International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: https://www.tandfonline.com/loi/tres20

A refined mapping of Arctic lakes using Landsat imagery

Homero Paltan, Jadu Dash & Mary Edwards

To cite this article: Homero Paltan, Jadu Dash & Mary Edwards (2015) A refined mapping of Arctic lakes using Landsat imagery, International Journal of Remote Sensing, 36:23, 5970-5982, DOI: 10.1080/01431161.2015.1110263

To link to this article: https://doi.org/10.1080/01431161.2015.1110263

International Journal of Remote Sensing, 2015
Vol. 36, No. 23, 5970–5982, http://dx.doi.org/10.1080/01431161.2015.1110263

Taylor & Francis
Taylor & Francis Group

A refined mapping of Arctic lakes using Landsat imagery

Homero Paltan[a,b]*, Jadu Dash[a], and Mary Edwards[a,c]

[a]*Geography and the Environment, University of Southampton, Southampton, UK;* [b]*School of Geography and the Environment, University of Oxford, Oxford, UK;* [c]*College of Natural Sciences and Mathematics, University of Alaska, Fairbanks, USA*

(*Received 23 May 2015; accepted 11 October 2015*)

Effective mapping of water bodies at regional scales is a challenge with respect to the description and monitoring of hydrological, climatic, and landscape processes. In a region as sensitive to climate change as the Arctic, inaccurate representation of lake cover has probably led to underestimation of the role of lakes as landscape constituents and thus of their contribution to biochemical cycles. To estimate lake cover reliably (and perhaps also its change through time), the scientific community necessitates techniques for mapping water bodies using satellite sensors that include rich historical data sets and have sufficiently fine spatial resolution. Here we applied a density-slicing detection technique to 617 cloud-free Landsat images for the summer months 2006–2011. We developed a comprehensive database of Arctic lakes with a detection accuracy of 80% and examined spatial patterns of lake distribution in relation to landscape properties. We mapped about 3,500,000 lakes; these cover nearly 6% of the Arctic land surface (about 400,000 km^2) and are typically small (<0.1 km^2). Lake density and lake fraction analyses show that lakes are most common in lowland permafrost areas with tundra vegetation. The method described here can also be used to map and monitor lake cover at regional to hemispheric scales and to monitor changes in lake cover over time.

Introduction

Lakes in the Arctic play an important role in terrestrial ecosystem processes and are major contributors to biogeochemical cycles. In particular, they are an important component in carbon exchange between the land surface and the atmosphere, functioning as both sources of emissions of greenhouse gases (carbon dioxide, CO_2; methane, CH_4) and sinks of organic and inorganic carbon (e.g. Bastviken et al. 2004; Cole et al. 2007; Tranvik et al. 2009).

It is likely that climate changes affecting the Arctic have altered several aspects of lake distribution and function. Lake ecosystems in the high Arctic have shown major changes over the 20th century (Smol et al. 2005). Shifts in hydrology, biology, and spatial patterns of lakes have been related to regional warming. For instance, depending upon the condition of the permafrost, surface thawing in permafrost areas may be linked with either lake formation or drainage (Smith et al. 2005). Modification of wetland spatial extent and ecological functions (Zhang 2005) may also affect the extent of small lakes and ponds.

The observed decadal increases in regional temperatures will likely continue; observations show that the Arctic is the region that has warmed the most (Collins et al. 2013),

*Corresponding author. Email: homero.paltanlopez@ouce.ox.ac.uk

and this is likely to accelerate physical, biogeochemical, and limnological changes in Arctic landscapes. However, the nature and rates of change of such processes are not spatially and temporally uniform across the Arctic. For instance, previous studies have demonstrated that thermokarst lakes in yedoma areas (those with deep, ice-rich silt deposits), such as occur in northeast Siberia, have more active methane fluxes than those in non-yedoma areas (Walter et al. 2006). The size and location of lakes within their catchments partly determine how snowmelt runoff mixes with lake water and affects nutrient loading (Rouse et al. 1997; McDonald et al. 2012). In order to have reliable regional estimates of key processes such as gas and nutrient fluxes there is a need for detailed information on lakes' size, shape, and association with particular landscapes. Thus, for both upscaling studies and assessment of climate-driven changes in the Arctic, accurate spatial representations of lake distributions and characteristics are essential.

