boundary cutoffs and selected the median water extent from the available multi-temporal mappings as the representative extent for each lake. When a lake presence was less certain (e.g., the case of ephemeral lakes), the Circa-2000 lake map was referenced again, and the historical lake extents from circa 2000 were used to help infer the possibility of lake existence in circa 2015 (see Sheng et al. (2016) for more technical details). As the outcome of this multi-scene composition, the final Circa-2015 lake data set is a single-layer vector mosaic representing the water extent during the lake stable season for each lake. To align with SWOT's observation goals, the subset of the Circa-2015 lake data set consisting of lakes equal to or larger than 1 ha was used as the PLD prior lake mask.

2.2. Additional Reservoir Polygons

To ensure that the prior lake mask includes major reservoirs as thoroughly as possible, we supplemented the Circa-2015 lake data set with another two global reservoir inventories. They are the Georeferenced global Dams And Reservoirs data set (GeoDAR) v1.1 (Wang et al., 2022a) and the Global Reservoir Inventory of the post-2000 impoundment (GREI-p2k) (Fan et al., 2024). GeoDAR v1.1 consists of 24,783 dam locations and their associated reservoir polygons when detectable. The dam points harmonized the Global Reservoir and Dam database (GRanD) v1.3 (Lehner et al., 2011) (see Section 2.4) and a georeferenced subset of the World Register of Dams (WRD) from the International Commission on Large Dams (ICOLD; https://www.icold-cigb.org). Reservoir polygons were retrieved for each of the dam points by jointly using the water masks of HydroLAKES v1.0, GRanD v1.3, and the Circa-2015 lake data set. This led to 21,515 reservoir polygons with a total area of 496,313.8 km^2, representing a total storage capacity of 7,216.1 km^3.

GREI-p2k contains 6,760 global reservoirs constructed after the year 2000. These post-2000 reservoirs were detected by comparing composite water occurrence probabilities before and after 2000, using the multi-decadal remote sensing products Global Surface Water (GSW) database (Pekel et al., 2016) and the Global Land Analysis and Discovery (GLAD) database (Pickens et al., 2020). Polygons of the verified post-2000 reservoirs were then retrieved using the maximum water occurrence maps of GSW and GLAD, such that each polygon represents the maximum inundation area of the reservoir from the construction to about 2020 and has a minimum size threshold of 0.5 km^2. These post-2000 reservoir polygons have a total area of 53,183.9 km^2, corresponding to a total storage capacity of 1,287.7 km^3.

2.3. SWORD

SWORD is the official a priori river database for SWOT (Altenau et al., 2021). It defines the global networks of mainstems and tributaries potentially visible to SWOT (i.e., wider than 50 m according to SWOT's river observation goal) (Biancamaria et al., 2016) and serves as the organization framework for the SWOT river vector

products. Because its primary data source is the Global River Widths from Landsat (GRWL) database (Allen & Pavelsky, 2018), SWORD also contains river reaches with mean annual flow widths as narrow as 30 m. In total, SWORD (version 17) consists of more than 225,600 river reaches (centerlines) with a median length of 10.4 km, comprising 11.1 million nodes with ∼200 m spacing. The SWOT river vector products, which contain WSE, width, slope, and discharge, will be disseminated at the scales of both river reach and node. In addition, SWORD also used multiple auxiliary data sets to provide a wide range of hydrological and morphological attributes such as reach sinuosity, average width, slope, natural and human-created obstructions, and the topology structure among the reaches and nodes. These attributes facilitate the processing of SWOT river products as well as their scientific applications. Here we used the latest SWORD version 17 to identify the PLD lakes that are directly connected to the river networks visible to SWOT, enabling intersecting water bodies to be considered in both lake and river products.

2.4. GRanD

GRanD is among the most comprehensive spatial repositories of large dams and reservoirs worldwide (Lehner et al., 2011). GRanD was constructed by harmonizing a collection of open-access dam and reservoir data, including the United Nations Food and Agricultural Organization (FAO) AQUASTAT (https://www.fao.org/aquastat/en/databases/dams) and multiple regional inventories and registers, to form a single, congruent global database. The latest version v1.3 contains 7,320 georeferenced dams and their associated reservoir polygons when possible. Each reservoir feature is also provided with over 50 attributes such as reservoir name, storage capacity, and purpose. While the primary goal is to inventory all reservoirs with a storage capacity greater than 0.1 km^3, GRanD v1.3 includes 3,992 smaller reservoirs, leading to a total storage capacity of 6,881 km^3 in the entire database. The reservoirs also include 119 regulated natural lakes such as Lake Victoria and Lake Ontario. While a more exhaustive inclusion of smaller and/or newer reservoirs is important, we used GRanD v1.3 to flag some of the largest reservoirs and regulated lakes as an a priori attribute for the operational PLD. GRanD polygons were not used explicitly to construct the geometry of the PLD lake mask.

2.5. HydroBASINS

HydroBASINS (Lehner & Grill, 2013) offers a global tessellation of hierarchically nested basins and subbasins at various scales, derived primarily from the HydroSHEDS hydrography data set at a grid resolution of 15 arc seconds (∼500 m at the equator) (Lehner et al., 2008). Following the Pfafstetter coding system (Verdin & Verdin, 1999), the basin hierarchy in HydroBASINS is broken down to 12 nesting levels. They range from level 1 containing 9 continental or subcontinental boundaries, to level 12 encompassing about 1.0 million subbasins at a scale of only tens of square kilometers. In other words, basins of lower levels progressively encompass the subbasins of higher levels. For clarity, the subbasins corresponding to each level are organized as a different data layer. We used the data layers at Pfafstetter levels 3 in HydroBASINS v1.c, which contains 291 basin polygons together with their associated Pfafstetter codes at level 2 (corresponding to 62 larger basins) and level 1 (corresponding to 9 continental and subcontinental divisions). These level-3 basin boundaries and their Pfafstetter codes were used to help structure the prior lake identifier (*lake_id* attribute) and partition the PLD into level-2 basin granules (Section 3.1).

2.6. SWOT Orbit Swaths

The SWOT mission is split in two phases related to two different orbits (JPL internal document, 2022e). The initial Calibration/Validation (Cal/Val) phase, up to 11 July 2023, was related to a 1-day orbit at an 857 km of altitude: by frequent revisits of specific sites, this phase enabled the calibration of radar system parameters in the shortest time; it also allowed the study of rapidly changing phenomena. The science or nominal orbit, on which SWOT has been placed since 21 July 2023, is a non-sun-synchronous 21-day orbit, at an 890.6 km of altitude. Combined with the swath of the satellite, this orbit allows a quasi-global coverage up to 78° north and south latitude and, with its 10-day sub-cycle, is a good compromise for the temporal sampling as a region may be observed between once per cycle (at the Equator) to more than 10 times per cycle (at the highest latitudes).

Unlike nadir-pointing altimetry missions, which provide measurements directly below the satellite, SWOT KaRIn makes observations with a 120 km wide swath, from approximately 10 to 60 km of its nadir, on both “right” and “left” sides. The terms “left” and “right” are defined as if one stands on the Earth surface at the spacecraft nadir

facing in the direction of the spacecraft velocity vector. The 10–60 km width swath files (for the CalVal orbit: https://www.aviso.altimetry.fr/fileadmin/documents/missions/Swot/sph_calval_swath.zip; for the nominal orbit: https://www.aviso.altimetry.fr/fileadmin/documents/missions/Swot/swot_science_orbit_sept2015-v2_10s_swath.zip) were used to compute the theoretical coverage of each prior lake by SWOT KaRIn during a 1-day Cal/Val orbit cycle or a 21-day nominal orbit cycle. These files provide the full swath per pass, with a "pass" defined as a half revolution of the Earth by the satellite from pole to pole (south to north latitudes for ascending passes, and north to south latitudes for descending passes). There are 28 passes for the 1-day Cal/Val orbit, and 584 passes for the 21-day nominal orbit.

It is worth noting that SWOT KaRIn provides Low Rate (LR) observations with low spatial resolution globally (including both ocean and land) and HR observations for terrestrial and coastal environments. But due to bandwidth issues of downlinking the data, HR observations do not take place everywhere across the land surface. In general, HR data are available in most of the hydrologically active regions and extend a few kilometers off the coast. Areas excluded from HR observations are the Caspian Sea, parts of the desert or areic regions, and some high-latitude cryospheric environments, where lakes are absent or relatively scare. The exact HR coverage, however, varies seasonally. Given this variability, lake coverage and observability attributes were calculated based on full SWOT swaths, ensuring that at least LR observations, which are consistently available, were considered.

2.7. Databases for Lake Names

Multiple databases or open-source online repositories were jointly used to populate lake names for the prior lake polygons as thoroughly as possible. These sources include the IGN Carthage database (BD CARTHAGE®) to cover lakes in France (https://services.sandre.eaufrance.fr/telechargement/geo/ETH/BDCarthage/FXX/2017), OpenStreetMap (OSM; https://www.openstreetmap.org), the Global Lakes and Wetlands Database (GLWD) (Lehner & Döll, 2004), the Natural Earth Data (scale 1:30,000,000) (https://www.naturalearthdata.com), the Vector Map Level 0 (VMap0) (https://mdl.library.utoronto.ca/collections/geospatial-data/vector-map-level-0-vmap0), and HydroLAKES (v1.0) (Messager et al., 2016).

3. Database Development

3.1. Overview of the Operational PLD

Conceptually, the operational PLD is comprised of two primary components (Figure 1): (a) the prior lake mask, which inventories the polygon geometries of global lakes potentially visible to SWOT (≥1 ha); and (b) the operational auxiliaries, which facilitate the linkage of SWOT observations to the prior lakes and compile the necessary prior information to compute lake storage change and generate lake products. Analogously, the prior lake mask establishes the data infrastructure, while the operational auxiliaries support the SAS in implementing its functionality.

Structurally, the operational PLD is a relational database (Figure 2) which ties the central *lake* table to four auxiliary tables: *lake_catchment*, *lake_influence*, *basin*, and *hypso_curve*. Following the terminology of data science, here "table" refers to an arrangement of records that may contain fields of both geometry (e.g., raw polygons) and non-geometry (e.g., other numeric and text prior attributes). The central *lake* table consists of the polygons of the prior lake mask and a set of attributes for each prior lake. The prior lake mask is used to link SWOT-observed water features to the prior lakes by intersecting their geometries. The other attributes store prior information to calculate lake water storage change and to help populate the vector products at two granule levels, including the standard lake single pass vector product (L2_HR_LakeSP) in continental-pass granules (CNES internal document, 2022b) and the standard lake average cycle product (L2_HR_LakeAvg) in Pfafstetter level-2 basin granules (CNES internal document, 2022a). The *lake* table also contains an attribute to link the prior lakes and prior rivers (from SWORD), such that pixels of lakes connected to the prior river network, the so-called "connected lakes," are also included in lake data processing. For clarity, we refer to the non-geometric attributes of the *lake* table and the entirety of the other ancillary tables (*lake_catchment*, *lake_influence*, *basin*, and *hypso_curve*) as operational auxiliaries, which collectively supplement the prior lake mask to enable the expected functions of the operational PLD.

The *lake_catchment* and *lake_influence* tables contain ancillary geometries to accelerate the assignment of SWOT observations to the prior lakes. The issue is that the spatial linkage between SWOT-observed water

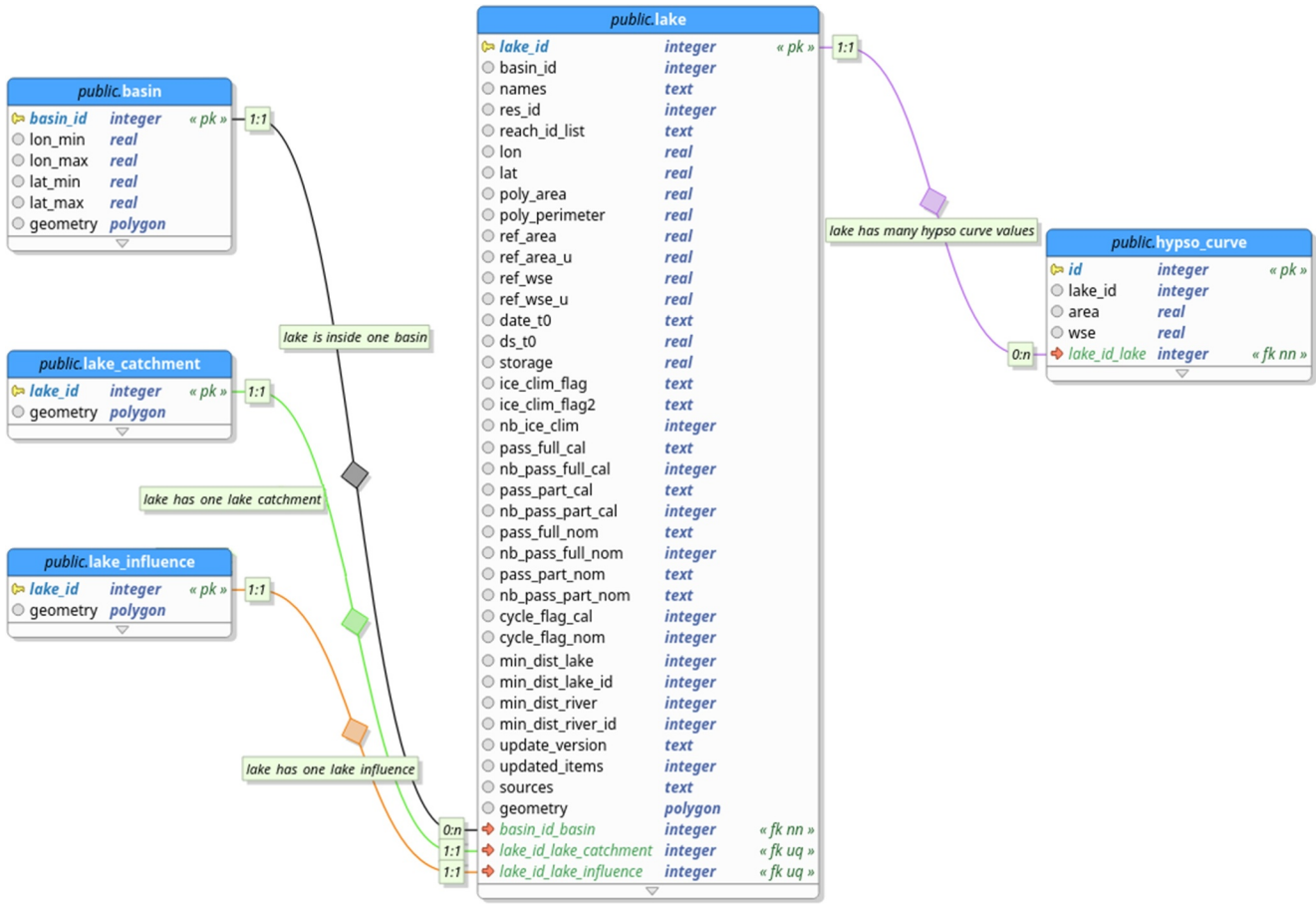

Figure 2. Structural model of the operational PLD. The diagram illustrates the relationships among the *lake*, *basin*, *lake_catchment*, *lake_influence*, and *hypso_curve* tables, along with the attributes within each table. Two-letter codes enclosed in double angle brackets denote: *pk* (primary key), *fk* (foreign key), *uq* (unique), and *nn* (not null). Cardinalities (*0*, *1*, *n*) define the theoretical association between tables and are indicated on both sides of each relationship. For example, in the relationship between the *lake* and *basin* tables, the "*0:n*" label next to the *lake* table indicates that each basin can be associated with anywhere from zero to *n* lakes (minimum *0*, maximum *n*). Conversely, the "*1:1*" label next to the *basin* table indicates that each lake is associated with exactly one basin (minimum *1*, maximum *1*).

features and the prior lakes does not always follow a one-to-one relationship. Particularly, complexities arise when an observed water feature is intersected by multiple prior lakes, leading to ambiguity regarding how the pixels in the observed water feature should be assigned to each of the prior lakes. To tackle the issue, these two assignment tables delineate a spatial partition for each prior lake stored in the *lake* table. Each lake assignment polygon defines the spatial domain within which the associated prior lake is allowed to "expand" before it infringes the domain of another prior lake. In other words, the lake assignment polygons disambiguate the vicinity of each prior lake so that lake assignment in complex spatial relationships can be eased (see examples in Section 4.5). The *lake_catchment* geometries provide a spatial partition that takes into account hydrological constraints and topography while *lake_influence* geometries take into account only distances between lakes (see more in Section 3.4).

The *hypso_curve* table stores the information of lake hypsometry for computing water storage changes, which will be added and periodically updated as SWOT data accumulate (see Section 5 for the versioning plan). This ancillary table will contain discrete WSE and water area points on the hypsometric curve (i.e., WSE-area relationship) for each prior lake. The curve will be fitted using the pairs of SWOT WSE and water area measurements collected from the first valid observation of each lake over a certain mission period. The hypsometric curve enables the calculation of water storage changes using an "incremental" approach (CNES internal document, 2023a; Crétaux et al., 2016), which takes into account the bathymetric variation between any two water levels. This method is more realistic than the "direct" approach, which employs an invariant bathymetric model (e.g., a trapezoidal prism or a truncated pyramid) between the levels.

The operational PLD is organized by HydroBASINS level-2 basins (Section 2.5), which results in 61 valid basin-granule PLD files. The *lake* table in each basin-granule PLD includes only the prior lakes intersected by the associated level-2 basin. The *lake_catchment* and *lake_influence* tables include the catchment and influence polygons intersecting this level-2 basin, respectively. The *basin* table delineates the full boundaries of HydroBASINS level-3 basins nested within this level-2 basin (Section 2.5), together with the associated basin Pfafstetter codes. This table is used to label the observed water features in different continents and basins. More details on the development of each table, except *hypso_curve*, are given in the following subsections.

3.2. Prior Lake Mask

As previously described (Section 2.1), the Circa-2015 lake data set (Sheng et al., 2016) was used as the primary source of the prior lake mask. To improve the representation of reservoirs, particularly those constructed after 2015, additional state-of-the-art global and regional reservoir databases (Section 2.2) were integrated to the Circa-2015 data set to form the final prior lake mask.

The reservoirs in GeoDAR v1.1 (Wang et al., 2022a) and GREI-p2k (Fan et al., 2024) databases that are not intersected by any Circa-2015 polygon were first added successively to the prior lake mask. The remaining GREI-p2k reservoirs were next investigated based on their spatial relationship with the updated prior lakes. High-resolution Esri and Google Earth imagery were also employed to assist in visual inspection. When a prior lake spatially conflicted with more than one GREI-p2k reservoir, we examined whether this prior lake overshoots the dam location and mistakenly spans multiple reservoirs. If verified, this prior polygon was manually split into multiple reservoirs. When a prior lake intersected with only one GREI-p2k reservoir, we examined whether this reservoir was substantially overrun by its intersecting prior lake. This possibility was identified when the GREI-p2k reservoir is well included (>75%) by the prior polygon but the latter is less well covered (<75%) by the former. We then visually inspected if this prior polygon mistakenly annexed the reservoir depicted by GREI-p2k; if verified, this prior polygon was truncated, allowing the GREI-p2k polygon to be added as a new reservoir without topological conflicts. For the rest of the cases (one GREI-p2k reservoir intersected by one or multiple prior lakes), we classified them based on the spatial agreement between the two data sources. If their overlapping area covered at least 50% of the lake area in both sources, we considered the GREI-p2k reservoir and its intersecting prior lake(s) to be in good agreement and thus excluded them from further investigation. Otherwise, the case was visually inspected, and when necessary, the prior polygon was split or replaced by the intersecting GREI-p2k reservoir.