Until recently, Arctic regional studies have relied upon lake cover data sets that do not have the necessary spatial resolution for the accuracy required of such estimates. The often used Global Lake World Database (GLWD) was derived from a recompilation of historical maps and extant data sets (Lehner and Döll 2004). However, this data set omits a large number of water bodies in the Arctic, with omissions particularly acute for the smallest lakes (<10 ha or 0.1 km^2; Grosse et al. 2008; Walter et al. 2006). The consequence of these omissions has not been precisely quantified, but a rough comparison between GLWD and ASTER scenes in the Kolyma region of northeast Siberia determined that about 50% of Arctic lakes are omitted in the GLWD (Walter, Smith, et al. 2007).

A pan-Arctic assessment examined the controls on the spatial distribution and abundance of northern lakes (>45°N) using the GLWD as lake input data (Smith, Sheng, and MacDonald 2007). The main patterns driving lake distributions were climate, geomorphology, substrate permeability, glaciation history, and, particularly, the presence of permafrost. They found lake density and average fractional areas to be 300–350% greater in glaciated terrain *versus* unglaciated, and 100–170% greater in permafrost areas *versus* permafrost-free terrains. Moreover, the presence of peatlands is associated with an increase of about 40–80% in lake density and 10–50% increases in area fraction compared with peat-free terrain. Perhaps surprisingly, measures of lake cover were found to be similar across continuous, discontinuous, and sporadic permafrost zones, whereas a modest decrease was seen in zones of isolated permafrost and a sharp drop in the absence of permafrost. However, omissions of smaller lakes from the GLWD suggests that our understanding of the structure of the surface-hydrological landscape and, in turn, the distribution of lakes in the Arctic may be incomplete (Walter et al. 2006; Smith, Sheng, and MacDonald 2007; Carroll et al. 2011; Cole et al. 2007).

Recently, another global database of the abundance and size distribution of lakes, the GLObal WAter BOdies (GLOWABO) database, has been developed (Verpoorter et al. 2014). This database was developed using imagery of the Enhanced Thematic Mapper Plus (ETM+) sensor on board the Landsat 7. By applying a novel algorithm that permitted pan-sharpened images to have 14.25 m spatial resolution, the authors detected about 117 million lakes >0.0002 km^2 worldwide. Boreal and Arctic regions showed the highest concentration, area, and perimeter of water bodies. However, this data set is not yet publicly available, so it is not possible to evaluate its accuracy and utility for the evaluation of temporal change in lake cover or lake–landscape correlations.

In order to map lakes across a vast area such as the Arctic, it is necessary to have a satellite mission that has both the spatial resolution to map extremely small lakes and adequate spatial and temporal coverage to provide enough cloud-free scenes during the ice-free period. The satellite sensors from the Landsat mission fulfilling the above criteria have been selected as a suitable data source for this task. In terms of methods to detect water bodies from optical remote-sensing data, a single-band, density-slicing technique has been shown to be more successful than other automated methods. For instance, a study conducted by Frazier and Page (2000) found that a maximum likelihood classification technique tends to exclude more atypical water pixels, leading to larger errors of omission. More recently, when monitoring lakes in the Yukon basin, Roach, Griffith, and Verbyla (2012) found that the density-slicing technique performs better than the classification tree and feature extraction methods to detect water bodies, as the former resulted in the most omission errors (approximately twofold greater) and the latter resulted in the most commission errors (approximately fourfold greater). Furthermore, that same study demonstrated that using only Landsat band 5 provides the most accurate detection of water bodies (rather than combining other spectral bands) in boreal regions. However, this technique has not been previously applied at a regional scale that includes a large number of image scenes. The objectives in the present study are i) to test the applicability of using a density-slicing technique and Landsat 5 TM imagery as a basis for detecting and monitoring lakes in the Arctic (a rough measure of the lower limit of the Arctic defined here as the region north of 65°N); and ii) to develop a database of lakes for the Arctic region that has high spatial resolution. Our goal is to provide an accurate representation of regional lake abundance and patterns of distribution across the Arctic that will enable better upscaling and thus contribute to the improvement of our understanding of Arctic processes.

Materials and methods

We developed the New Arctic Lakes Geodatabase (NALGDB) using 617 Landsat 5 TM cloud- free images from the summer months of 2006–2011. The time span was necessary to ensure complete coverage of the Arctic with cloud-free Landsat images. In addition, 18 images from 2003–2006 were used to secure as complete a coverage as possible. Even so, some areas in Greenland and north Taimyr (Russia) are either omitted or under-represented because of the lack of cloud-free images. In each Landsat scene, water pixels were extracted using a density-slicing technique for the shortwave infrared band (band 5: 1.55–1.75 μm). Following the delineation of water pixels, areas under water were vectorized and elongated water bodies such as rivers were removed by applying a compactness function, which differentiated rounded features from elongated ones.

The accuracy of NALGDB was assessed using the commonly used ‘error matrix’-based approach (Foody 2002). This assessment was conducted by comparing randomly selected areas with variable lake cover across the Arctic (from the data set) against lakes manually digitized from high-resolution imagery obtained from Google Earth (a detailed description of the method is provided in the Supplemental Material (SM)). Visual inspection also suggests that the omission of areas of Greenland and Taimyr does lead to underestimation of lake cover/density. This is less important for Greenland, but Taimyr has significant lake cover and the lake count there is markedly underestimated.

The spatial analysis of lake distribution was performed by coupling the lake data set with supporting information under a GIS platform. The thematic data included in the analysis were i) vegetation type, ii) permafrost cover, iii) topography, and iv) surface geology (i.e. glaciated or unglaciated, and yedoma or non-yedoma). In addition, a data set of national boundaries was overlaid to estimate lake abundance per country. The vegetation classification broadly follows the Kaplan et al. (2003) classification scheme: cushion-forb-lichen-moss tundra, prostrate-shrub tundra, erect dwarf-shrub tundra, low- and high-shrub tundra, cold deciduous forest, cold evergreen needle-leaf forest, and cold-temperate evergreen needle-leaved and deciduous mixed forest. We added two extra categories: permanent snow and ice and barrens (sparsely vegetated to non-vegetated). Permafrost extent is classified as continuous (90–100%), discontinuous (50–90%), sporadic (10–50%), or isolated patches (0–10%) (Brown et al. 2011). For terrain, highlands and lowlands are separated by a threshold value of 300 m.a.s.l. (obtained from the USGS GTOPO 30 DEM). We obtained areas that were glaciated during the Last Glacial Maximum (LGM) as defined the ICE-5G(VM2) model by Peltier (2004). A map of yedoma extent was produced by digitizing the Siberian extent from Walter, Edwards, et al. (2007) and for North America from Kanevskiy et al. (2011; see SM3). We recognize the latter as an over-estimate. Finally, the spatial distribution of lakes was defined by obtaining regional values of lake density (number of lakes per unit area), lake fraction (portion of land covered by lake – water), and mean size of lakes. These values were assessed in relation to the supplementary information layers.

Results

The NALGDB identifies ~3,500,000 lakes that have a minimum mapping area of 0.0036 km^2 (see Table 1), with 80% accuracy in lake detection. Arctic lakes cover an area of ~350,000 km^2 or about 6% of the total land area. The median lake size is 0.0149547 km^2. Thirty-two per cent of lakes have a surface area of <0.01 km^2 (1 ha), 53% between 0.01 and 0.1 km^2 (1–10 ha), 13% 0.1–1 km^2 (10–100 ha), 1.25% 1–10 km^2 (100–1000 ha), and 1% > 10 km^2. Our data set contains 142 'great lakes' each with a surface area of over 100 km^2. The results show that lakes in the Arctic are typically small (<0.1 km^2). The detailed distribution of densities and lake fractions is presented in Figure 1.