The improved prior lakes were next compared with the remaining GeoDAR reservoirs. The procedure was overall similar to the one for GREI-p2k, except that we employed a more qualitative approach in comparing GeoDAR and prior polygons, and the comparison was focused on large reservoirs only. This was because many small and medium-sized reservoir polygons in GeoDAR are already sourced from the Circa-2015 lake data set, and the other polygons sourced from HydroLAKES and GRanD usually exhibit coarser shorelines (Wang et al., 2022a). Nevertheless, when a GeoDAR polygon showed clear superiority in representing the reservoir integrity (e.g., with improved shoreline connectivity and reduced surface water patchiness), the GeoDAR polygon was used to replace the intersecting prior lake(s). In occasional cases where a single source was not a sufficient solution, we performed manual digitization to modify and merge multiple sources. The data sources and harmonizing methods were reflected in the attribute *source* of the prior lake mask (Table 2).

Additional regional improvements were further made on the prior lake mask after the integration of global reservoir databases. In particular, we included nearly 7,000 reservoirs in the Crateús and Banabuiú basins of Brazil to refine the completeness and accuracy of reservoir mapping in this hotspot region. These Brazilian reservoirs were mapped from Landsat surface reflectance images using water index spectral thresholds as in Fisher et al. (2016) to represent the interannual water area maximum during 2008–2019. We also improved the mapping of several critical lakes in semi-arid western Africa, which are typically covered by aquatic vegetation and difficult to delineate using global algorithms. These lakes were extracted following a supervised classification of Sentinel-2 images using the Active Learning for Cloud Detection (ALCD) algorithm (Papa et al., 2023) and/or spectral thresholding of the modified NDWI (MNDWI) with an ad hoc threshold for each lake (de Fleury et al., 2023). Other supplementary data sources, such as the Circa-2000 lake map (Sheng et al., 2016), GLAKES (Pi et al., 2022), GSW (Pekel et al., 2016), HydroLAKES (Messager et al., 2016), and OSM, were used to incorporate another several hundred lakes and reservoirs, collectively contributing to less than 0.5% of the global

Table 2
Key Attributes in the Operational PLD Tables

Attribute	Description
***Lake* table**	
lake_id	A ten-digit integer identifier (ID) for each prior lake, structured as follows: a three-digit basin ID (matching *basin_id*), a six-digit ordinal index specific to the lake, and a one-digit code representing the lake's type.
basin_id	A three-digit integer ID for the HydroBASINS Pfafstetter level-3 basin containing the prior lake. This ID matches the first three digits of *lake_id*.
names	Known name(s) of the prior lake. If one lake has multiple names, they are separated by semicolons.
res_id	The reservoir ID from the Global Reservoir and Dam database (GRanD v1.3), if the prior lake intersects a GRanD reservoir.
*reach_id_list**	A list of IDs of the SWORD reaches intersecting the prior lake (i.e., type 3 reaches). If multiple reaches exist, their IDs are separated by semicolons.
lat/lon	The latitude and longitude (in decimal degree) of the centroid of the prior lake.
poly_[area/perimeter]	The area (in km^2) and perimeter (in km) of the prior lake polygon.
ref_area/ref_area_u**	An intermediate reference area of the prior lake and its associated uncertainty (in m^2), corresponding to the same water state as *ref_wse*.
ref_wse/ref_wse_u**	An intermediate reference water surface elevation (WSE) of the prior lake, along with its associated uncertainty (in m). The WSE value is referenced to the EGM 2008 geoid model.
*date_t0**	The initial reference date and time (in Coordinated Universal Time, UTC) from which water storage changes are calculated for SWOT Level 2 lake vector data products. This corresponds to the first valid SWOT observation of each prior lake.
*ds_t0**	The storage change (in m^3) between the initial reference WSE at *date_t0* and the intermediate reference WSE (*ref_wse*) in the prior lake.
*storage**	The maximum variation in water storage (in m^3) for each prior lake, defined as the difference in storage corresponding to the highest and lowest valid WSEs detected by SWOT.
ice_clim_[flag/flag2]	An integer flag characterizing ice presence on the lake during each day of the calendar year based on climatological data: 0: Never frozen 1: Can be frozen 2: Always frozen *ice_clim_flag* is a text sequence representing ice flags from 1 January to 30 June, and *ice_clim_flag2* covers the period from 1 July to 31 December.
nb_ice_clim	The number of days per calendar year the prior lake is ice-covered (*ice_clim_[flag/flag2]* = 1 or 2) based on climatological data.
pass_[full/part]_[cal/nom]	A list of IDs of the SWOT swath passes (separated by semicolons) that fully or partially cover the prior lake during a calibration or nominal orbit cycle.
nb_pass_[full/part]_[cal/nom]	The count of SWOT pass IDs corresponding to *pass_[full/part]_[cal/nom]*.
cycle_flag_[cal/nom]	An integer flag characterizing the KaRIn observation scenario for each prior lake during a calibration or nominal orbit cycle: 0: Never observed 1: Only partially observed 2: Fully observed after aggregating partial observations in multiple passes 3: Fully observed by a single pass at least once
min_dist_[lake/river]	The geodesic distance (in m) from the prior lake to its nearest prior lake or river. This attribute is used as a criterion for PLD updates (Section 5).
min_dist_[lake/river]_id	*lake_id or reach_id* corresponding to *min_dist_[lake/river]*.
update_version	The version of the last update for each prior lake. Null if no update was made.

Table 2
Continued

Attribute	Description
update_items	A five-digit integer flag indicating what has been updated since the last version (see Table S1 in Supporting Information S1 for details).
sources	The data source(s) of each prior lake polygon.
***Lake_catchment* table**	
lake_id	The ID of each prior lake encompassed by the lake catchment polygon.
***Lake_influence* table**	
lake_id	The ID of each prior lake encompassed by the lake influence polygon.
***Hypso_curve* table**	
id	The ID for each WSE-area pair
lake_id	The ID of the prior lake associated with the WSE-area pair
wse/area**	The WSE (in m) and water area (in m^2) values for each discrete point on the hypsometric curve of the prior lake. The curve is fitted using available SWOT measurements of lake WSE and water area starting from the first valid observation.
***Basin* table**	
basin_id	The ID of the HydroBASINS Pfafstetter level-3 basin, matching the first three digits of *lake_id* for the lakes within the basin.
lat_[min/max]	The minimum and maximum latitudes (in decimal degree) of the basin boundary
lon_[min/max]	The minimum and maximum longitudes (in decimal degree) of the basin boundary

Note. Attributes marked with "*" are expected to be populated and updated as SWOT data accumulate and improve (Section 5).

lake area. With the aid of manual digitization and editing, these refinements added previously omitted lakes, such as those on small islands, and improved lake area integrity by reducing patchiness and enhancing completeness. Additional editing was also performed to separate other identified polygons that represent distinct lakes divided by waterfalls, dams, and weirs. The input data for each prior lake are documented in the *sources* attribute (Table 2).

While many dams have been constructed since 2015, there are also reservoirs recently removed or decommissioned. To remove these reservoirs from the prior lake mask, we referred to multiple dam registers and inventories that contain project status and timeline information, including the ICOLD WRD, GRanD, and the AQUASTAT Geo-referenced Database of Dams (https://www.fao.org/aquastat/en/databases/dams), to identify the locations of decommissioned, removed, or subsumed dams as thoroughly as possible. We then visually inspected each of the identified locations against the latest high-resolution Esri and Google Earth images to ensure the removal of any prior reservoir polygon if water impoundment is no more verifiable. If a dam was subsumed by a new one with a larger/elevated reservoir area, we modified the prior polygon through careful digitization so that it represents the boundary of the expanded reservoir. In an extraordinary case, the 66-year-old Kakhovka Dam in Ukraine was destroyed on 6 June 2023, draining its reservoir that previously inundated more than 2,100 km^2 along the Dnieper River (Naddaf, 2023). Later Landsat images confirmed that the dam remains breached, and the exposed lakebed contains a large quantity of seasonal ponds, wetlands, channels, and floodwater. Given this complexity, lumping their water storage changes into one prior polygon is what we perceived to be the best strategy for now. In addition, the PLD has the potential of being applied for other satellite missions to reconstruct historical water records, and there is also a possibility of rebuilding the dam after the war (Stone, 2024). After considering these factors, we decided to keep the prior polygon for the Kakhovka Reservoir in the current version of the PLD.

Finally, the updated prior lake mask was post-processed to connect polygons meeting at a common vertex by introducing a small connecting channel at the vertex. Meanwhile, polygons sharing a common border were separated by creating a thin sliver along the border. This post-processing slightly reduced the number of original prior polygons, but it improved the connectivity of lake surface area and eliminated topological conflicts.

Table 3
PLD Lake Abundance for Each Pfafstetter Level-1 Continental Divisions

Continent code (ID)	Continent name	Lake count	Total lake area (km^2)	Mean/median size (ha)	Lake density
1 (AF)	Africa	73,781 (1.25%)	243,728.5 (9.29%)	330.3/3.2	0.8%
2 (EU)	Europe and Middle East	504,435 (8.55%)	301,419.3 (11.49%)	59.8/3.7	1.7%
3 (SI)	Siberia	1,113,265 (18.88%)	289,579.0 (11.04%)	26.0/4.0	2.2%
4 (AS)	Central and Southeast Asia	446,773 (7.58%)	234,546.1 (8.94%)	52.5/2.5	1.1%
5 (AU)	Australia and Oceania	57,338 (0.97%)	73,324.7 (2.79%)	127.9/3.8	0.7%
6 (SA)	South America	250,215 (4.24%)	164,364.8 (6.26%)	65.7/3.6	0.9%
7 (NA)	North America and Caribbean	1,512,137 (25.64%)	826,763.1 (31.51%)	54.7/3.9	5.2%
8 (AR)	North American Arctic	1,899,238 (32.20%)	481,354.8 (18.34%)	25.3/3.5	7.8%
9 (GR)	Greenland	40,759 (0.69%)	8,948.3 (0.34%)	22.0/3.3	0.4%
Global	–	5,897,941 (100%)	2,624,028.6 (100%)	44.5/3.6	1.9%

Note. Lake geometric properties in this study were computed using geodesic methods based on the World Geodetic System 1984 (WGS84) datum.

3.3. Attributes in the *Lake* Table

The attributes in the *lake* table (Table 2) provide multi-theme information for each prior lake polygon, which covers basic lake identities, relationship with drainage basins and prior rivers, reference WSE and water area for deriving lake storage change, and SWOT overpass statistics to enable data processing and product distribution. Accompanying the attribute definitions in Table 2, we provide additional details that are necessary for understanding the attribute format, purpose, and populating method.

3.3.1. Lake Identities

The primary key, *lake_id*, is a ten-digit integer in the format CBBNNNNNNT, where C is a one-digit continent code (Table 3), BB a two-digit basin code, NNNNNN a zero-padded, six-digit sequence indicating the lake's ordinal index within its basin, and T a one-digit code denoting the water body type (Table 4). The first three digits (CBB) are based on the Pfafstetter coding system used in HydroBASINS. The continental code (C) corresponds to level-1 divisions (Table 3), and BB concatenates the codes of level-2 and level-3 basins representing increasing drainage details. This hierarchy organizes the global prior lakes to 291 subbasins at the scale of Pfafstetter level 3 (see Section 3.5), and the assignment was based on geometric intersection with HydroBASINS boundaries. In cases where a lake intersects two or more basins (e.g., due to imprecise basin boundaries), it was assigned to the basin containing its centroid. Similarly, if a lake is not located within any basin, it was assigned to the nearest basin. Following the Pfafstetter coding system, prior lakes for each level-3 basin are then indexed from 000001 to a maximum of 999999 based on a random order.

The last digit in *lake_id* classifies global prior lakes to two water body types based on their geometric connectivity with prior rivers (Table 4). The same water body type codes are also used for the primary key in SWORD (Altenau et al., 2021). As listed in Table 4, each prior lake was categorized to either a connected lake ($T = 3$) or a disconnected lake ($T = 2$). A connected lake is defined as any prior lake polygon intersected by one or more prior reaches and is included in both river and lake data processing. It is worth noting that this connectivity was determined specifically in relation to SWORD. This means a "disconnected" prior lake may still be hydrologically connected to a river, but the river may be too narrow to be detected by SWOT and is therefore not inventoried in SWORD.

Table 4
Water Body Type Codes in the Prior Lake and River IDs

Type code (T)	Water body type
1	River (only applicable to SWORD)
2	Disconnected lake
3	Connected lake
4	Dam (only applicable to SWORD)
5	No topology (only applicable to SWORD)

Figure 3 illustrates an example to help interpret the hierarchy of *lake_id*. This example covers the Pfafstetter coding system in the North American continent ($C = 7$), which encompasses eight level-2 basins. One of them (first $B = 4$) contains the Mississippi River Basin (second $B = 2$) at level 3. There are 109,835 prior lakes in the Mississippi River Basin, which all share "742" as the first three digits in *lake_id*. Among them is an example lake

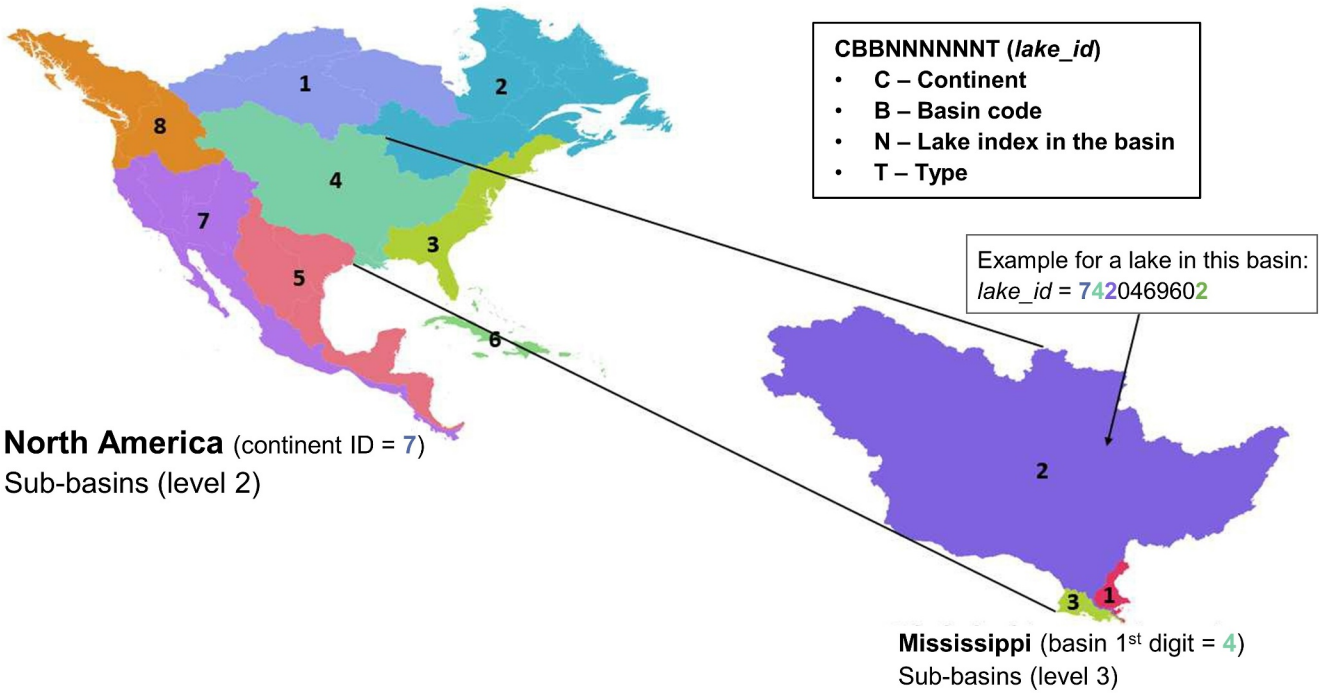

Figure 3. Hierarchical structure of the 10-digit *lake_id* for prior lakes. The example is given for a disconnected prior lake ($T = 2$) in a Pfafstetter level-3 basin (BB = 42, the Mississippi River Basin) of the North American continent ($C = 7$).

"7420469602," indicating that this lake is indexed to be the 46960th in the basin (NNNNNN = 046960) and is disconnected from any prior rivers in SWORD ($T = 2$).

The *names* attribute inventories the known names of global prior lakes as thoroughly as possible for the convenience of PLD and SWOT science data users. The lake names were populated through a "spatial join" from multiple open-source atlases and databases, including the IGN Carthage database for France, OSM, GLWD, the Natural Earth Data, VMap0, and HydroLAKES (v1.0) (Section 2.7; Table 1). All names are in capital letters to avoid accents and other spelling discrepancies. The same name can be shared by several prior polygons if they are disconnected portions of the same lake due to either mapping issues or seasonal variation. From this aspect, the *names* attribute is potentially useful for dissolving patchy water bodies within known lakes to improve the integrity of their prior extents and the completeness of storage change estimates. About 152,240 lake names were assigned to ~329,360 prior lake polygons, accounting for ~6% of the global lakes by count and ~62% by area.

The prior lake polygons include both natural lakes and artificial reservoirs. While classifying lake typology is not the priority of the operational PLD, the *lake* table does provide a *res_id* attribute, which flags about 7,300 large reservoirs using the IDs of GRanD v1.3. These IDs were populated by intersecting the prior lakes with GRanD reservoir polygons. If a prior lake intersects more than one reservoir, only the ID of the GRanD reservoir containing the prior lake centroid was used. Although GRanD focuses on the world's largest reservoirs (e.g., with storage capacity exceeding 0.1 km^3), this flag allows for a preliminary attribution of SWOT-measured water storage changes to either climate or human regulation. More comprehensive information about reservoirs and other lake types is available in the forthcoming "scientific PLD."

3.3.2. Relations With SWOT-Visible Rivers

The *reach_id_list* attribute identifies each river-connected prior lake by the IDs of the intersecting SWORD reaches. Similar to the structure of the prior lake ID, the ID of each prior reach (*reach_id*) is a 11-digit integer in the format CBBBBBRRRRT, with the last digit T being a type code consistent with that in *lake_id* (more details on *reach_id* are given by Altenau et al. (2021)). In other words, all prior reaches listed in the *reach_id_list* attribute of the *lake* table share the same last digit (T) of 3 indicating "lake on river", and similarly, all prior lakes with a valid *reach_id_list* value also have a last digit of 3, indicating they are river-connected lakes (Table 4). In

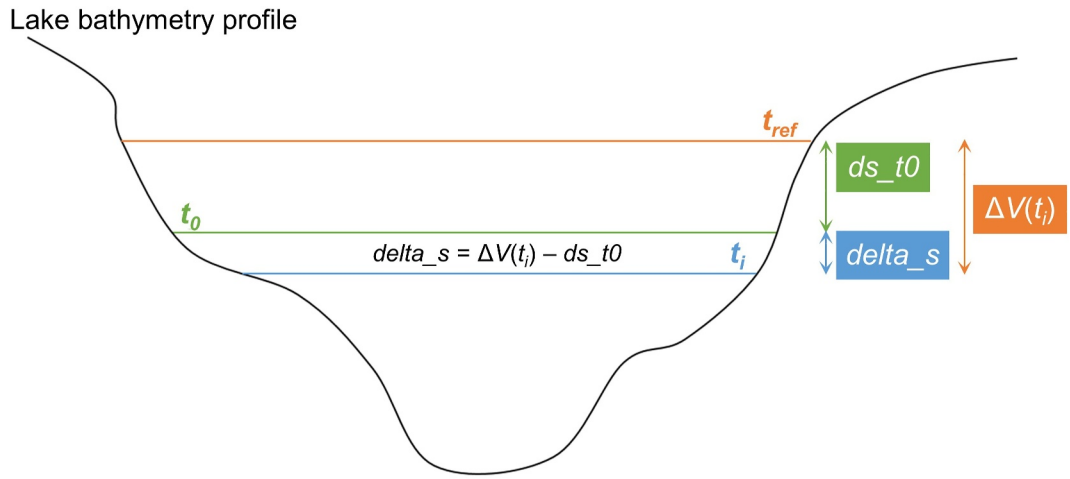

Figure 4. Illustration of different water level states used in lake storage change calculations.

SWORD v17 used for identifying river-connected lakes, prior reaches were segmented to comply with prior lake boundaries, so any type 3 prior reach and its nodes are, in principle, spatially contained by the associated type 3 prior lake.

For each of the prior reaches in *reach_id_list*, SWOT-detected water pixels are considered to be part of both a river and a lake (such as a reservoir). These water pixels are kept for both lake and river data processing whereas pixels on the other reaches are eliminated from further lake processing. The *reach_id_list* attribute also facilitates a potential synergy of SWOT lake and river data products. One example is the LakeFlow algorithm (Riggs et al., 2023), which uses both products and the concept of lake-river mass conservation to improve the estimates of lake inflow and outflow. The *reach_id_list* attribute identified 47,022 prior river reaches connected to 17,894 prior lakes, accounting for 42.3% of the global lake area. More advanced information on lake drainage topology and lake-river connectivity will be available in the forthcoming "scientific PLD."

3.3.3. Prior Information for Computing Lake Storage Change

An essential function of the operational PLD is to assist the SAS in turning repeated lake area and WSE observations from SWOT to lake water storage variation. For this purpose, the *lake* table includes several attributes associated with reference water states for each prior lake, based on which water storage changes (i.e., the output variable *delta_s* in the lake vector products) can be computed. These attributes, hereafter "storage reference attributes," start with *date_t0*, which defines the date and time (in Coordinated Universal Time, UTC) of the first valid SWOT observation of each prior lake. The WSE and water area at *date_t0* set up the initial reference state for computing *delta_s*. In other words, even though lake storage algorithms in the SAS vary in bathymetrical model (linear or quadratic) and integration approach (direct or incremental), the output *delta_s* conceptually always represents the storage change from the observed state (i.e., WSE and water area at a given time t_i) to the initial reference state defined by *date_t0* (see Figure 4).

For practical reasons in hypsometric modeling, the calculation is first performed for the lake storage change ($\Delta V(t_i)$) between t_i and an intermediate reference state defined by the *ref_wse* and *ref_area* attributes. The *ref_wse* attribute quantifies the median WSE of the prior lake during a given SWOT observation period, and *ref_area* stores the inundation area corresponding to *ref_wse*. Their associated uncertainties are given in *ref_area_u* and *ref_wse_u*, which are needed for propagating storage change errors. The storage difference between these two reference states (i.e., the intermediate state and the initial state at *date_t0*) is provided in the *ds_t0* attribute. This way, *delta_s* can be derived by subtracting *ds_t0* from $\Delta V(t_i)$ (Figure 4). Technical details on lake storage change algorithm and error propagation are beyond the scope of this paper but are available in the Algorithm Theoretical Basis Document (CNES internal document, 2023a).

It is important to note that the reference state at *date_t0* does not necessarily correspond to the minimum level of the prior lake. However, the *lake* table provides another attribute *storage*, which aims to quantify the storage

change between the maximum and minimum WSEs observed for each lake. This attribute estimates the magnitude of possible storage variation per prior lake during SWOT's mission lifetime, which can help assess the scales of any observed water storage change relative to the maximum storage variation. So far, we have applied SWOT HR data up to November 22, 2024, to populate several storage reference attributes. Specifically, we used the latest release (Version 2/C) of the L2_HR_LakeSP_Prior product (CNES internal document, 2022b; accessible at https://podaac.jpl.nasa.gov/dataset/SWOT_L2_HR_LakeSP_2.0), which provides single-pass WSE and area measurements (in the *wse* and *area_total* attributes, respectively) for each prior lake when feasible (Section 4.5). Outlier WSE values for each prior lake were filtered using a 2-sigma method, and the earliest acquisition time after data filtering was assigned to *date_t0*. The *ref_wse* attribute was then calculated as the median WSE from the filtered data. Since the quality of *area_total* is still being refined, *ref_area* was temporarily populated as the area of the prior lake polygon. The *ds_t0* attribute was computed as the storage change between the initial reference state at *date_t0* and the intermediate reference state (*ref_wse*, *ref_area*). These attributes have been populated for more than 5.3 million prior lakes. The *storage* attribute will be calculated as SWOT data accumulate and improve. For other prior lakes, *ref_area* was assigned by the lake polygon area while other attributes were marked as "null." Updates to these storage reference attributes will be implemented according to the PLD versioning plans described in Section 5.

3.3.4. SWOT Overpasses and Lake Coverage

Lastly, the *lake* table contains a few more attributes that describe SWOT's theoretical data coverage of each prior lake in relation to orbit passes. These attributes inform how well each prior lake can be observed by KaRIn under a single swath pass or after aggregating multiple passes during a calibration (*cal*) or nominal (*nom*) orbit cycle. The *pass_full* and *pass_part* attributes list the IDs of the passes covering each prior lake fully and partially, respectively. Their values were configured by intersecting the prior lakes and the orbit passes with swaths covering 10–60 km from nadir (Section 2.6). These two attributes were then used to quantify how many times each lake can be observed completely (*nb_pass_full*), partially (*nb_pass_part*), or both during an orbit cycle (see Section 4.4). Using this information, the *cycle_flag* attribute further summarizes SWOT's lake coverage into four scenarios. Scenario "0" flags the prior lakes that may never be observed by SWOT. This was determined by the lakes where both *pass_full* and *pass_part* values are empty. Scenario "1" indicates that the lake is only partially observed even after aggregating all passes over a cycle, and scenario "2" indicates that the lake can be fully observed by SWOT, but only after pass aggregation over a cycle. In both scenarios, *pass_part* has valid pass IDs while *pass_full* is empty. Finally, scenario "3" flags all prior lakes that are fully observed by at least one single pass. This was determined by the prior lakes where *pass_full* has valid pass IDs regardless of *pass_part*.

3.3.5. Lake Ice Flag

The goal of *ice_clim_flag* (climatological ice flag) is to help the data user make decisions on removing potentially ice-affected SWOT lake products, and to allow the SAS to calculate the *ice_clim_f* attribute (i.e., a climatological flag indicating whether the lake is ice-covered on the day of the observation based on *ice_clim_flag*) in the lake vector product (CNES internal document, 2022b). Climatological ice flags are estimated ice conditions for a typical year, averaging ice conditions between 1 January 2010 and 1 January 2020. Here we briefly describe the two steps taken to develop the lake ice flag.

Development of a lake ice fraction empirical model. To develop a priori ice conditions for all prior lakes, we applied an empirical lake ice fraction model by matching same-day ice fractional data derived from Landsat 5, 7, and 8 images, whenever cloud-free conditions were observed, with daily surface air temperature from ERA5 climate reanalysis data (Copernicus Climate Change Service, 2017). The lake ice fraction was calculated based on the lake ice detection algorithm (SLIDE) (Yang, Pavelsky, et al., 2022) for each prior lake polygon. By modeling the lake ice fraction with daily-mean air temperature, we identified the following logistic regression:

$$\log(\text{odds}(P_{\text{ice}})) = -0.46 \cdot \text{SAT}_{30} - 0.02 \cdot \text{SAT}_{30} \cdot \text{Period} + 0.85 \quad (1)$$

where P_{ice} denotes the lake ice area fraction; SAT_{30} denotes the prior 30-day mean surface air temperature; and Period, a categorical variable, denotes whether the calculation was carried out during the breakup months (Period = 1 when Julian day is between [70, 227]; Period = 0 otherwise). Adding the variable Period allowed the

model to accommodate the difference in ice dynamics during the breakup and freeze-up, a difference that has been previously identified in other types of freshwater bodies (Lacroix et al., 2005).

Estimating lake ice flag. For each point geometry representing the prior lake centroid, and for each day during the period between 1 January 2010 and 1 January 2020, we estimated the ten-year mean lake ice fraction by inputting daily mean surface air temperature from ERA5 reanalysis database (variable: mean_2m_air_temperature) to the empirical lake ice model above. Then, a climatological mean lake ice fraction was estimated by averaging lake ice fraction across the 10 years for each Julian day. Finally, the continuous ice fraction was converted to three discrete integer values to represent ice conditions for SWOT ice flag: mean ice fraction <0.2: 0; 0.2 ≤ mean ice probability <0.8: 1; and mean ice probability ≥0.8: 2.

This flag can suggest likely ice cover conditions at the given time of year for a given prior lake based on modeled historical ice conditions. However, factors such as interannual variability in ice phenology, thaw events during cold seasons, and non-stationarity in climate mean that users are encouraged to seek ice conditions that are more recent and locally relevant whenever those sources are available. When no other sources are available, the climatological flag provides a reasonable expectation of the average ice condition.

3.4. *Lake_catchment* and *Lake_influence* Tables

The *lake_catchment* and *lake_influence* tables store two types of assignment polygons/domains for each prior lake referenced in the *lake* table. By definition, a lake assignment polygon should encompass the associated prior lake along with its water fluctuation zone, and meanwhile, the assignment polygons of different prior lakes should not overlap. This way, when it is unclear how a SWOT-detected water region should be assigned to the prior lakes using the prior lake geometry alone, the assignment polygons can help determine the rule for executing lake assignment (see Sections 3.1 and 4.5). In addition to the geometries, each assignment polygon is also indexed by the ID of the encompassed prior lake, *lake_id* (Section 3.3.1), which links the *lake_catchment* and *lake_influence* tables to the *lake* table.

We considered two rationales for constructing lake assignment polygons. The first rationale follows the concept of lake hydrological catchment, which defines the sub-basin between the outlets of a prior lake and its immediate upstream prior lake(s). If a prior lake is in the headwater (meaning no lakes further upstream), the catchment is then the entire watershed upstream to the outlet of this lake. Since water dynamics in a lake are confined by its own catchment boundary, this rationale complies with the ideal definition of lake assignment polygons described above. To implement this rationale, we applied the algorithm recently developed for the global Lake Topology and Catchment (Lake-TopoCat) database (Sikder et al., 2023) on the prior lake mask using the 90-m-resolution MERIT-Hydro hydrography data (Yamazaki et al., 2019). The algorithm generated fine-detailed catchments for each of the prior lakes, which compose the geometries of the *lake_catchment* table. A regional example is given for part of western Africa in Figure 5a.

The second rationale relies on geometric vicinity. Specifically, we employed the Voronoi tessellation (Aurenhammer, 1991) to partition the continental surface into proximal regions based on the geodesic distance to the prior lakes, and the resultant regions, also known as Voronoi cells or Thiessen polygons, are the geometries for the *lake_influence* table (see the example of Figure 5b). Mathematically, the Voronoi tessellation decomposes a plane with a finite number of objects, or the so-called "seeds," into the same number of Thiessen polygons. Each Thiessen polygon corresponds to one seed object, for example, a prior lake in our case, and every virtual point within this polygon is closer to its seed prior lake than to any other prior lake. Because of this proximal characteristic, Thiessen polygons are often regarded as the "areas of influence" in computational geometry and have been widely applied in hydrology, meteorology, and geo-statistics, such as interpolating rainfall from gauge station measurements (Evans & Jones, 1987). Although these influence features do not follow the exact lake catchment boundaries (Figure 5), it is important to note that assignment polygons are not required for every case of lake assignment. When they are indeed needed, the Thiessen polygons provide a computationally efficient alternative to ease the linkage of SWOT observations to the prior lakes. An example of when lake assignment polygons are required and how they function to ease lake linkage is given in Section 4.5.

While lake catchment features carry a clearer physical meaning than the Thiessen-based lake influence features, we opted to provide both assignment tables in this version of the PLD. The latest release of the SWOT lake vector data product (version C) was produced based on the *lake_influence* table, but in recognition of the merits of lake

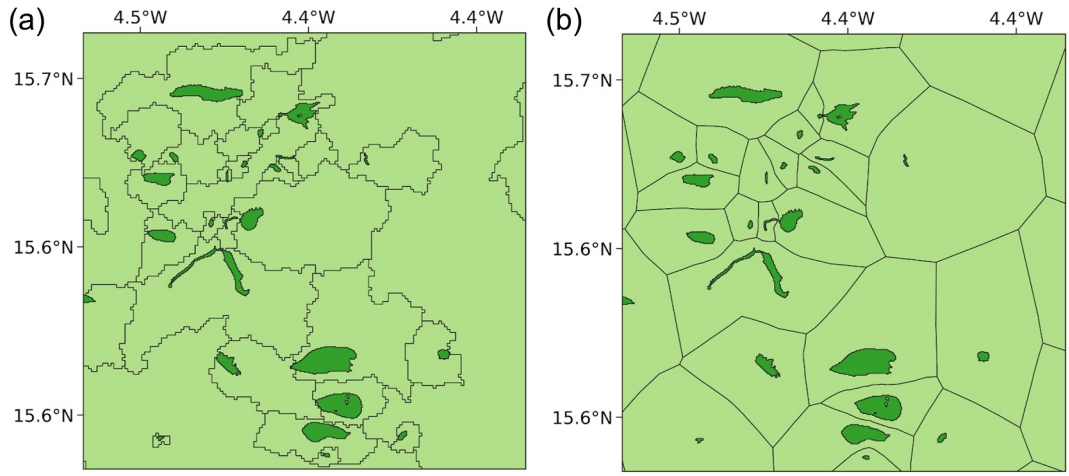

Figure 5. An example of SWOT prior lakes in part of western Africa (deep green) and their associated assignment domains (light green). (a) Lake hydrological catchments as in the *lake_catchment* table. (b) Lake influence features as in the *lake_influence* table.

catchments, the forthcoming releases of the lake vector lake product may be processed using the *lake_catchment* table.

3.5. *Basin* Table

The *basin* table contains the geometries of Pfafstetter level-3 basins corresponding to each level-2 basin granule (see Section 3.1 for PLD organization). The basin boundaries were retrieved from the HydroBASINS database, with a total number of 291 level-3 basins on the global continents except Antarctica. Each basin feature in this table is provided with five attributes as listed in Table 2. The *basin_id* attribute is the basin identifier, containing the level-3 Pfafstetter code from HydroBASINS. The value of this attribute matches the first three digits in *lake_id* (i.e., CBB) of the *lake* table, which links each prior lake to its associated basin. The basin geometries and *basin_id* values are used to separate the water features observed by SWOT, including those not intersected by any prior lakes, to different continents and basins, which is needed for populating the lake vector products at different granule scales.

4. Results and Discussion

4.1. Prior Lake Abundance and Distribution

As the primary component of the SWOT PLD, the prior lake mask contains 5,897,941 polygons (Figure 6), mostly representing the intermediate water extents of global lakes during their stable seasons. While our goal was to inventory lakes with a minimum size of 1 ha, the prior mask also included 42 lakes smaller than 1 ha, sourced from OSM and user requests. The global prior lakes have a total area of 2,624,028.6 km^2, covering about 2% of the Earth's land surface excluding Antarctica. The Caspian Sea was excluded from the PLD due to its large size and dual characteristics of both lake and ocean (Zimnitskaya & Geldern, 2011).

Table 3 summarizes lake abundance in each of the nine Pfafstetter-1 continental divisions. Lake counts range from fewer than 80,000 per division in Africa (AF), Australia and Oceania (AU), and Greenland (GR) to more than 1 million in Siberia (SI), North America and Caribbean (NA), and North American Arctic (AR). In general, divisions with a higher number of lakes also tend to exhibit a greater total lake area and higher lake density. Despite a global average of 1.9%, lake density varies substantially from only 0.4%–0.8% in GR, AU, and AF, to 5.2% in NA and as high as 7.8% in AR. At a continental scale, lake abundance is negatively correlated to aridity and positively correlated to the degree of glaciation or periglacial processes (except GR). Divisions with less lake abundance, however, tend to have a greater mean lake size (e.g., 127.9 ha in AU and 330.3 ha in AF), implying fewer but larger lakes are more likely to develop in arid regions. On the other hand, the lake-dense circum-Arctic regions (AR and SI) are dominated by smaller lakes with an average size of 25–26 ha, substantially below the

10.1029/2023WR036896

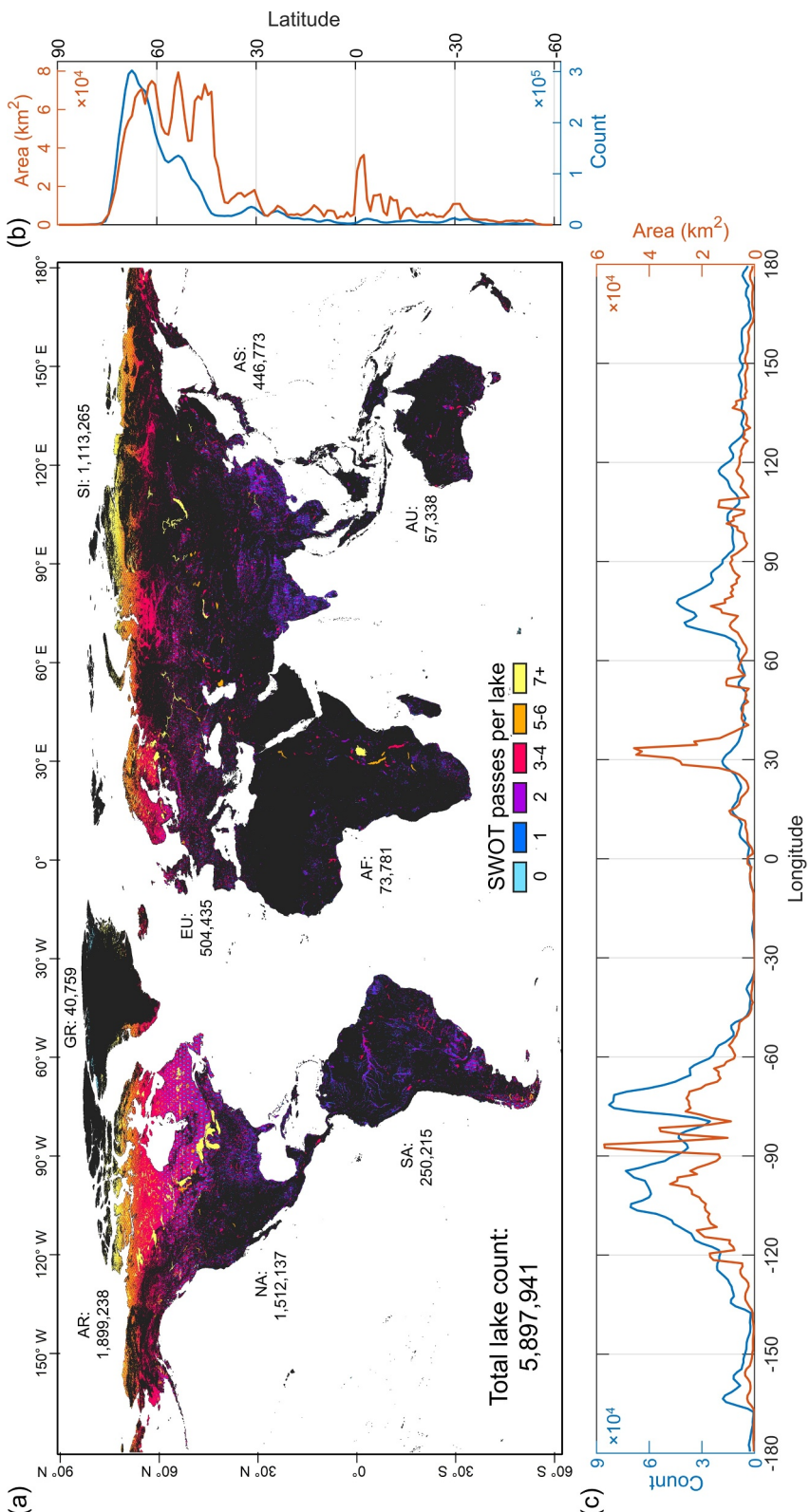

Figure 6. Global map and distribution of the SWOT prior lakes. (a) Global map of prior lakes, with numbers labeling the count of lake polygons per Pfafstetter level-1 division and colors displaying the number of SWOT overpasses per lake during each 21-day orbit cycle. (b) Count and total area of the PLD lakes per latitudinal degree. (c) Count and total area of the PLD lakes per longitudinal degree. The location of each lake polygon was determined by the latitude and longitude coordinates of the centroid of the lake polygon. Values in both latitudinal and longitudinal histograms (b and c) were smoothed by a 3-degree average window to enhance esthetic appearance and take into account that some lakes can span multiple 1-degree intervals.

global average 44.5 ha. In comparison, the median lake sizes are more consistent among the continents and range subtly between 2.5 and 4.0 ha.