The spatial pattern of overall numbers of lakes in the Arctic is as follows: ~1,900,000 in Eurasia, ~1,600,000 in North America, and ~50,000 in Greenland (a slight under-estimate; see above and SM). The greatest densities of lakes (>0.125 lakes/km^2) and lake fractions (>12.5% of land area) occur in lowland zones of continuous permafrost (Figure 1*(a)* and 1*(b)*). In Eurasia, the southern Yamal Peninsula (65–70°E), areas in Eastern Siberia (correlated with the extent of yedoma), northern Scandinavia, and the northwest Russian plain/Karelia region are particularly lake-rich. The North American

Table 1. Size classes of lakes in the Arctic.

Size class (km^2)	Number of lakes	% of the total	Total lake area (km^2)	Mean size (km^2)	Median size (km^2)	Lake fraction (%)
0.0036–0.01	1,130,262	32.116	7,252.59	0.006417	0.009024	0.10
0.01–0.1	1,875,177	53.282	46,159.78	0.02462	0.024708	0.62
0.1–1	467,886	13.295	126,899.33	0.27122	0.196615	1.70
1–10	43,931	1.248	98,935.00	2.25205	1.644046	1.33
10–100	1,916	0.054	43,224.03	22.55951	15.362864	0.58
>100	142	0.004	72,506.85	510.61162	169.72507	0.97
Total	3,519,314		394,978	-	-	5.30

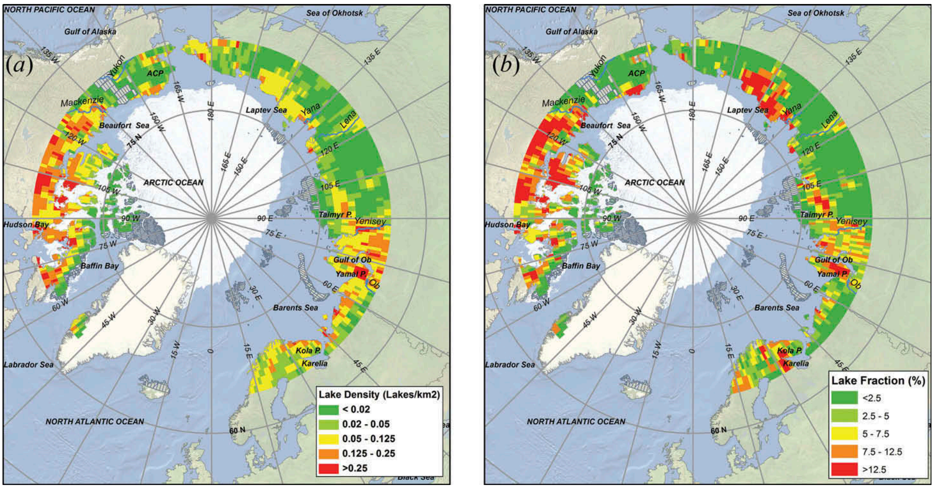

Figure 1. Distribution of lakes in the Arctic. (a) NALGDB estimates of lake density (number of lakes/km^2). Hatched areas indicate where no observations were possible because of cloud cover. White areas are snow and ice (no observations). ACP – Alaskan Arctic Coastal Plain; Kola P. – Kola Peninsula; Taimyr P. – Taimyr Peninsula; Yamal P. – Yamal Peninsula. Major rivers and seas are indicated. (b) Estimates of lake fraction (% land cover).

Arctic has higher lake density overall, however: both the Canadian Shield and the Alaskan Arctic Coastal Plain have large numbers of lakes. There is a general tendency for lake density and fraction both to be higher in areas underlain by permafrost and/or covered by tundra, in yedoma areas, and in areas of lowland terrain. Areas with the lowest lake densities are largely located in the southern Eurasian and Alaskan regions of our study area. As a considerable proportion of lakes in the Arctic is small (85%), it is likely that the regional distribution of small lakes also follows the described pattern of lake density.