With a targeted minimal lake size of 1 ha, the prior mask reveals an unprecedented detail of global lake distribution. About 65% of the total lake count or nearly 40% of the total lake area is clustered in the sparsely populated high-latitude regions above 55°N (Figure 6b), where glacial activities prevailed in the last ice age. Lakes are particularly ubiquitous across the Canadian Shield and Scandinavia as a result of glacial erosions during the Pleistocene (Shilts et al., 1987; Smith et al., 2007) and the boreal permafrost lowlands (e.g., in Siberia and Alaska) associated with thermokarst (Kokelj & Jorgenson, 2013; Manasypov et al., 2014; Smith et al., 2005; Wik et al., 2016). While lake count gradually declines southward, lake area continues to plateau till 40°N, owing to the presence of some of the largest lakes in the world such as the Laurentian Great Lakes, Lake Balkhash, and Lake Baikal. As a result, more than 70% of the global lake area is concentrated above 40°N, a latitudinal belt accounting for only one-third of the global landmass (excluding Antarctica).

In comparison, the temperate and tropical zones between 40°N and 40°S are home to about 85% of the global human population (estimated based on the Gridded Population of the World (GPW v4) (CIESIN, 2018)) but contain only 15.5% of the global lake count or 27.3% of the lake area, highlighting the unequal spatial distribution of lake water resources. Longitudinally, 63.6% of the global lakes (or 57.4% by area) are distributed in the land-lacking western hemisphere (Figure 6c) due to disproportionate lake densities in Alaska, the Canadian Shield, the Amazon floodplain, and alpine Patagonia. A spike in lake area is also seen around 30°E, associated with Lake Victoria and a few elongated large lakes in the East African Rift System such as Lakes Tanganyika and Malawi. Another cluster of lake abundance occurs in the longitudinal belt of 60°E to 90°E, due to thousands of thermokarst lakes across the West Siberian Lowlands and glacial lakes of the Tibetan Plateau.

4.2. Comparison With Other Global Lake Masks

We compare the PLD prior lake mask with HydroLAKES, GLAKES, and the entirety of the Circa-2015 lake data set, to further understand the PLD's capability in supporting SWOT's science goal for global lake monitoring. The comparison emphasizes the characteristics of lake size distribution, shoreline fractality, and lake mask accuracy across different landscapes, in addition to summary statistics on global lake abundance. While the prior lake mask is, to a large extent, a subset of the Circa-2015 lake data set (Section 2.1), we include Circa-2015 for comparison in order to understand the abundance of small lakes that are inventoried but beyond SWOT's science goal (<1 ha).

As shown in Figure 7a, all data sets concur that the distribution of the Earth's lake area is asymmetric and lake abundance increases as lake size decreases. When lakes are larger than a scale of ∼100 ha and smaller than ∼1,000,000 ha, the abundance-size relationship conforms to a power-law or Pareto distribution, where the cumulative number of lakes larger than a certain size decreases linearly with increasing lake size in logarithmic space. This power-law relationship supports the perception that lakes behave as self-similar, scale-invariant fractals at least within a certain size range (Downing et al., 2006; Goodchild, 1988; Seekell et al., 2013). Beyond the range, lakes deviate from a power-law distribution, as observed in the lower and upper tails of the abundance-size relationship in Figure 7a. Earlier studies attributed the upper-tail deviation to the increasing randomness (case-specific sizes) of very large lakes and the lower-tail deviation to the incomplete or uncertain representation (sampling biases) of small lakes (Lehner et al., 2011, 2024; Messager et al., 2016; Seekell & Pace, 2011). However, since ∼100 ha (1 km^2) is one to two orders of magnitude of the minimum lake size in these global data sets, the lower-tail deviation is unlikely to result from sampling biases of small lakes, but instead confirms that lakes do not behave as fractals below a certain cutoff area (Kyzivat et al., 2019; Mandelbrot, 1982). Cael and Seekell (2016) explained that such a lower cutoff area exists because topographic characteristics at sub-kilometer scales are less self-similar (echoing the result of a terrain sensitivity analysis by Seekell et al. (2013)), and the development of small lakes is more subject to external dynamics that are scale dependent. Pi et al. (2022) also noted that lakes <100 ha, despite accounting for only ∼15% of the global lake area, dominated lake area variability over the past four decades, further highlighting the unique roles of small lakes in representing regional geomorphic processes and regulating surface water dynamics.

For the above reasons, the capability of characterizing small lake abundance is critical to the SWOT PLD. Following Messager et al. (2016) for HydroLAKES, we used 1,000,000 ha as the upper cutoff area for the power-law relationships of all lake data sets. With the upper cutoff fixed, we tested how the power-law fitting slope varies with a changing lower cutoff area and considered the stationary point on this relationship as the optimal

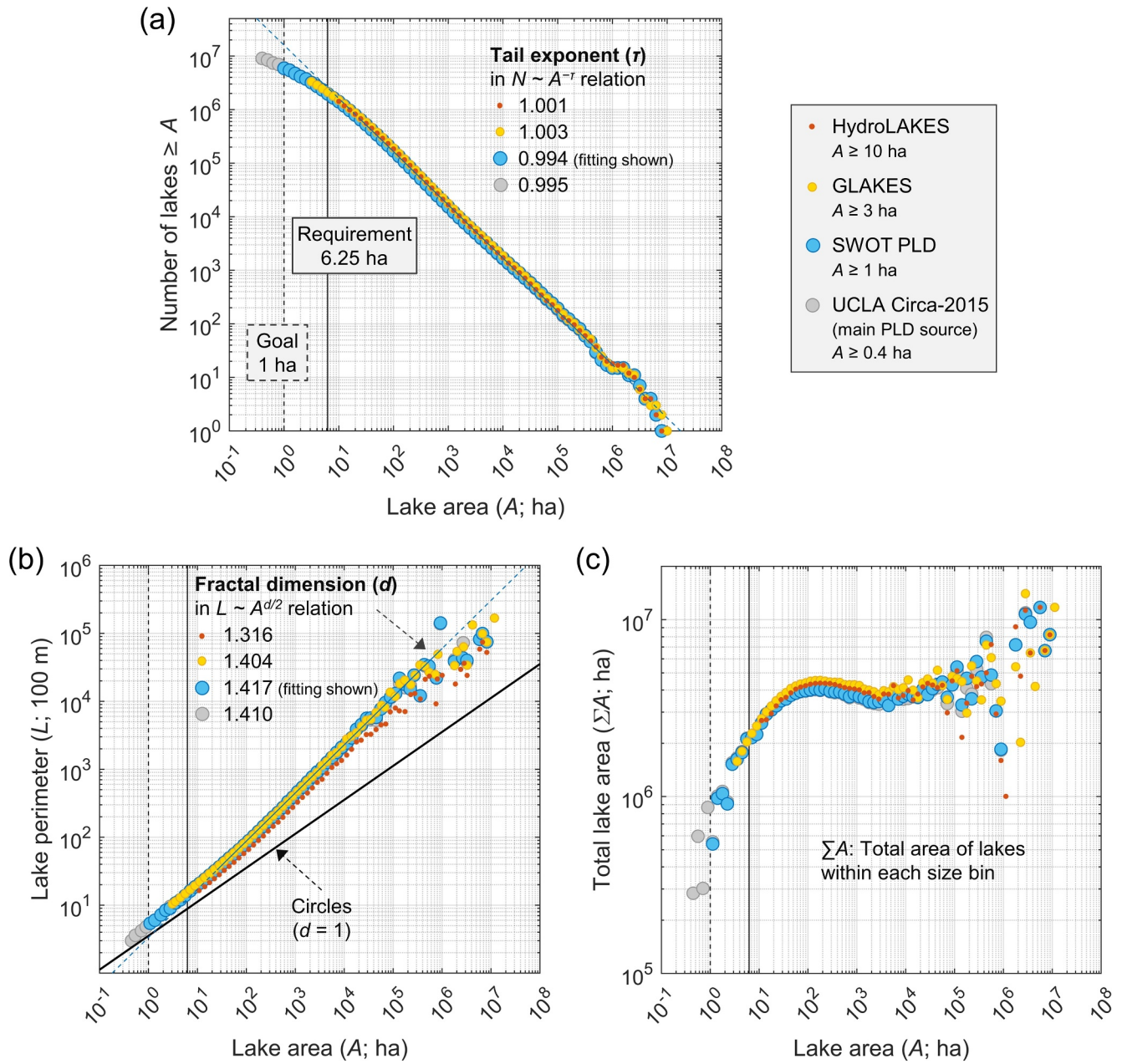

Figure 7. Comparison of lake size distribution and morphometric characteristics among different global lake masks (SWOT PLD, HydroLAKES, GLAKES, and the UCLA Circa-2015 Global Lake Dataset). (a) Lake abundance-size relationship, depicting how the cumulative number of lakes (N) larger than a certain lake size (A) decreases as A increases. Lakes with sizes between a lower cutoff area (25–40 ha depending on data sets) and an upper cutoff area (1,000,000 ha) are power-law (Pareto) distributed, with tail exponents (τ) of different data sets highly consistent with 1.05 predicted by percolation theory. (b) Lake perimeter-area scaling relationships, where dots represent median perimeter lengths (L) and areas (A) of the lakes falling within each of the size bins. As lakes within the Pareto range are fractals, their perimeter-area scaling conforms to power laws, with fractal dimensions (d) of different data sets close to theoretical prediction 4/3. The solid black line represents a hypothetical baseline scenario where lakes are circles of different sizes ($d = 1$). In both panels (a) and (b), the power-law fitting line is only shown for the PLD (blue), and the line appears solid between the lower and upper cutoff areas (where the power law applies) and dashed beyond this range. Details on cutoff areas and fitting statistics for each data set are given in Table S2 in Supporting Information S1. (c) Total areas of lakes falling within each of the size bins. In all plots, lakes of different sizes are grouped into logarithmically spaced bins (x-axis), with 0.1 intervals in both directions from 1 ha. Solid and dashed black vertical lines mark the lake sizes for the SWOT observation requirement (6.25 ha) and goal (1 ha), respectively. The Caspian Sea was excluded from the statistics.

lower cutoff for each data set (see Text S1 in Supporting Information S1 for details). Our result show that all data sets have fairly consistent lower Pareto cutoffs, ranging from ~25 ha for the PLD and Circa-2015, ~32 ha for HydroLAKES, to ~40 ha for GLAKES. Then, using the subset of lakes between the lower and upper cutoffs, we

fitted a power-law function for each of the data sets (fitting for the PLD shown in Figure 7a, and equations for all datasets in Table S2 in Supporting Information S1), which rendered similar tail exponents around 1.00, close to 1.05 predicted by percolation theory (Cael & Seekell, 2016). While this consistency suggests that the four data sets are comparable in representing the abundance of larger lakes, the major difference is their ability to characterize smaller lakes that deviate from a Pareto distribution. As shown in Figure 7a, the pattern of how this deviation develops becomes increasingly clear as the lower tail extends from the minimum lake size of 10 ha in HydroLAKES, 3 ha in GLAKES, 1 ha in PLD, to 0.4 ha in Circa-2015. Put in the context of SWOT, the lower-tail deviation is to the extent that there are 23% fewer lakes meeting SWOT's science requirement (≥6.25 ha) than would be expected if the lakes conformed to power law across the entire size range, and the deviation is amplified to 63% for lakes meeting SWOT's science goal (≥1 ha).

Besides size distribution, lake perimeter-area scaling relations are plotted in Figure 7b to compare shoreline convolutedness (complexity) among the data sets. As fractals are self-similar and scale-invariant, their perimeters and areas are related to each other by power laws (Cheng, 1995), and the exponent, equivalent to the slope of perimeter-area scaling in logarithmic space, defines half of the fractal dimension (d) measuring how convoluted the fractal shorelines are relative to non-fractal, regular features ($d = 1$, e.g., simple circles; Text S2 in Supporting Information S1). As expected, the perimeters and areas of lakes within the lower and upper cutoffs in all data sets follow power-law relationships. The fitted d ranges from 1.32 for HydroLAKES to 1.42 for PLD (fitting shown in Figure 7b), which are overall consistent with 4/3 predicted by percolation theory (Cael & Seekell, 2016). The smaller d for HydroLAKES was likely because the scales of some source data, such as the MODerate resolution Imaging Spectro-radiometer (MODIS) MOD44W water mask (Carroll et al., 2009), might underrepresent shoreline complexity, in combination with additional shoreline smoothing during data post-processing (Messager et al., 2016) (Figure 8). As fractality decrease among lakes smaller than the lower cutoffs, the lake masks with finer resolutions, particularly the PLD and the Circa-2015 data set, reveal a subtle transition of d toward 1 (Figure 7b), echoing the finding of Cael and Seekell (2016) based on high-resolution Swedish lakes that the shapes of small lakes are less convoluted and less fractal-behaving. This comparison highlights the advantage of the PLD in representing shoreline morphology for both sizable and small lakes.

We further compare the lake masks using their summary statistics (Table 5) and discuss the implications of discrepancies among them. The total lake count in the PLD (~5.9 million ≥1 ha) is nearly double that in GLAKES (3.4 million ≥3 ha) and more than quadruple that in HydroLAKES (1.4 million ≥10 ha). These multi-fold differences reflect an unparalleled ability of the PLD to characterize the sheer number of small but SWOT-visible lakes. This improvement is exemplified by two high-latitude lake-rich regions: one in the Kanin Peninsula of Russia dotted with circularly shaped thermokarst lakes and bogs (Figure 8a), and the other in the interior of the Canadian Shield, which is dominated by more convoluted, elongated lakes largely controlled by structural geology (Figure 8b). In both examples, the PLD shows superiority in representing local lake density, geolocations, and shoreline morphology. The Circa-2015 lake data set includes another 3.0 million lakes in the world beyond SWOT's observation goal (<1 ha), although these tiny ponds add <1% to the total lake area. Despite a significantly greater lake population in the PLD, its total lake area (2,624,028.6 km^2) is 6.5% lower than that of GLAKES (2,805,886.2 km^2) and exceeds that of HydroLAKES (2,556,555.5 km^2) by only 2.6%.

For more detailed comparisons, we broke down these statistics into area classes determined by the minimum lake sizes of each of the data sets as well as SWOT's science requirement (Table 5). For lakes <6.25 ha and ≥3 ha (the minimum size in GLAKES), the abundance in the PLD exceeds that in GLAKES by ~2% for both count and area. Based on our visual comparison, we attribute this difference to an overall greater omission error for such small lakes in GLAKES (e.g., in Figures 8a and 8b, small lakes in GLAKES often show slightly more conservative shorelines than those in the PLD, even though the former was mapped based on all-time maximum water occurrence). This result also suggests that the efficacy of the self-adaptive mapping algorithm used for the Circa-2015 data set (Sheng et al., 2016) is on par with that of the non-parametric expert system used for GSWO (Pekel et al., 2016) (also see Section 4.3). For lakes of large classes, however, the abundance in the PLD becomes 6%–9% less than that in GLAKES and 1%–4% less than that in HydroLAKES. It is worth noting that HydroLAKES contains 51,114 lakes with polygon geodesic areas slightly smaller than 10 ha, which we also included for comparison (assuming each of their areas is 10 ha). If these lakes were excluded, the subset of the PLD lakes ≥10 ha surpasses HydroLAKES by 2.1% in lake count but remains 3.0% lower in total lake area. To investigate if such lower abundance is skewed to any individual lakes or size groups, we calculated how the total lake areas are

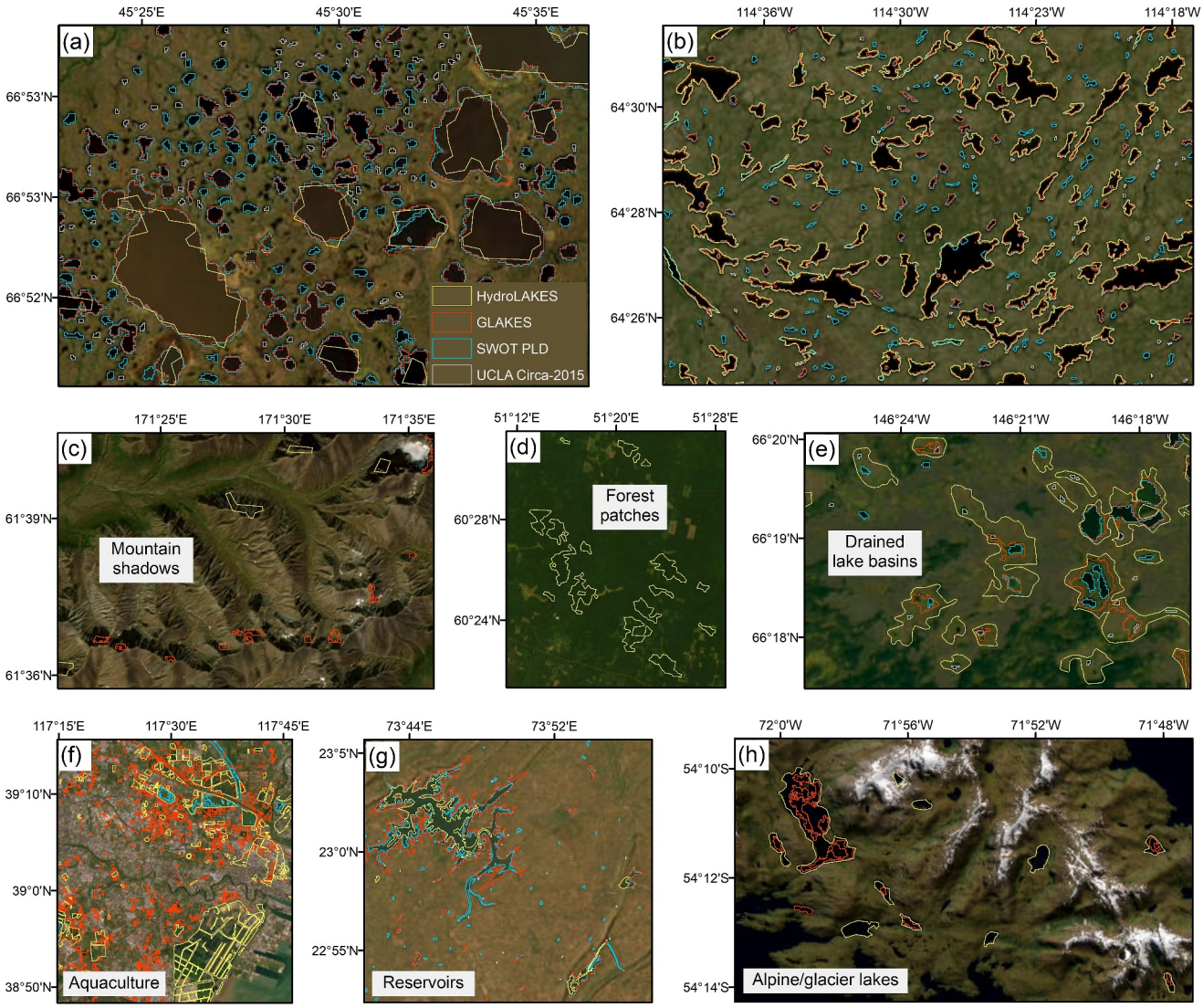

Figure 8. Regional comparisons among the PLD, HydroLAKES, GLAKES, and the Circa-2015 lake data set. (a) Thermokarst lakes in the southern Kanin Peninsula, the Nenets Autonomous Okrug, Russia. (b) Structurally controlled lakes in the Canadian Shield, Northwest Territories, Canada. (c) Commission errors (mountain shadows misclassified as lakes in HydroLAKES and GLAKES) in Kamchatka Krai, Siberia, Russia. (d) Commission errors (forest patches misclassified as lakes in HydroLAKES) in southern Komi Republic, Russia. (e) Thermokarst lakes and drained thaw lake basins in the Yukon River Valley, eastern Alaska. (f) Aquaculture ponds near the Bohai coastline, Tianjin, China. (g) Reservoirs in eastern Gujarat, India. (h) Alpine and glacier lakes in the southern Andes. Background images are from the latest high-resolution Esri Imagery basemap.