In relation to vegetation types, the NALGDB identifies the greatest number of lakes in the erect dwarf-shrub tundra and the low- and high-shrub tundra vegetation units (about one million lakes in each unit – see Table 2). These areas, with the exception of the mountainous western Scandinavian Peninsula, are strongly correlated with the distribution of continuous permafrost. However, they are mostly dominated by small lakes, which results in a relatively low lake fraction. Many of the largest lakes are associated with the Canadian Shield, where vegetation cover is erect dwarf-shrub tundra and cold evergreen needle-leaf forest.

The continuous permafrost zone contains nearly 85% of all Arctic lakes (see Table 3). From the values of mean lake size, it would appear that water bodies in continuous permafrost areas are typically large (~0.4 km^2). However, given that about 84% of all lakes are small, low numbers of larger lakes bias the mean estimates, meaning that mean values are not particularly useful in describing regional lake-cover patterns.

In yedoma terrain, most of the lakes are small, with 80% of them having an area of <0.1 km^2. In Siberia about 400,000 lakes were identified, covering nearly 67,000 km^2 and representing about 7% of the total yedoma land surface. For North America, about 17,000 lakes were identified; these forming <1% of the total North American yedoma area (255 km^2). However, these estimates rely on the accuracy of the spatial depiction of

Table 2. Summary of the distribution of lakes in the Arctic per vegetation unit.

Vegetation unit	Extent (km^2)	Number of lakes*	% of the total	Total lake area (km^2)	Mean size (km^2)	Median size (km^2)	Lake density ($lakes/km^2$)	Lake fraction (%)
Barren	97,348	334,410.00	8.76	3,435.43	0.102826	0.017165	3.44	3.53
Cold deciduous forest	188,563	220,844.00	5.79	35,780.61	0.162018	0.022020	1.17	18.98
Cold evergreen needle-leaf forest	1,468,069	497,525.00	13.03	90,906.41	0.182717	0.017093	0.34	6.19
Cool-temperate evergreen needle-leaf and mixed forest	3,366	2,554.00	0.07	267.03	0.104557	0.016657	0.76	7.93
Cushion-forb, lichen, and moss tundra	478,541	275,345.00	7.21	30,526.22	0.110865	0.016930	0.58	6.38
Erect dwarf-shrub tundra	1,370,578	1,009,124.00	26.44	129,852.36	0.128678	0.019460	0.74	9.47
Low- and high-shrub tundra	1,970,209	930,259.00	24.37	142,326.30	0.152996	0.017714	0.47	7.22
Permanent snow and ice	73,304	1,781.00	0.05	1,845.11	1.035999	0.012398	0.02	2.52
Prostrate dwarf-shrub tundra	724,437	545,090.00	14.28	62,819.45	0.115246	0.016910	0.75	8.67

Note: *This total figure may differ from other values presented in this study, as the same lake may lie in two countries.

Table 3. Summary of the distribution of lakes in the Arctic by permafrost zone.

Permafrost	Extent (km^2)	Number of lakes*	% of total	Total lake area (km^2)	Mean size (km^2)	Median size (km^2)	Lake density ($lakes/km^2$)	Lake fraction (%)
Continuous (90–100%	6,015,219	2,897,806.00	81.47	338,122.58	0.411513	0.018100	0.48	5.62
Discontinuous (50–90%)	585,485	321,622.00	9.04	59,332.89	0.114481	0.015901	0.55	10.13
Sporadic (10–50%)	176,784	103,027.00	2.90	11,351.83	0.110831	0.014082	0.58	6.42
No Permafrost	675,068	234,573.00	6.59	67,636.21	0.240521	0.015239	0.35	10.02

Note: *This figure may differ from other values presented in this study as some lakes may be found in two units.
**Values are underestimated as satellite imagery was incomplete for this country (See Supplemental Material).

Table 4. Summary of the distribution of lakes in the Arctic by country.