distributed across detailed size bins (Figure 7c). The patterns are highly consistent among the data sets: as lakes grow in size, their population (count) decreases monotonically, but the total lake area exhibits a three-phase change. In phase one, the total lake area increases as lake size grows toward ~100 ha, suggesting that for smaller lakes, the area gain due to size growth outpaces the area loss due to population decline. In phase two, the total lake area gradually decreases and stabilizes as lake size grows to ~10,000 ha, suggesting that for medium-sized lakes, the area gain from size growth begins to lag behind the area loss due to population decline, although the former tends to compensate for the latter as lake size continues to grow. In phase three, the total lake area increases once more (despite larger variability) when lake size exceeds 10,000 ha, suggesting the dominant influence of large lakes on area statistics, albeit with a limited population. Regardless of this multi-phase pattern, lake abundance in the PLD remains lower than in the other datasets across most size bins, with this discrepancy becoming increasingly systematic and pronounced as lake size increases to medium scales (~1,000–10,000 ha)

Table 5
Statistical Comparison Among SWOT PLD, HydroLAKES, GLAKES, and the UCLA Circa-2015 Global Lake Dataset

Statistics		HydroLAKES	GLAKES	SWOT PLD	Circa-2015
Minimum lake size (ha)		10	3	1	0.4
Lake count	All	1,427,687	3,426,388	5,897,941	9,040,051
	1–3 ha	0	0	2,590,239	2,659,767
	3–6.25 ha	0	1,303,279	1,327,132	1,345,113
	6.25–10 ha	0	616,371	574,887	579,279
	≥10 ha	1,427,687[a]	1,506,738	1,405,641	1,408,160
Lake area (km^2)	All	2,556,555.5	2,805,886.2	2,624,028.6	2,614,909.9
	1–3 ha	0.0	0.0	46,094.8	47,296.9
	3–6.25 ha	0.0	56,581.3	57,609.6	58,354.9
	6.25–10 ha	0.0	48,654.0	45,296.6	45,641.6
	≥10 ha	2,556,555.5[a]	2,700,650.9	2,475,027.4	2,443,111.3

Note. The Caspian Sea was excluded from the statistics. The PLD contains slightly fewer lakes than Circa-2015 in each size category because of geometric post-processing to improve lake surface connectivity. Additionally, the PLD includes a few dozen lakes smaller than 1 ha from OSM and user requests (Section 4.1), which are included in the statistics for "All." [a]Likely due to projection difference and number rounding, HydroLAKES includes 51,114 lakes that are close to but smaller than 10 ha based on their polygon geodesic areas. These lakes collectively cover a total area of 4,983.2 km^2. For simplicity, we assumed each of these HydroLAKES lakes to be 10 ha and included them in the analysis.

(Figure 7c). For size bins ≥6.25 ha, both lake count and total lake area in the PLD are 10% (median) less than those in GLAKES and 5% (median) less than those in HydroLAKES.

The observed discrepancy in the abundance for lakes ≥6.25 ha reflects differences in mapping standard, quality, timespan, and reference sources among the data sets. A higher lake abundance in GLAKES is expected because its polygons represent all-time water area maximum during 1984–2019 whereas most lakes in the PLD depict representative water extents during circa 2015. Although both data sets were derived from Landsat imagery, differences in mapping period and standard likely led to GLAKES having not only larger areas but also a greater number of lakes within each size group. This is particularly notable given that not all intermittent lakes were inundated during circa 2015. While the Circa-2015 lake data set was supplemented by recently constructed reservoirs (Sections 2.2 and 3.2), natural lakes that disappeared before or emerged after circa 2015 are not included in the PLD. On the other hand, HydroLAKES was a compilation of eight independent lake sources with publication dates spanning a decade (Messager et al., 2016). Variation among these data sources may contribute to a slightly higher abundance (for lakes ≥10 ha) in HydroLAKES than the PLD.

For instance, the acquisition time of SWBD (February 2000), a major source of HydroLAKES for 56°S to 60°N, may explain the smaller areas in the reservoirs of northwestern India (Figure 8g), where water levels were low during the drier monsoon season. In another relevant case, a number of important reservoirs in western Africa were built after February 2000. These include the Ziga Reservoir (impounded around July 2000 based on the GSW database (Pekel et al., 2016)) in Burkina Faso (12.5°N, 1.1°W) that is absent from HydroLAKES 1.0. On the other hand, this acquisition time of SWBD coincided with the warmer season in the southern hemisphere. Meanwhile, SWBD as a radar-derived product (Slater et al., 2006) is less sensitive to surface spectral disturbance such as remnant lake ice and snow. Both factors might lead to a more complete inventory of glacier lakes in HydroLAKES across the southern Andes (Figure 8h).

Another example in Figure 8e highlights a portion of the Yukon River Valley in Alaska, where thermokarst lakes and their drained lake basins develop dynamically atop the permafrost (Grosse et al., 2013). While the PLD polygons appear highly consistent with the recent thermokarst lake extents, HydroLAKES depicts the much larger drained thaw lake basins. These outdated lake boundaries are sourced from the US National Hydrography Dataset (U.S.-Geological-Survey, 2013) and contribute partially to an overestimated area abundance in HydroLAKES.

In addition, part of the higher abundance in HydroLAKES and GLAKES may be ascribed to commission errors (false positives) such as mountain shadows and forest patches, as shown in the examples of Figures 8c and 8d. Such commission errors were largely eliminated from the PLD owing to a rigorous QA/QC procedure (Sheng et al., 2016) (Section 2.1). Other factors such as lake definition and mapping objective could also lead to discrepancies in lake abundance. In Figure 8f, GLAKES and HydroLAKES include a large quantity of aquaculture ponds in coastal China, which were often not considered as lakes in the PLD.

4.3. Uncertainty in Small Lake Abundance

While the PLD enhanced the characterization of small lakes compared to other global data sets, its "exclusive, representative, and exhaustive" mapping criteria were contingent on 30-m resolution Landsat-8 imagery, a mapping standard based on representative (rather than maximum) water occurrence, and a mapping algorithm that tends to identify only pixels with a larger water fraction as water (Lyons et al., 2013) (Section 2.1). These factors represent the major sources of uncertainty in the current version of the PLD. Particularly, these factors do not favor inclusion of littoral zones, where water occurrence is more likely to be intermittent and pixel spectral mixture between vegetation and water often prevails. The impacts of these factors also tend to amplify on smaller lakes (Ogilvie et al., 2018). As lake size decreases, the relative proportion of peripheral pixels increases, leading to a more substantial underestimation of the lake area or, if the shoreline is also convoluted, a complete omission of the lake.

To evaluate the impacts of these factors, especially on the characterization of small lakes, we benchmarked the PLD against several independent, census-quality regional lake data sets. These data sets were retrieved from (a) the hydrographic features of the CanVec Series (1:50,000 scale) (Natural Resources Canada, 2019) for Canada (hereafter "CanVec") (Figure S1 in Supporting Information S1), (b) Virtuellt Vattendrags Nätverk (virtual watercourse network) data set (Nisell et al., 2007) for Sweden (hereafter "ViVaN") (Figure S2 in Supporting Information S1), and (c) NHDWaterbody features in the National Hydrography Dataset Plus High Resolution (NHDPlus HR) (Moore et al., 2019; US Geological Survey, 2022) for the US (hereafter "NHD") (Figure S3 in Supporting Information S1). Detailed data processing is given in Text S3 and Table S3 in Supporting Information S1. Lake features in these data sets were built from high-resolution aerial photos, geospatial data from state and local agencies, and public archives and maps. The inventoried lakes include both perennial and intermittent water bodies with a minimum size of 0.01 ha or finer. Given such details, we consider these benchmark data sets to be the canonical lake "census" for each of the regions.

Synthesizing all three regions (Canada, Sweden, and the US), Figure 9a compares the lake abundance-size curves derived from the census and the PLD (comparisons for individual regions are given in Figure S4 in Supporting Information S1). The margin between the two curves informs an integrated impact of all factors described above. To differentiate the factor impacts, we partitioned the margin by two more abundance-size curves: one based on the subset of the census lakes that are classified as "perennial," to quantify the contribution of water occurrence permanency, and the other based on the Circa-2015 data set to assess the impact of geometric post-processing (e.g., adding new reservoirs and improving water connectivity; Section 3.2). Here we use SWOT's science goal, 1 ha, as the baseline lake size for comparison. According to the census, there are 5,116,374 lakes larger than 1 ha in the three regions, aggregating to a total water area of 1,384,376.7 km^2 (see Text S3 in Supporting Information S1 for caveats in data preprocessing). Taking these numbers as the benchmark, the PLD captures 63.1% of the lake abundance and 91.1% of the total lake area, confirming that underrepresentation in the PLD is skewed toward smaller lakes (see regional maps in Figures S1–S3 in Supporting Information S1). Excluding intermittent lakes reduces the census abundance by 8.8% (or 2.2% in lake area). While this reduction seems to suggest a limited impact of water permanency, we visually inspected NHD and CanVec against Google Earth imagery and found many lakes classified as "perennial" are, in fact, intermittent or ephemeral. This issue of water permanency misclassification has also been noted for the NHD by the US Environmental Protection Agency (2020). Additionally, when a lake is truly perennial, the census polygon often represents the water occurrence closer to the maximum extent rather than the representative extent. Both factors suggest that the impact of water permanency inferred from the census classification is likely conservative. As expected, the post-processing of Circa-2015 reduced the lake count in the PLD but increased the lake area, although the overall impact is limited to ~1% or less. The largest margin occurs between the perennial census lakes and Circa-2015, with 27.0% in terms of lake count and 7.0% in lake area. Detailed statistics, including those for separate regions, are given in Table S4 in Supporting Information S1.

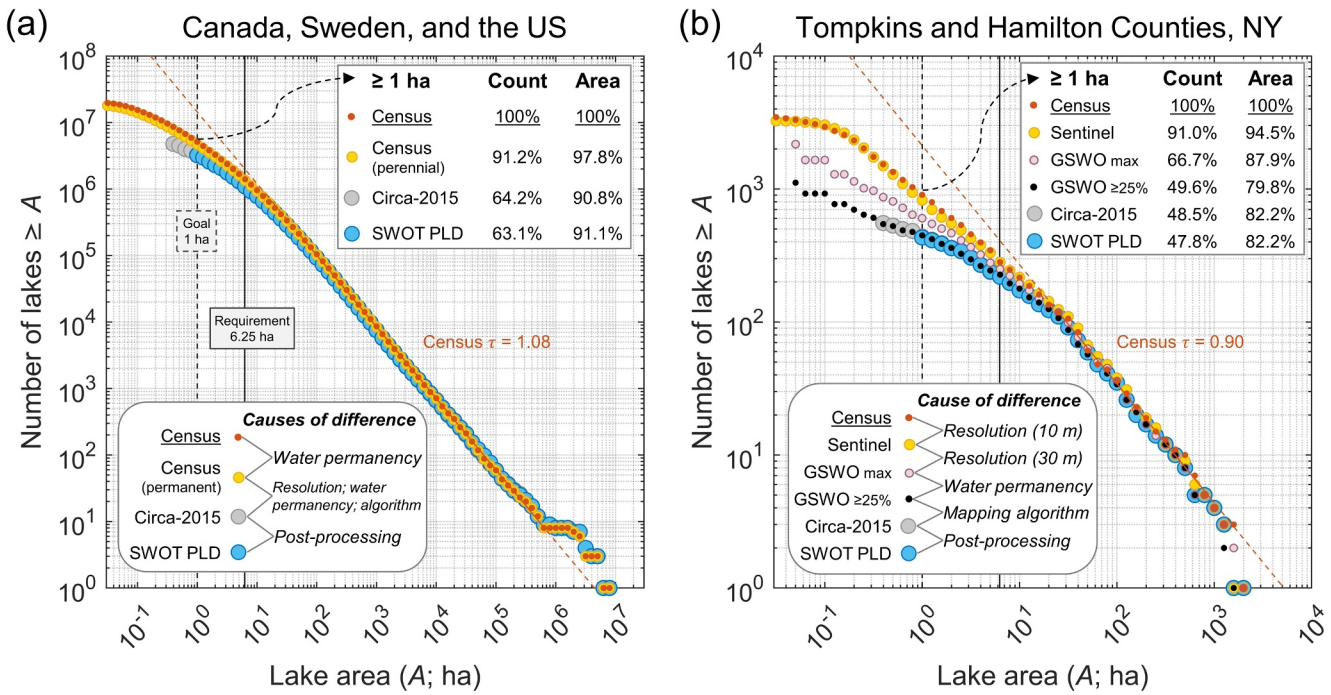

Figure 9. Comparison of lake abundance-size relationships among different data sets in the benchmark regions. (a) Canada, Sweden, and the US combined. (b) Tompkins and Hamilton Counties in the state of New York, US. In both plots, lakes of different sizes are grouped into logarithmically spaced bins (*x*-axis), with 0.1 intervals in both directions from 1 ha. Census data (red dots) refer to NHD for the US, CanVec for Canada, and ViViaN for Sweden and include lakes that are either perennial or intermittent, while census "(perennial)" (yellow dots in panel a) include only perennial lakes if differentiable (see Table S3 in Supporting Information S1). In both panels, the abundance (count) and total area of lakes larger than 1 ha relative to those of the census data are listed for each data set in the rectangle legend, and the main causes of the difference between the data sets are identified successively in the rounded rectangle. The red fitting line shows the Pareto distribution (tail exponent τ labeled) based on the census data between the lower and upper cutoff areas. The fitting line appears solid within this area range and dashed beyond the range. Detailed statistics on data comparison and Pareto fitting are given in Tables S4 and S5 in Supporting Information S1.

To further disentangle the factors contributing to the census-PLD difference, we focus on two counties (Hamilton and Tompkins) in the state of New York, where a high-accuracy lake mask produced from 10-m resolution Sentinel imagery is also available (Liu et al., 2024). This lake mask (Figure S5 in Supporting Information S1) was generated with a novel size-adaptive object mapping algorithm called Optical-SAR Pond Object Mapper (OptiSAR-POM, Liu et al., 2024), using both Sentinel-1 SAR backscatter and Sentinel-2 spectral reflectance images during high water level periods. Again, taking the census statistics for lakes ≥1 as the benchmark, the PLD captures 47.8% of the regional lake abundance and 82.2% of the lake area (Figure 9b), which are overall consistent with the margin calculated for the entire US region (42.1% for abundance and 94.1% for area) (see Figure S4c and Table S4 in Supporting Information S1 for statistics of individual regions). With the help of additional data sets including this Sentinel-based lake mask, we were able to break down the margin for these two New York counties into more specific factors below.

To first isolate the impact of spatial resolution, we generated another lake mask that represents all-time water area maximum within the two counties using the latest Landsat-based GSWO data set (hereafter "GSWO max," Text S3 and Figure S5 in Supporting Information S1). This way, the census data, the Sentinel-based lake mask (10-m resolution), and GSWO max (30-m resolution) share a consistent maximum water occurrence standard, and their discrepancies are attributed predominantly to different image or mapping resolutions. As shown in Figure 9b (detailed statistics in Table S5 in Supporting Information S1), GSWO max captures 66.7% of the lake abundance and 87.9% of the lake area. These scales may represent the largest possible capability of Landsat imagery in characterizing SWOT-observable lakes (>1 ha). While Landsat's 30-m spatial resolution may see lakes as small as 0.1 ha, a practical limiting factor is the rising proportion of spectrally mixed pixels as lakes become smaller. Ogilvie et al. (2018) and Perin et al. (2021) suggested that regardless of the mapping method, pixel spectral mixture may restrict the efficacy of mapping water bodies below ~3–5 ha using Landsat imagery. By upgrading

the resolution to 10 m, the captured lake abundance in the Sentinel-based mask rises substantially to 91.0% (and 94.5% in lake area), indicating a promising potential of Sentinel images for refining small prior lake abundance.

Lake abundance continues to decrease from 66.7% in GSWO max to 48.5% in Circa-2015 (or from 87.9% to 82.2% in terms of lake area). As both data sets are Landsat-based, we interpret this decrease as a combined result of inconsistent water permanency standards (all-time maximum vs. representative extents) and, to a lesser extent, different lake mapping algorithms. As described in Section 2.1, the Circa-2015 data set was mapped using an NDWI thresholding algorithm (Sheng et al., 2016) with validated relative errors of 2%–8% for lakes of ∼1 ha and <1%–2% for lakes larger than 10 ha (Lyons et al., 2013). GSWO was produced from a non-parametric expert system using a decision tree as the inference engine, with overall omission errors less than 5% and commission errors less than 1% (Pekel et al., 2016). These accuracies are fairly comparable. By examining the boundaries of several stable perennial lakes, the mappings of both data sets appear visually similar as well. We therefore consider the impact of different mapping algorithms to be less substantial than the impact of resolution. Then, to isolate the impact of water permanency, we experimented with a series of water occurrence thresholds on the GSWO data set (Text S3 in Supporting Information S1). After removing river segments, the map of water occurrence greater than 25% (hereafter "GSWO ≥25%") exhibits an abundance-size curve that closely resembles that of Circa-2015, with a marginal discrepancy of only 1.1% in lake abundance or 2.4% in lake area (Figure 9b). This resemblance is also echoed by the similarity in the spatial pattern of lake distribution between GSWO ≥25% and Circa-2015 (Figure S5 in Supporting Information S1), suggesting that the "representative" lake extents used in Circa-2015 (and thus largely the PLD as well) correspond to the upper quartile (75%) of the all-time lake inundation extents. For these reasons, we attribute the margin between GSWO max and GSWO ≥25%, that is, 17.1% in abundance and 8.1% in area, to the inconsistency of water permanency standards, and the remaining minor margin between GSWO ≥25% and Circa-2015 (1.1% and 2.4%) to the difference of lake mapping algorithms. Finally, the post-processing of Circa-2015 had a negligible impact on the PLD, which reduced the lake abundance by <1% and had no impact on the lake area in these two counties.

As observed unanimously in each of the regions (Figure 9b and Figure S4 in Supporting Information S1), the abundance-size curves of different lake data sets converge as lake size increases, suggesting a diminishing uncertainty in the abundance of larger lakes. Following the method in Section 4.2, we parameterized the Pareto relationship between lake abundance and size per data set per region (Text S1 in Supporting Information S1). We found that despite variation among the regions, Pareto relationships from lake data sets of the same region, including the lower cutoff areas, are fairly consistent (see detailed statistics in Tables S4 and S5 in Supporting Information S1). The lower cutoff varies from ∼25–32 ha for the two New York counties, ∼40–80 ha for Canada, to ∼400–500 ha for Sweden, with the exception for the US where the lower cutoff for NHD extends to ∼6 ha, much finer than that for the PLD (∼200 ha). The tail exponents appear more consistent within each of the regions, ranging from 0.89–0.95 for the two New York counties to 1.10–1.14 for Sweden. Such consistency further suggests that the uncertainty in lake abundance is predominantly sourced from small lakes that do not behave as self-similar fractals.

To summarize the results from the three benchmarking regions (Tables S4 and S5 in Supporting Information S1), the PLD captures 63.1% of the abundance (or 91.1% of the total area) for lakes larger than SWOT's science goal (1 ha). Inferred from the two counties in the state of New York, 63.8% of the underrepresented lake abundance (or 67.7% of the underestimated lake area) may be attributed to the 30-m pixel resolution of Landsat-8 imagery (e.g., increasing spectral mixture for smaller lakes), about one-third (or 45.6% in area) to the mapping standard using representative lake extents, and the remaining 3.4% (or −13.3% in area) to the accuracy of the mapping algorithm and geometric post-processing. The uncertainty tends to decrease as lake size increases. For instance, the proportion of lake abundance characterized by the PLD increases to 78.2% (or 93.1% in area) for lakes larger than SWOT's science requirement (6.25 ha), which is consistent with previous studies showing that Landsat-based mapping errors tend to concentrate on water bodies smaller than 3–5 ha (Ogilvie et al., 2018; Perin et al., 2021). The proportion of lake abundance characterized by the PLD increases further to 87.5% (or 95.4% in area) for lakes larger than the lower cutoff area (a scale of ∼50 ha) of the Pareto distribution. The remaining margin for large lakes (e.g., ≥50 ha) is partially explained by the effect that the PLD might slightly underestimate the areas on some lakes (due to factors above), rather than omitting them entirely. For example, if the census identifies a lake as 50.2 ha but the PLD mapped it as 49.8 ha, this lake would be excluded from the abundance statistics for lakes ≥50 ha. This effect of lake area underestimation, as opposed to lake omission, may become an increasingly dominant cause of the remaining census-PLD margin as it diminishes with increasing lake size.

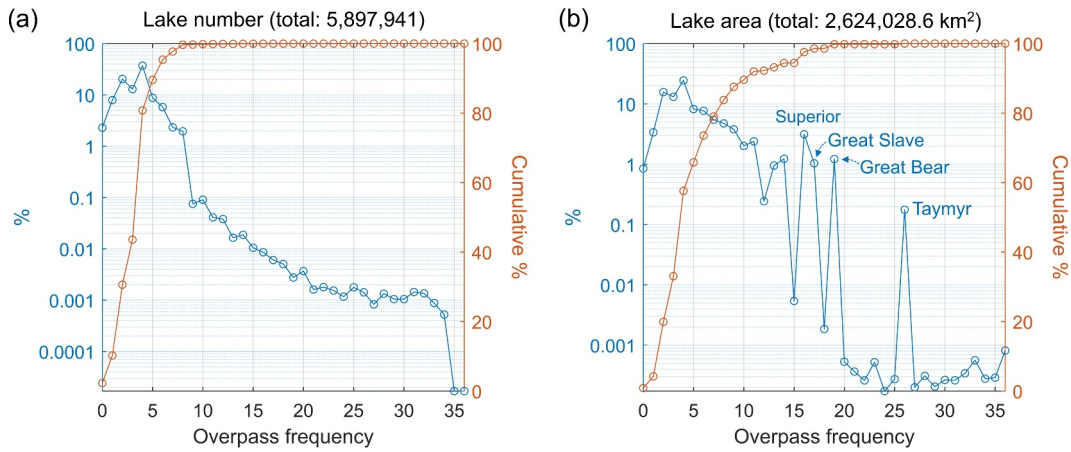

Figure 10. Distributions of lake overpass frequency within a SWOT nominal orbit cycle (21 days). (a) Density (left *y*-axis, blue) and cumulative distribution (right *y*-axis, orange) of overpass frequency in terms of lake count. Note the left *y*-axis has a logarithmic scale while the right *y*-axis has a linear scale. (b) Same as panel (a) but in terms of lake area. Outliers of lake area contribution (jumps on the area density curve) are predominantly caused by a few very large/wide lakes as labeled on panel (b). These individual lakes span multiple SWOT passes, leading to highly frequent but typically spatially incomplete observations, which contribute disproportionately to the area probability density curve.

4.4. Lake Spatiotemporal Coverage

SWOT data coverage of the land surface is determined jointly by orbit characteristics, the KaRIn swath width (2×50 km), the nadir gap width (20 km) between the two swaths, and the orbit crossover density which varies with latitude (Biancamaria et al., 2016). In addition, the spatiotemporal coverage of prior lakes also depends on the size and shape of each lake. With all these factors considered, Figure 6a shows the theoretical frequency of KaRIn observations over each prior lake during every 21-day orbit cycle, which was calculated by summing the counts of unique overpasses in both *pass_ full_nom* and *pass_part_nom* attributes (Section 3.3.4). As summarized in Figure 10, 97.7% of the global lakes, covering 99.2% of the total lake area, are observed by SWOT at least once per orbit cycle. Nearly 70% of the global lakes, covering 80.1% of the lake area, are observed at least weekly on average (i.e., three times per cycle). About 2.3% of the global lakes, or 0.8% of the lake area, may never be observed due to a combination of nadir gaps, orbit intertrack gaps, and exclusion from extreme polar regions beyond the orbit's latitudinal range (78°S to 78°N). This lake coverage complies with the SWOT science requirements, which states that "SWOT shall collect data over a minimum of 90% of all ocean and land area covered by the orbit inclination for 90% of the operation time" (JPL internal document, 2018).

Despite complexity in the global pattern (Figure 6a), lake observation frequency tends to increase with higher latitudes and larger lake sizes. As latitude increases, orbit crossovers densify and overlap among adjacent swaths increases. This gradually leads to the closure of orbit intertrack gaps at 25°S/N and then the closure of nadir gaps at about 62°S/N. As lake size increases, the chance of one lake being observed by multiple passes also increases. As a result, unobserved lakes between 10°S and 10°N make up 5.8% of the local lakes (or 0.6% in terms of lake area), and the proportion decreases to approximately 0.1% or less (in both count and area) at the latitudes above 60°S/N, including regions beyond 78°S/N. Since lake abundance is skewed toward higher latitudes in both count and area (Section 4.1), these factors also explain why SWOT's coverage gap for global lake area (0.8%; Figure 10b) is significantly smaller than that for the entire land surface (3.6%) (Biancamaria et al., 2016).

Globally, the median SWOT observation frequency is maintained at about twice per orbit cycle for lakes smaller than 150 km^2 between 50°S and 50°N. The median frequency increases to three times per cycle for larger lakes over this latitudinal band and for lakes of any size located between 50°–60°S/N. The median frequency increases further to four times per cycle above 60°N. Meanwhile, some of the highest observation frequencies are found in the world's largest lakes regardless of latitudinal distribution. For example, the majority of the 200 largest lakes in the world (Figure 6a) are observed from 5–6 times per cycle (e.g., Lake Nasser, Lake Tanganyika, and Lake Kariba in the tropics) to more than 15 times per cycle (every 1–2 days on average, e.g., Great Slave Lake, Great Bear Lake, and Lake Taymyr in the Pan-Arctic region; Figure 10b).

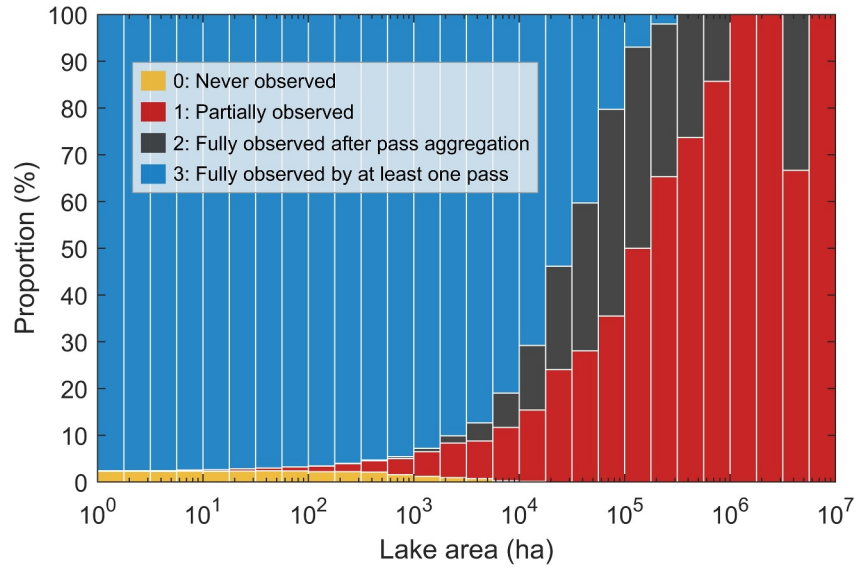

Figure 11. SWOT spatial coverage (observation scenarios in *cycle_ flag_nom*) for prior lakes of different size groups during each nominal orbit cycle. Lakes of varying sizes are grouped into logarithmically spaced bins with 0.25 intervals starting from 1 ha. *Y*-axis shows the stacked proportions (%) of observation scenarios within each size group.

A higher overpass frequency does not always warrant a better spatial coverage. Nearly 6% of the prior lakes, constituting 68.2% of the global lake area, fall on or cross the edge of at least one swath overpass. These lakes will appear incomplete in some of the granules of the single-pass product (L2_HR_LakeSP). However, with a higher overpass frequency, there is an increasing chance that the lake can be fully observed by at least one pass per cycle, or the aggregation of multiple passes can lead to a full extent to represent the average inundation condition during the cycle. The latter reflects the value of the cycle-average product (L2_HR_LakeAvg). To evaluate how lakes are spatially covered per cycle, we calculated the percentage of lakes for each of the *cycle_ flag_nom* scenarios (Section 3.3.4) and analyzed how the percentages vary in lake size. As shown in Figure 11, smaller lakes, albeit overall less frequently observed, are easier to be seen with a full extent. About 95% or more of the lakes smaller than 10 km^2 (1,000 ha) are fully observed at least once per cycle (scenario 3). As lake size increases, the proportion of scenario 3 monotonically declines, while the proportions of lakes that are fully observed only after pass aggregation (scenario 2) and those that remain partially observed after pass aggregation (scenario 1) increase at similar rates. The three scenarios reach comparable proportions at around 300–600 km^2, beyond which scenario-3 lakes are no longer the majority. The proportion of scenario-2 lakes peak at nearly 45% between 600 and 1,000 km^2 before beginning to decline. Lakes larger than this size range are increasingly dominated by scenario 1 until reaching 10,000 km^2, beyond which all lakes, except Great Bear Lake (31,983.1 km^2), can only be partially observed despite very high overpass frequencies. This is because very large lakes tend to include orbit or nadir gaps, except for the case of Great Bear Lake, where these gaps are closed due to its high latitude (~66°N). The proportion of lakes that are never observed (scenario 0) remains below 2.4% regardless of lake size, with 2.6% of these lakes located in polar regions beyond the orbit's latitudinal range. Most confined to SWOT's orbit or nadir gaps, unobserved lakes are typically small (e.g., 97.4% are <1 km^2), and the proportion diminishes to zero for lakes larger than ~200 km^2 (Figure 11).

Summarizing all lake sizes, Figure 12 shows that 97.4% of the global prior lakes, constituting 51.4% of the total lake area, are fully observed at least once during a nominal cycle, and 2.3% (or 0.8% by area) are never seen. Fewer than 1,000 prior lakes, accounting for 11.3% of the global lake area, achieve full coverage after aggregating multiple passes per cycle, whereas the remaining 0.3% (16,189) lakes, accounting for 36.4% of the global lake area, cannot be fully covered during a cycle. These partially observed lakes, which are skewed toward large sizes, may require supplementary data from other sensors or auxiliary water probability maps, such as GSWO (Pekel et al., 2016), to recover complete water areas. Across all Pfafstetter-1 (sub)continents, the proportion of partially observed lakes remains below 1%. In the high-latitude subcontinents SI and AR, over 99% of the prior lakes are

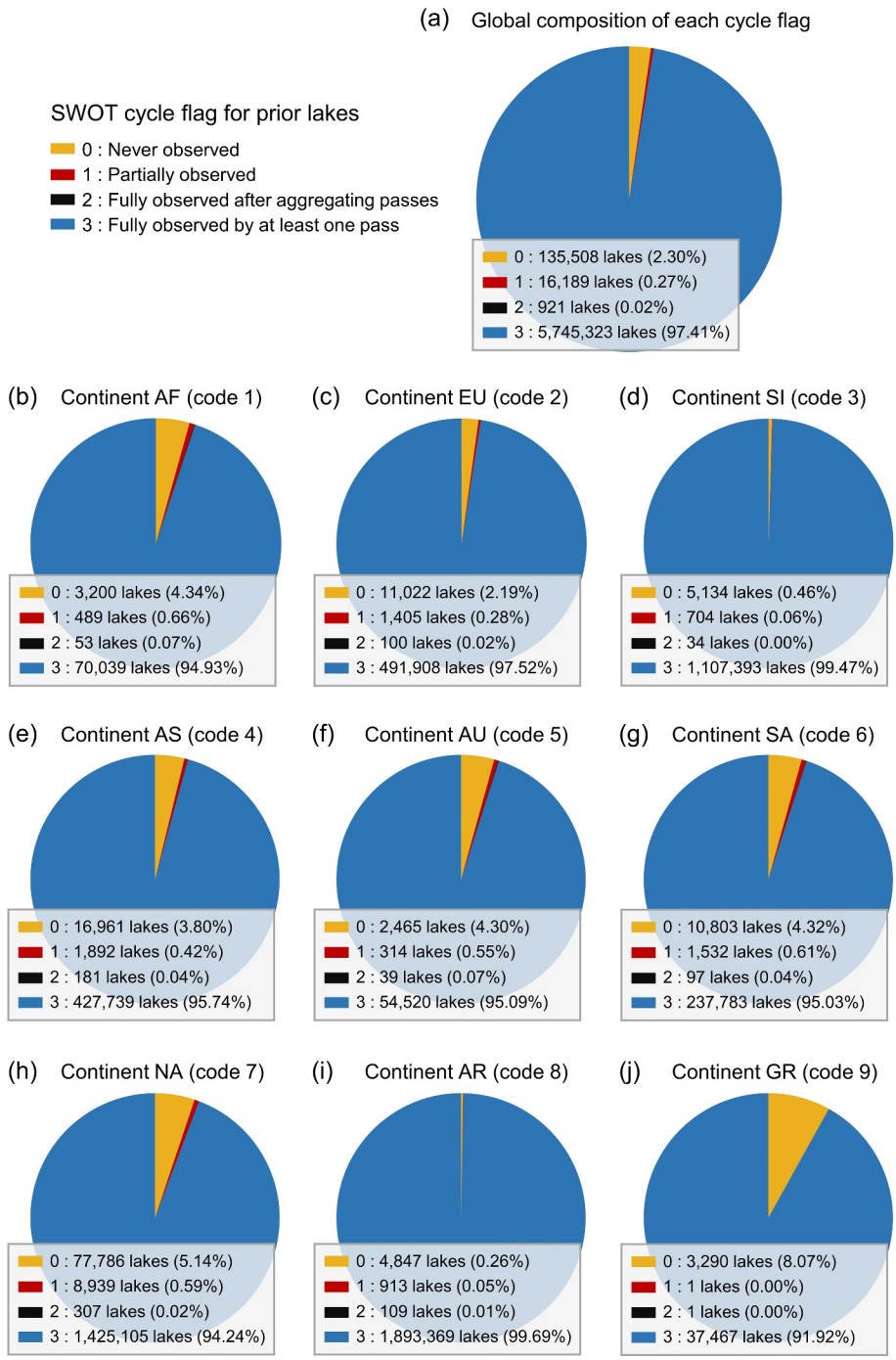

Figure 12. Distributions of lake coverage scenarios during a nominal SWOT orbit cycle for each continent/subcontinent. (a) Composition of cycle flags for global prior lakes. (b–j) Compositions of cycle flags for prior lakes in each of the nine continental divisions (Table 3).

fully observed by at least one pass, while in GR, this proportion is only 91.9% because the remaining 8.1% of the lakes, accounting for 4.5% of the regional lake area, are not covered by the orbit inclination. Notably, approximately a quarter of the prior lakes in AU are observed only by LR products. Despite this regional limitation, LR observations remain valuable for monitoring the dynamics of larger lakes and playas in Australia's drylands.

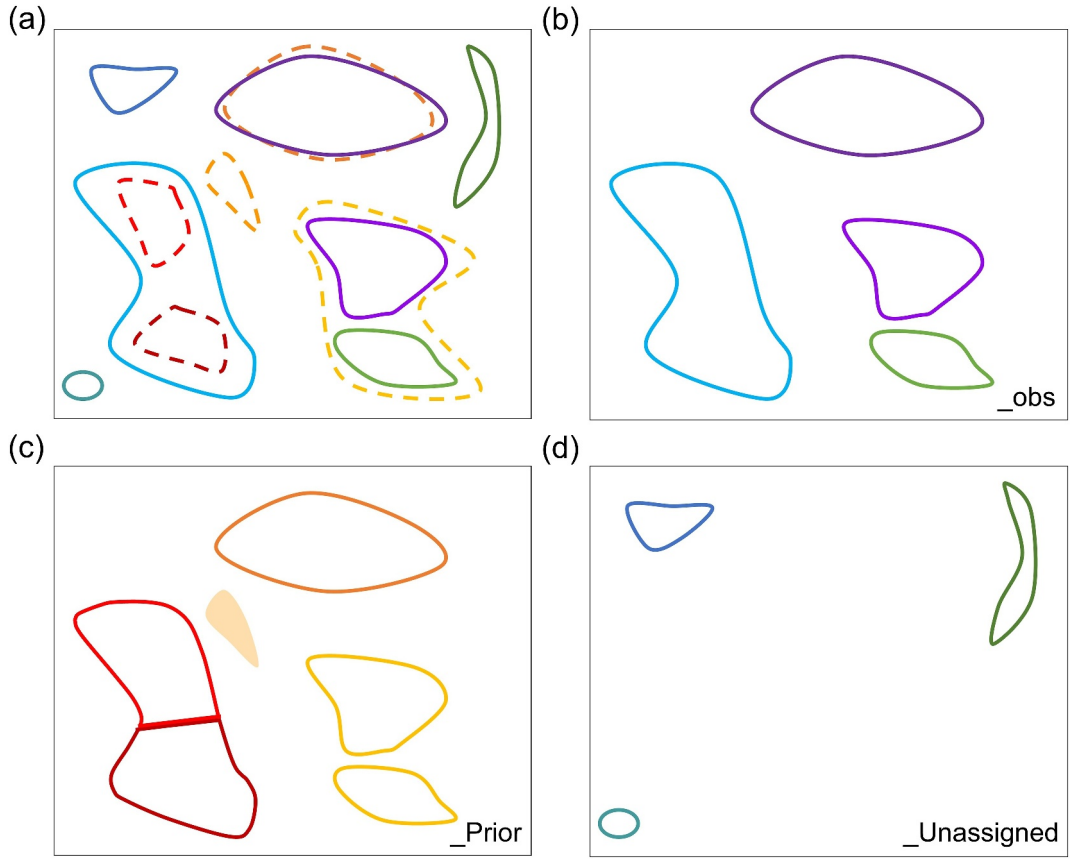

Figure 13. Schematic illustration of how the PLD is used to organize SWOT-observed water features into the three vector files of the L2_HR_LakeSP product. (a) Observed water features (solid) and prior lakes (dashed) in a hypothetical region. Different colors represent different water features or prior lakes. (b) Result of the observation-oriented file (L2_HR_LakeSP_Obs). (c) Result of the PLD-oriented file (L2_HR_LakeSP_Prior). The unobserved prior lake is an empty geometry with only prior attributes, shown as a filled polygon. An observed feature intersecting two prior lakes is partitioned to two features (red and dark red), whereas two observed features intersecting the same prior lake (yellow) are dissolved to a multipart feature. (d) Result of the observation-oriented unassigned file (L2_HR_LakeSP_Unassigned).

4.5. Example of Linking SWOT Observations

Here we provide a conceptual example to demonstrate how the operational PLD assists the SAS in linking KaRIn observations to the prior lakes and generating the L2_HR_LakeSP vector product. More technical details are given in the Algorithm Theoretical Basis Document (CNES internal document, 2023a). As introduced in Section 1, the lake processing pipeline starts from the subset of the pixel cloud (L2_HR_PIXC) after the removal of pixels associated with prior rivers. The remaining non-river pixels are segmented to distinct water regions based on statistical clustering of the pixel height, and the pixel geolocations are further regularized by the average height per water region to reduce noise from the interferogram (Desroches et al., 2016). The resulting pixels with height-constrained geolocations, L2_HR_PIXCVec (CNES internal document, 2022c), are used to vectorize water regions, and the attributes such as water area and average WSE are computed for each vectorized water feature. These processes are directly based on SWOT observations and are independent of the PLD.

The observed water features are next compared with the prior lake polygons to establish spatial linkage between them. Depending on the relationship, the observed water features are organized into three product files (Figure 13): L2_HR_LakeSP_Obs, L2_HR_LakeSP_Prior, and L2_HR_LakeSP_Unassigned. As illustrated in Figure 13a, observed water features (solid) and prior lake polygons (dash) do not always exhibit a one-to-one relationship. A linkage is considered valid if an observed feature intersects at least one prior lake with sufficient overlap, typically defined as 2% or larger (CNES internal document, 2023b). In this case, the water feature is

Figure 14. Example of lake assignment using the operational PLD. (a) SWOT-observed water features in a hypothetical case. (b) Associated prior lakes. Prior lakes 232142722 (*lake_id*) and 232132172 are not observed by this overpass and will be gathered by L2_HR_LakeSP_Prior as empty geometries with only prior attributes. The observed water feature in the central left is linked to no prior lake and will be gathered by L2_HR_LakeSP_Unassigned. The observed water feature in the upper right intersects both prior lakes 232008092 and 232009412. It will be a single feature in L2_HR_LakeSP_Obs but will be split into two separated features in L2_HR_LakeSP_Prior. The observed feature associated with prior lake 232123812 will be gathered by both L2_HR_LakeSP_Obs and L2_HR_LakeSP_Prior with identical geometry. (c) Closeup of the case where one observed feature intersects two prior lakes and how the pixels of this feature are reorganized by the assignment polygons in the *lake_influence* table.

considered a lake and stored in L2_HR_LakeSP_Obs (Figure 13b). Otherwise, the feature is gathered in L2_HR_LakeSP_Unassigned (Figure 13d). Both product files are observation-oriented, meaning that the water features maintain the geometries as observed by SWOT, and the output attributes, such as area and WSE, are the same as those of the input observed features.

To enable storage change calculation, each observed water feature must be linked to a reference water state. However, reference states are available only for prior lakes (Section 3.3.3), which often exhibit complex topological relations with observed features. This spatial inconsistency requires water features in L2_HR_LakeSP_Obs to be reorganized (grouped or split) according to the prior lakes, so that the resulting features and the prior lakes have a one-to-one relationship. The resulting features are gathered in L2_HR_LakeSP_Prior (Figure 13c). This process is straightforward when the original feature intersects only one prior lake. In this case, the geometry of the water feature remains unchanged, and the intersected prior lake with its storage reference attributes (Section 3.3.3) is assigned to this water feature. When a prior lake intersects more than one water feature, all intersected features are grouped to a multipart geometry (i.e., an entity composed of several distinct polygons that represent only one set of attributes), and this prior lake is assigned to the multipart feature.

A more complicated case is one observed water feature intersecting multiple prior lakes. When this occurs, the assignment polygons of the intersected prior lakes in either the *lake_catchment* table or the *lake_influence* table can be utilized to split the observed water feature. Figure 14 illustrates an example using the *lake_influence* table. In this example, an observed feature in the northeast overlaps two prior lakes (*lake_id* 232008092 and 232009412). To partition this feature, each of its PIXCVec pixels is assigned to the prior lake whose area of influence contains the pixel (Figure 14c). Since the influence areas are Thiessen polygons (Section 3.4), this assignment essentially groups the water pixels based on the closest prior lake. The pixels are then re-vectorized based on their prior lake assignment to form separate water features, and the corresponding WSE and water areas are recalculated. Eventually, water storage change for each feature is computed using the storage reference attributes of the prior lake assigned to the feature. Any prior lake that is not observed under an overpass, such as an intermittent lake during the dry season, is also added to L2_HR_LakeSP_Prior but as an empty geometry with only prior attributes. Water storage change is not calculated for L2_HR_LakeSP_Unassigned where features are not linked to any prior lake, thus lacking a reference water state to effectively derive storage change.

5. Versioning Plan

The operational PLD introduced in this paper (version 2.01), while being the first version following peer review, has evolved through a series of improvements over earlier versions. With the accumulation of SWOT observations throughout the mission lifetime, the PLD will continue to be updated recursively to improve the functionality and quality according to the versioning plan configured below.

5.1. Five Update Levels

We envision five levels (*Levels* 0 to 4) of PLD update depending on the quality of the prior lake polygons and the attributes computed from the SWOT lake vector data products. *Level* 0 refers to user-provided inputs. We encourage data users who identify any major issue with the PLD, such as a misrepresented lake extent, the omission of a permanent lake, or an incorrectly merged or split lake polygon, to report to us via email with an explanation of the identified issue and suggestion for improvement. *Level* 1 continues to refine storage reference attributes, including *ref_wse(_u)*, *ref_area(_u)*, *date_t0*, *ds_t0*, and *storage*. As described in Section 3.3.3, *ref_wse(_u)* has been populated using available SWOT data for most prior lakes, while *ref_area* is currently assigned based on the area of the prior lake polygon. As SWOT data accumulate and quality improves, both attributes will be updated using the *wse(_u)* and *area_total(_u)* values from the LakeSP_Prior product (CNES internal document, 2022b), corresponding to the 50th percentile of the time series for each prior lake during the available mission period (see timeline in Section 5.3). Accordingly, *ds_t0* will be updated with reference to the first valid SWOT observation of the prior lake at *date_t0*, and *storage* will be calculated as the maximum water storage variation during the mission period. *Level* 2 generates and updates the *hypso_curve* table. The *hypso_curve* table will be generated by fitting the (*wse*, *area_total*) pairs in the LakeSP product, beginning with the first valid observation for the prior lake (Section 3.1). Each time the table is updated, the fitting will be recomputed using all (*wse*, *area_total*) pairs available from the first valid observation to the end of the update cycle. *Level* 3 updates the geometry of each existing prior lake. This will be done by intersecting the polygons associated with the three highest *wse* values of this prior lake in the LakeSP_Prior product. *Level* 4 adds new prior lakes that are absent from the previous PLD version. New prior lakes will be obtained from the water features that are observed to be persistent in the LakeSP_Unassigned product.

5.2. Three Priority Categories

Along with the five update levels, we will classify the prior lakes into three categories (P1 to P3) based on how complex the update scenario can be, with consideration of the prior lake geometry, SWOT data coverage, and relationship with SWOT-observed water features. These classes will be used to guide the update priority. *Class* P1 updates the "easiest" prior lakes and is defined as any lake that satisfies the following criteria: (a) having a size compliant with the SWOT science requirement (*poly_area* > 6.25 ha); (b) being fully observed by SWOT at least once per cycle (*cycle_flag* = 3); (c) being fairly isolated from other water bodies (*min_dist_lake* > 300 m and *min_dist_river* > 300 m); and (d) exhibiting low complexity in relation to SWOT observation, that is, one prior lake generally corresponds to one SWOT-observed lake feature. *Class* P2 updates the prior lakes that meet criteria (b) and (d) of Class P1, but are smaller in size (*poly_area* > 1 ha) and closer to other water bodies (*min_dist_lake* or *min_dist_river* < 300 m). *Class* P3 updates the other prior lakes.

5.3. General Timeline

The *Level* 1 update began in December 2024, during which storage reference attributes *ref_wse(_u)*, *date_t0*, and *ds_t0* were computed for more than 5.3 million prior lakes across all priority categories, using the LakeSP_Prior product version C derived from processing PIC0 and PIC2. This update resulted in PLD v2.01, as introduced in this paper. The remaining *Level* 1 updates, including populating *ref_area(_u)*, filling missing *date_t0* values, recalculating *ref_wse(_u)* and *ds_t0*, and computing *storage*, are scheduled for completion around December 2025 using LakeSP version D. The *Level* 2 update, focusing on generating the *hypso_curve* table, is also expected around the same time. Updates for *Level* 3 (geometry) and *Level* 4 (new lakes) will be considered thereafter These expected PLD updates will reflect an improved understanding of global lake distribution and dynamics as SWOT observations accumulate. In return, the updated PLD will also refine the processing algorithms for SWOT lake vector products. Beyond supporting SWOT data production, the PLD, with its high-resolution lake mask and multiple operational auxiliaries, can benefit a wide range of disciplines such as limnology, hydrological modeling, ecology, and climate science.

Data Availability Statement

The operational SWOT PLD is openly accessible through the Hydroweb.next platform (https://hydroweb.next.theia-land.fr) under Etalab 2 license. GeoDAR is described in Wang et al. (2022a) and version 1.1 is accessible through Wang et al. (2022b) under CC BY 4.0 on the Zenodo data repository. GREI-p2k is described in Fan et al. (2024) and accessible through Song (2024) under CC BY 4.0 on the Science Data Bank. GRanD is described in Lehner et al. (2011), and version 1.3 is currently archived and accessible through the Global Dam Watch website (Mulligan et al., 2021; https://www.globaldamwatch.org). HydroBASINS is described in Lehner and Grill (2013), and version 1.c is archived and accessed through the HydroSHEDS website (https://www.hydrosheds.org/products/hydrobasins). SWORD is described in Altenau et al. (2021) and version 17 is openly accessible through Altenau et al. (2025) on the Zenodo data repository. SWOT Cal/Val and science orbit swaths are accessed through the AVISO Satellite Altimetry Data website (https://www.aviso.altimetry.fr/en/missions/current-missions/swot/orbit.html).

Acknowledgments
J. W., G. H. A., and Y. S. acknowledge the support of NASA Surface Water and Ocean Topography (SWOT) Science Team Grant 80NSSC20K1143. Y. S. acknowledges NASA SWOT Science Team Grant NNX16AH85G and USGS/NASA Landsat Science Team Grant G12PC00071. L. C. S. acknowledges NASA SWOT Science Team Grant 80NSSC20K1144. The authors also extend their gratitude to David Seekell at Atle Investment Services, Sweden and B.B. Cael at the National Oceanography Center, UK for their assistance in accessing the ViVaN (Virtuellt Vattendrags Nätverk, "virtual watercourse network" for Sweden) dataset.

References

Abbott, B. W., Bishop, K., Zarnetske, J. P., Minaudo, C., Chapin, F. S., Krause, S., et al. (2019). Human domination of the global water cycle absent from depictions and perceptions. *Nature Geoscience*, *12*(7), 533–540. https://doi.org/10.1038/s41561-019-0374-y

Adrian, R., O'Reilly, C. M., Zagarese, H., Baines, S. B., Hessen, D. O., Keller, W., et al. (2009). Lakes as sentinels of climate change. *Limnology & Oceanography*, *54*(6), 2283–2297. https://doi.org/10.4319/lo.2009.54.6_part_2.2283

Allen, G. H., & Pavelsky, T. M. (2018). Global extent of rivers and streams. *Science*, *361*(6402), 585–587. https://doi.org/10.1126/science.aat0636

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. D., & Bendezu, L. (2021). The surface water and ocean topography (SWOT) mission river database (SWORD): A global river network for satellite data products. *Water Resources Research*, *57*(7), e2021WR030054. https://doi.org/10.1029/2021WR030054

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. D., & Bendezu, L. (2025). SWOT River Database (SWORD) (Version v17) [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.14727521

Aurenhammer, F. (1991). Voronoi diagrams - A survey of a fundamental geometric data structure. *Computing Surveys*, *23*(3), 345–405. https://doi.org/10.1145/116873.116880

Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT mission and its capabilities for land hydrology. *Surveys in Geophysics*, *37*(2), 307–337. https://doi.org/10.1007/s10712-015-9346-y

Boy, F., Crétaux, J.-F., Boussaroque, M., & Tison, C. (2022). Improving Sentinel-3 SAR mode processing over lake using numerical simulations. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 5220518. https://doi.org/10.1109/Tgrs.2021.3137034

Busker, T., de Roo, A., Gelati, E., Schwatke, C., Adamovic, M., Bisselink, B., et al. (2019). A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrology and Earth System Sciences*, *23*(2), 669–690. https://doi.org/10.5194/hess-23-669-2019

Cael, B. B., & Seekell, D. A. (2016). The size-distribution of Earth's lakes. *Scientific Reports*, *6*(1), 29633. https://doi.org/10.1038/srep29633

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

Cheng, Q. (1995). The perimeter-area fractal model and its application to geology. *Mathematical Geology*, *27*(1), 69–82. https://doi.org/10.1007/Bf02083568

CIESIN. (2018). *Gridded population of the world, version 4 (GPWv4): Population count, revision 11 (version revision 11)*. NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4JW8BX5

CNES internal document. (2022a). *Level 2 KaRIn high rate lake average vector product (short name: L2_HR_LakeAvg), revision A. SWOT Product Description*. SWOT-TN-CDM-0676-CNES. Retrieved from https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-docs/web-misc/swot_mission_docs/pdd/SWOT-TN-CDM-0676-CNES_Product_Description_L2_HR_LakeAvg_20220930_RevA.pdf

CNES internal document. (2022b). *Level 2 KaRIn high rate lake single pass vector product (short name: L2_HR_LakeSP), revision A. SWOT Product Description*. SWOT-TN-CDM-0673-CNES. Retrieved from https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-docs/web-misc/swot_mission_docs/pdd/SWOT-TN-CDM-0673-CNES_Product_Description_L2_HR_LakeSP_20220930_RevA.pdf

CNES internal document. (2022c). *Level 2 KaRIn high rate pixel cloud vector attribute product (short name: L2_HR_PIXCVec), revision A. SWOT Product Description*. SWOT-TN-CDM-0677-CNES. Retrieved from https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-docs/web-misc/swot_mission_docs/pdd/SWOT-TN-CDM-0677-CNES_Product_Description_L2_HR_PIXCVec_20220930_RevA.pdf

CNES internal document. (2023a). *Level 2 KaRIn high rate lake single pass science algorithm software: Level 2 processing (short name: SAS_L2_HR_LakeSP: Level 2 processing), initial release. SWOT Algorithm Theoretical Basis Document*. SWOT-NT-CDM-1753-CNES. Retrieved from https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-docs/web-misc/swot_mission_docs/atbd/SWOT-NT-CDM-1753-CNES_ATBD_LakeSP_20230726_Initial_w-sigs.pdf

CNES internal document. (2023b). *Level 2 KaRIn high rate lake tile auxiliary parameter file (short name: Param_L2_HR_LakeTile). SWOT Auxiliary Data Description. 2023*. to be released.

Cooley, S. W., Ryan, J. C., & Smith, L. C. (2021). Human alteration of global surface water storage variability. *Nature*, *591*(7848), 78–81. https://doi.org/10.1038/s41586-021-03262-3

Copernicus Climate Change Service (C3S). (2017). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate.

Crétaux, J.-F., Abarca-del-Rio, R., Berge-Nguyen, M., Arsen, A., Drolon, V., Clos, G., & Maisongrande, P. (2016). Lake volume monitoring from space. *Surveys in Geophysics*, *37*(2), 269–305. https://doi.org/10.1007/s10712-016-9362-6

Crétaux, J.-F., Jelinski, W., Calmant, S., Kouraev, A., Vuglinski, V., Berge-Nguyen, M., et al. (2011). SOLS: A lake database to monitor in the near real time water level and storage variations from remote sensing data. *Advances in Space Research*, *47*(9), 1497–1507. https://doi.org/10.1016/j.asr.2011.01.004

de Fleury, M., Kergoat, L., & Grippa, M. (2023). Hydrological regime of Sahelian small waterbodies from combined Sentinel-2 MSI and Sentinel-3 synthetic aperture radar altimeter data. *Hydrology and Earth System Sciences*, *27*(11), 2189–2204. https://doi.org/10.5194/hess-27-2189-2023

Desroches, D., Fjørtoft, R., Gaudin, J.-M., Ruiz, C., & Blumstein, D. (2016). Precise geolocation of water bodies in SWOT HR InSar data. In *Proceedings of EUSAR 2016: 11th European conference on synthetic aperture radar*. Germany.

Downing, J. A., Prairie, Y. T., Cole, J. J., Duarte, C. M., Tranvik, L. J., Striegl, R. G., et al. (2006). The global abundance and size distribution of lakes, ponds, and impoundments. *Limnology & Oceanography*, *51*(5), 2388–2397. https://doi.org/10.4319/lo.2006.51.5.2388

Durand, M., Fu, L.-L., Lettenmaier, D. P., Alsdorf, D. E., Rodriguez, E., & Esteban-Fernandez, D. (2010). The surface water and ocean topography mission: Observing terrestrial surface water and oceanic submesoscale eddies. *Proceedings of the IEEE*, *98*(5), 766–779. https://doi.org/10.1109/JPROC.2010.2043031

Evans, D. G., & Jones, S. M. (1987). Detecting Voronoi (area-of-influence) polygons. *Mathematical Geology*, *19*(6), 523–537. https://doi.org/10.1007/Bf00896918

Fan, C., Song, C., Wang, J., Sheng, Y., Lin, Y., Yuan, C., et al. (2024). Emerging global reservoirs in the new millennium: Abundance, hotspots, and total water storage. *Science Bulletin*, *69*(14), 2179–2182. https://doi.org/10.1016/j.scib.2024.04.043

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, *45*(2), 2005RG000183. https://doi.org/10.1029/2005rg000183

Fergus, C. E., Lapierre, J.-F., Oliver, S. K., Skaff, N. K., Cheruvelil, K. S., Webster, K., et al. (2017). The freshwater landscape: Lake, wetland, and stream abundance and connectivity at macroscales. *Ecosphere*, *8*(8), e01911. https://doi.org/10.1002/ecs2.1911

Fisher, A., Flood, N., & Danaher, T. (2016). Comparing Landsat water index methods for automated water classification in eastern Australia. *Remote Sensing of Environment*, *175*, 167–182. https://doi.org/10.1016/j.rse.2015.12.055

Fu, L.-L., Pavelsky, T., Crétaux, J.-F., Morrow, R., Farrar, J. T., Vaze, P., et al. (2024). The surface water and ocean topography mission: A breakthrough in radar remote sensing of the ocean and land surface water. *Geophysical Research Letters*, *51*(4), e2023GL107652. https://doi.org/10.1029/2023GL107652

Goodchild, M. F. (1988). Lakes on fractal surfaces: A null hypothesis for lake-rich landscapes. *Mathematical Geology*, *20*(6), 615–630. https://doi.org/10.1007/BF00890580

Grosse, G., Jones, B., & Arp, C. (2013). Thermokarst lakes, drainage, and drained basins. In J. F. Shroder, R. Giardino, & J. Harbor (Eds.), *Treatise on geomorphology*, *Glacial and periglacial geomorphology* (Vol. 8, pp. 325–353). Academic Press.

Herdendorf, C. E. (1984). *Inventory of the morphopmetric and limnologic characteristics of the large lakes of the world, technical bulletin OHSU-TB-17*. The Ohio State University Sea Grant Program.

JPL internal document. (2018). *Surface Water and Ocean Topography Mission (SWOT) Project Science Requirements Document, rev B*. JPL D-61923. Retrieved from https://swot.jpl.nasa.gov/system/documents/files/2176_2176_D-61923_SRD_Rev_B_20181113.pdf

JPL internal document. (2022a). *Level 1B KaRIn high rate single look complex data product (short name: L1B_HR_SLC), revision A. SWOT Product Description*. JPL D-56410. Retrieved from https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-docs/web-misc/swot_mission_docs/pdd/D-56410_SWOT_Product_Description_L1B_HR_SLC_20220727_RevA.pdf

JPL internal document. (2022b). *Level 2 KaRIn high rate river average vector product (short name: L2_HR_RiverAvg), revision A. SWOT Product Description*. JPL D-56414. Retrieved from https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-docs/web-misc/swot_mission_docs/pdd/D-56414_SWOT_Product_Description_L2_HR_RiverAvg_20220927a_RevA.pdf

JPL internal document. (2022c). *Level 2 KaRIn high rate river single pass vector product (short name: L2_HR_RiverSP), revision A. SWOT Product Description*. JPL D-56413. Retrieved from https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-docs/web-misc/swot_mission_docs/pdd/D-56413_SWOT_Product_Description_L2_HR_RiverSP_20220916a_RevA.pdf

JPL internal document. (2022d). *Level 2 KaRIn high rate water mask pixel cloud product (short name: L2_HR_PIXC), revision A. SWOT Product Description*. JPL D-56411. Retrieved from https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-docs/web-misc/swot_mission_docs/pdd/D-56411_SWOT_Product_Description_L2_HR_PIXC_20220727b_RevA.pdf

JPL internal document. (2022e). *Reference orbit track (RefOrbitTrack). SWOT Auxiliary Data Description*. JPL D-105500.

JPL internal document. (2023). *Level 2 KaRIn high rate river single pass science algorithm software (short name: SAS_L2_HR_RiverSP), initial release. SWOT Algorithm Theoretical Basis Document*. JPL D-105505. Retrieved from https://archive.podaac.earthdata.nasa.gov/podaac-ops-cumulus-docs/web-misc/swot_mission_docs/atbd/D-105505_SWOT_ATBD_L2_HR_RiverSP_20230713__w-sigs.pdf

Kokelj, S. V., & Jorgenson, M. T. (2013). Advances in thermokarst research. *Permafrost and Periglacial Processes*, *24*(2), 108–119. https://doi.org/10.1002/ppp.1779

Kyzivat, E. D., Smith, L. C., Pitcher, L. H., Fayne, J. V., Cooley, S. W., Cooper, M. G., et al. (2019). A high-resolution airborne color-infrared camera water mask for the NASA ABoVE campaign. *Remote Sensing*, *11*(18), 2163. https://doi.org/10.3390/rs11182163

Lacroix, M. P., Prowse, T. D., Bonsal, B. R., Duguay, C. R., & Menard, P. (2005). River ice trends in Canada. In *13th workshop on ice covered rivers, Hanover, NH*.

Lehner, B., Beames, P., Mulligan, M., Zarfl, C., De Felice, L., van Soesbergen, A., et al. (2024). The global dam watch database of river barrier and reservoir information for large-scale applications. *Scientific Data*, *11*(1), 1069. https://doi.org/10.1038/s41597-024-03752-9

Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, *296*(1–4), 1–22. https://doi.org/10.1016/j.jhydrol.2004.03.028
Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15), 2171–2186. https://doi.org/10.1002/hyp.9740
Lehner, B., Liermann, C. R., Revenga, C., Vorosmarty, C., Fekete, B., Crouzet, P., et al. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*(9), 494–502. https://doi.org/10.1890/100125
Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, *89*(10), 93–104. https://doi.org/10.1029/2008EO100001
Li, J., & Sheng, Y. (2012). An automated scheme for glacial lake dynamics mapping using Landsat imagery and digital elevation models: A case study in the Himalayas. *International Journal of Remote Sensing*, *33*(16), 5194–5213. https://doi.org/10.1080/01431161.2012.657370
Liu, D., Zhu, X., Holgerson, M., Bansal, S., & Xu, X. (2024). Inventorying ponds through novel size-adaptive object mapping using Sentinel-1/2 time series. *Remote Sensing of Environment*, *315*, 114484. https://doi.org/10.1016/j.rse.2024.114484
Luo, S., Song, C., Ke, L., Zhan, P., Fan, C., Liu, K., et al. (2022). Satellite laser altimetry reveals a net water mass gain in global lakes with spatial heterogeneity in the early 21st century. *Geophysical Research Letters*, *49*(3), e2021GL096676. https://doi.org/10.1029/2021GL096676
Lyons, E. A., & Sheng, Y. (2018). LakeTime: Automated seasonal scene selection for global lake mapping using Landsat ETM plus and OLI. *Remote Sensing*, *10*(1), 54. https://doi.org/10.3390/rs10010054
Lyons, E. A., Sheng, Y., Smith, L. C., Li, J., Hinkel, K. M., Lenters, J. D., & Wang, J. (2013). Quantifying sources of error in multitemporal multisensor lake mapping. *International Journal of Remote Sensing*, *34*(22), 7887–7905. https://doi.org/10.1080/01431161.2013.827343
Manasypov, R. M., Pokrovsky, O. S., Kirpotin, S. N., & Shirokova, L. S. (2014). Thermokarst lake waters across the permafrost zones of western Siberia. *The Cryosphere*, *8*(4), 1177–1193. https://doi.org/10.5194/tc-8-1177-2014
Mandelbrot, B. B. (1982). *The fractal geometry of nature*. WH Freeman.
McFeeters, S. K. (1996). The use of the normalized difference water index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714
Mendonca, R., Muller, R. A., Clow, D., Verpoorter, C., Raymond, P., Tranvik, L. J., & Sobek, S. (2017). Organic carbon burial in global lakes and reservoirs. *Nature Communications*, *8*(1), 1694. https://doi.org/10.1038/s41467-017-01789-6
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. *Nature Communications*, *7*(1), 13603. https://doi.org/10.1038/ncomms13603
Moore, R. B., McKay, L. D., Rea, A. H., Bondelid, T. R., Price, C. V., Dewald, T. G., & Johnston, C. M. (2019). User's guide for the national hydrography dataset plus (NHDPlus) high resolution: U.S. geological survey open-file report 2019–1096 (p. 66). https://doi.org/10.3133/ofr20191096
Mulligan, M., Lehner, B., Zarfl, C., Thieme, M., Beames, P., van Soesbergen, A., et al. (2021). Global dam watch: Curated data and tools for management and decision making. *Environmental Research: Infrastructure and Sustainability*, *1*(3), 033003. https://doi.org/10.1088/2634-4505/ac333a
Naddaf, M. (2023). Ukraine dam collapse: What scientists are watching. *Nature*, *618*(7965), 440–441. https://doi.org/10.1038/d41586-023-01928-8
Natural Resources Canada. (2019). CanVec hydrography: Waterbody features. Retrieved from https://ftp.maps.canada.ca/pub/nrcan_rncan/vector/canvec/fgdb/Hydro
Nie, Y., Sheng, Y., Liu, Q., Liu, L., Liu, S., Zhang, Y., & Song, C. (2017). A regional-scale assessment of Himalayan glacial lake changes using satellite observations from 1990 to 2015. *Remote Sensing of Environment*, *189*, 1–13. https://doi.org/10.1016/j.rse.2016.11.008
Nisell, J., Lindsjö, A., & Temnerud, J. (2007). Rikstäckande virtuellt vattendrags nätverk for flödebaserad modelling ViVaN. Department of aquatic science and assessment Sveriges Lantbrukuniveristet report 2007:17. Swedish with English Summary.
Ogilvie, A., Belaud, G., Massuel, S., Mulligan, M., Le Goulven, P., & Calvez, R. (2018). Surface water monitoring in small water bodies: Potential and limits of multi-sensor Landsat time series. *Hydrology and Earth System Sciences*, *22*(8), 4349–4380. https://doi.org/10.5194/hess-22-4349-2018
Oki, T., & Kanae, S. (2006). Global hydrological cycles and world water resources. *Science*, *313*(5790), 1068–1072. https://doi.org/10.1126/science.1128845
Papa, F., Crétaux, J.-F., Grippa, M., Robert, E., Trigg, M., Tshimanga, R. M., et al. (2023). Water resources in Africa under global change: Monitoring surface waters from space. *Surveys in Geophysics*, *44*(1), 43–93. https://doi.org/10.1007/s10712-022-09700-9
Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes [Letter]. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584
Perin, V., Tulbure, M. G., Gaines, M. D., Reba, M. L., & Yaeger, M. A. (2021). On-farm reservoir monitoring using Landsat inundation datasets. *Agricultural Water Management*, *246*, 106694. https://doi.org/10.1016/j.agwat.2020.106694
Pi, X., Luo, Q., Feng, L., Xu, Y., Tang, J., Liang, X., et al. (2022). Mapping global lake dynamics reveals the emerging roles of small lakes. *Nature Communications*, *13*(1), 5777. https://doi.org/10.1038/s41467-022-34140-9
Pickens, A. H., Hansen, M. C., Hancher, M., Stehman, S. V., Tyukavina, A., Potapov, P., et al. (2020). Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series. *Remote Sensing of Environment*, *243*(15), 111792. https://doi.org/10.1016/j.rse.2020.111792
Riggs, R. M., Allen, G. H., Brinkerhoff, C. B., Sikder, M. S., & Wang, J. (2023). Turning lakes into river gauges using the LakeFlow algorithm. *Geophysical Research Letters*, *50*(10), e2023GL103924. https://doi.org/10.1029/2023GL103924
Ryan, J. C., Smith, L. C., Cooley, S. W., Pitcher, L. H., & Pavelsky, T. M. (2020). Global characterization of inland water reservoirs using ICESat-2 altimetry and climate reanalysis. *Geophysical Research Letters*, *47*(17), e2020GL088543. https://doi.org/10.1029/2020GL088543
Schindler, D. W. (2009). Lakes as sentinels and integrators for the effects of climate change on watersheds, airsheds, and landscapes. *Limnology & Oceanography*, *54*(6), 2349–2358. https://doi.org/10.4319/lo.2009.54.6_part_2.2349
Schwatke, C., Dettmering, D., Bosch, W., & Seitz, F. (2015). DAHITI - An innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences*, *19*(10), 4345–4364. https://doi.org/10.5194/hess-19-4345-2015
Seekell, D. A., & Pace, M. L. (2011). Does the Pareto distribution adequately describe the size-distribution of lakes? *Limnology & Oceanography*, *56*(1), 350–356. https://doi.org/10.4319/lo.2011.56.1.0350
Seekell, D. A., Pace, M. L., Tranvik, L. J., & Verpoorter, C. (2013). A fractal-based approach to lake size-distributions. *Geophysical Research Letters*, *40*(3), 517–521. https://doi.org/10.1002/grl.50139
Sheng, Y., Song, C., Wang, J., Lyons, E. A., Knox, B. R., Cox, J. S., & Gao, F. (2016). Representative lake water extent mapping at continental scales using multi-temporal Landsat-8 imagery. *Remote Sensing of Environment*, *185*, 129–141. https://doi.org/10.1016/j.rse.2015.12.041

Shilts, W. W., Aylsworth, J. M., Kaszycki, C. A., & Klassen, R. A. (1987). Canadian shield. In W. L. Graf (Ed.), *Geomorphic systems of North America* (Vol. 2(Centennial Special), pp. 119–161). Geological Society of America. https://doi.org/10.1130/dnag-cent-v2.119
Shugar, D. H., Burr, A., Haritashya, U. K., Kargel, J. S., Watson, C. S., Kennedy, M. C., et al. (2020). Rapid worldwide growth of glacial lakes since 1990. *Nature Climate Change*, *10*(10), 939–945. https://doi.org/10.1038/s41558-020-0855-4
Sikder, M. S., Wang, J., Allen, G. H., Sheng, Y., Yamazaki, D., Song, C., et al. (2023). Lake-TopoCat: A global lake drainage topology and catchment database. *Earth System Science Data*, *15*(8), 3483–3511. https://doi.org/10.5194/essd-15-3483-2023
Slater, J. A., Garvey, G., Johnston, C., Haase, J., Heady, B., Kroenung, G., & Little, J. (2006). The SRTM data "finishing" process and products. *Photogrammetric Engineering & Remote Sensing*, *72*(3), 237–247. https://doi.org/10.14358/Pers.72.3.237
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 lake distribution. *Permafrost and Periglacial Processes*, *18*(2), 201–208. https://doi.org/10.1002/ppp.581
Smith, L. C., Sheng, Y., MacDonald, G. M., & Hinzman, L. D. (2005). Disappearing Arctic lakes. *Science*, *308*(5727), 1429. https://doi.org/10.1126/science.1108142
Song, C. (2024). Global reservoir inventory of the post-2000 impoundment (GREI-p2k) [Dataset]. *Science Data Bank*. https://doi.org/10.57760/sciencedb.15520
Song, C., Sheng, Y., Wang, J., Ke, L., Madson, A., & Nie, Y. (2017). Heterogeneous glacial lake changes and links of lake expansions to the rapid thinning of adjacent glacier termini in the Himalayas. *Geomorphology*, *280*(1), 30–38. https://doi.org/10.1016/j.geomorph.2016.12.002
Stone, R. (2024). Laid to waste: Ukraine scientists are tallying the grave environmental consequences of the Kakhovka Dam disaster. *Science*, *383*(6678), 18–23. https://doi.org/10.1126/science.zbde496
Tranvik, L. J., Downing, J. A., Cotner, J. B., Loiselle, S. A., Striegl, R. G., Ballatore, T. J., et al. (2009). Lakes and reservoirs as regulators of carbon cycling and climate. *Limnology & Oceanography*, *54*(6), 2298–2314. https://doi.org/10.4319/lo.2009.54.6_part_2.2298
U.S. Environmental Protection Agency. (2020). *Resource and programmatic assessment for the navigable waters protection rule: Definition of "Waters of the United States."*. U.S. Environmental Protection Agency and Department of the Army. Retrieved from https://www.epa.gov/sites/default/files/2020-01/documents/rpa_-_nwpr_.pdf
U.S. Geological Survey. (2013). National hydrography geodatabase: Alaska. Retrieved from http://nhd.usgs.gov/
U.S. Geological Survey. (2022). *USGS national hydrography dataset plus high resolution national release 1 FileGDB*. U.S. Geological Survey, 20220707. Retrieved from https://www.sciencebase.gov/catalog/item/62c6050cd34eeb1417baff15
Verdin, K. L., & Verdin, J. P. (1999). A topological system for delineation and codification of the Earth's river basins. *Journal of Hydrology*, *218*(1–2), 1–12. https://doi.org/10.1016/S0022-1694(99)00011-6
Verpoorter, C., Kutser, T., Seekell, D. A., & Tranvik, L. J. (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, *41*(18), 6396–6402. https://doi.org/10.1002/2014gl060641
Wang, J., Sheng, Y., & Tong, T. S. D. (2014). Monitoring decadal lake dynamics across the Yangtze Basin downstream of Three Gorges Dam. *Remote Sensing of Environment*, *152*, 251–269. https://doi.org/10.1016/j.rse.2014.06.004
Wang, J., Song, C., Reager, J. T., Yao, F., Famiglietti, J. S., Sheng, Y., et al. (2018). Recent global decline in endorheic basin water storages. *Nature Geoscience*, *11*(12), 926–932. https://doi.org/10.1038/s41561-018-0265-7
Wang, J., Walter, B. A., Yao, F., Song, C., Ding, M., Maroof, A. S., et al. (2022a). GeoDAR: Georeferenced global dams and reservoirs dataset for bridging attributes and geolocations. *Earth System Science Data*, *14*(4), 1869–1899. https://doi.org/10.5194/essd-14-1869-2022
Wang, J., Walter, B. A., Yao, F., Song, C., Ding, M., Maroof, A. S., et al. (2022b). GeoDAR: Georeferenced global dams and reservoirs dataset for bridging attributes and geolocations (Version 1.1; Version 1.0) [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.6163413
Wik, M., Varner, R. K., Anthony, K. W., MacIntyre, S., & Bastviken, D. (2016). Climate-sensitive northern lakes and ponds are critical components of methane release. *Nature Geoscience*, *9*(2), 99–105. https://doi.org/10.1038/ngeo2578
WMO. (2022). *The 2022 GCOS implementation plan. GCOS-244*. World Meteorological Organization.
Wu, Q., Ke, L., Wang, J., Pavelsky, T. M., Allen, G. H., Sheng, Y., et al. (2023). Satellites reveal hotspots of global river extent change. *Nature Communications*, *14*(1), 1587. https://doi.org/10.1038/s41467-023-37061-3
Wurtsbaugh, W. A., Miller, C., Null, S. E., DeRose, R. J., Wilcock, P., Hahnenberger, M., et al. (2017). Decline of the world's saline lakes. *Nature Geoscience*, *10*(11), 816–821. https://doi.org/10.1038/ngeo3052
Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT hydro: A high-resolution global hydrography map based on latest topography dataset. *Water Resources Research*, *55*(6), 5053–5073. https://doi.org/10.1029/2019wr024873
Yang, X., O'Reilly, C. M., Gardner, J. R., Ross, M. R. V., Topp, S. N., Wang, J., & Pavelsky, T. M. (2022). The color of Earth's lakes. *Geophysical Research Letters*, *49*(18), e2022GL098925. https://doi.org/10.1029/2022GL098925
Yang, X., Pavelsky, T. M., Bendezu, L. P., & Zhang, S. (2022). Simple method to extract lake ice condition from Landsat images. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 4202010. https://doi.org/10.1109/Tgrs.2021.3088144
Yao, F., Livneh, B., Rajagopalan, B., Wang, J., Crétaux, J.-F., Wada, Y., & Berge-Nguyen, M. (2023). Satellites reveal widespread decline in global lake water storage. *Science*, *380*(6646), 743–749. https://doi.org/10.1126/science.abo2812
Yao, F., Livneh, B., Rajagopalan, B., Wang, J., Yang, K., Crétaux, J., et al. (2024). Leveraging ICESat, ICESat-2, and landsat for global-scale, multi-decadal reconstruction of lake water levels. *Water Resources Research*, *60*(2), e2023WR035721. https://doi.org/10.1029/2023wr035721
Zimnitskaya, H., & Geldern, J. V. (2011). Is the Caspian Sea a sea; and why does it matter? *Journal of Eurasian Studies*, *2*(1), 1–14. https://doi.org/10.1016/j.euras.2010.10.009