Country	Extent – as total land area (km^2)	Number of lakes*	% of the total	Total lake area (km^2)	Mean size (km^2)	Lake density (lakes/km^2)	Lake fraction (%)
Canada	2,345,237	1,491,291	42.32	194,176.09	0.130207	0.64	8.28
Finland	121,903	46,377	1.32	5,022.01	0.108287	0.38	4.12
Greenland**	1,948,583	16,684	0.47	2,240.30	0.134278	0.01	0.11
Iceland *no data*	50,928	*no data*	*no*	*data*	*no data*	*no data*	*no data*
Norway	111,800	61,623	1.75	6,376.58	0.103477	0.55	5.70
Russia**	4,718,672	1,716,248	48.70	81,903.57	0.047722	0.36	1.74
Sweden	129,729	57,994	1.65	8,138.29	0.140330	0.45	6.27
United States (Alaska) **	605,845	134,026	3.80	13,787.23	0.102870	0.22	2.28

Note: *This figure may differ from other values presented in this study as some lakes may lie in two countries. **Values may be underestimated as satellite imagery was incomplete for this country.

surficial deposits, which is simplified in both source maps and thus should be considered approximations.

Lastly, the country-based analysis shows that (unsurprisingly) Russia and Canada, the nations with the greatest extent of Arctic land, have the greatest numbers of lakes (1.7 and 1.4 m, respectively), and numbers of lakes are typically correlated with country size (except for Greenland). However, Canadian lakes are typically larger than Russian, and thus in Canada lakes occupy a greater portion of the territory (i.e. 8.28% Canadian territory *versus* 1.74% in Russia). In spite of having a relatively small number of lakes compared with other countries, the Scandinavian countries show high lake fractions and densities, emphasizing the importance of lakes as landscape elements in these countries. National figures per each country are presented in Table 4.

Discussion

The NALGDB is a new data set for the study of Arctic lakes. We defined water bodies as lakes according to surface features that meet our minimum mapping area and morphological characteristics defined by compactness properties (see SM). We recognize that other lake-like features, such as lagoons, might have been added to the database; however, their abundance is relatively low in this region and will have limited impact on the overall statistics presented in this study. A detailed limnological and geomorphological classification of individual elements is beyond the scope of this study.

With an overall detection accuracy of 80%, our results appear robust. However, further investigation shows that our method performs better in tundra regions than in forested areas (see SM). In the tundra regions as well as 'barren' areas (e.g. the high alpine) the accuracy is even greater than 80%. Such figures enrich the reliability of this data set, given that tundra and barren areas occupy about 75% of our study area. The most common source of commission error is from some palsas, mires, and fens being detected as lakes. Ephemeral bodies resulting from thaw processes, which drained or evaporated over a season, may also be a source of commission errors, but these are likely to be restricted largely to floodplains. The omission errors were mostly from lakes in areas adjacent to rivers. In some cases these were grouped together with the river as one water body and were removed from the database based on the compactness threshold.

Comparisons with previous estimates of Arctic lake cover

Our lake estimates exceed those of the GLWD in terms of lake number, lake density, and lake area. This is mostly due to the use of data of higher spatial resolution. While there may be some error in the identification and distribution of very small lakes, as they lie at the limits of the spatial resolution of the satellite data, there is a clear dominance of small lakes in the Arctic. Our results corroborate those of McDonald et al. (2012) and Verpoorter et al. (2014; GLOWABO), who also identified large numbers of small lakes at an even higher spatial resolution, the former in the USA and the latter globally.

While there are numerous small lakes globally, in the GLOWABO database (Verpoorter et al. 2014) the abundance–size relationship of lakes fails to follow a power law that underlies previous estimates of global lake numbers (e.g. Downing et al. 2006) . It might be suspected that very small lakes are proportionally more abundant in the Arctic than in other regions, because of the occurrence of numerous small ponds linked to extensive wetlands and the presence of permafrost (Lehner and Döll 2004; McDonald et al. 2012). A test of the abundance–size relationship with our NALGDB data set suggests a power law may hold for the larger size classes (Figure 2), but the frequency of lakes in the smallest measured size class falls markedly. While this is partly related to the limits of detectability that reduce the bin size of the smallest class, making it not strictly comparable, we conclude that there is a likely lower limit for the power law, even in the Arctic, making a statistically based estimation of the number of even smaller lakes likely to be unreliable. Any estimates for the Arctic should be made at

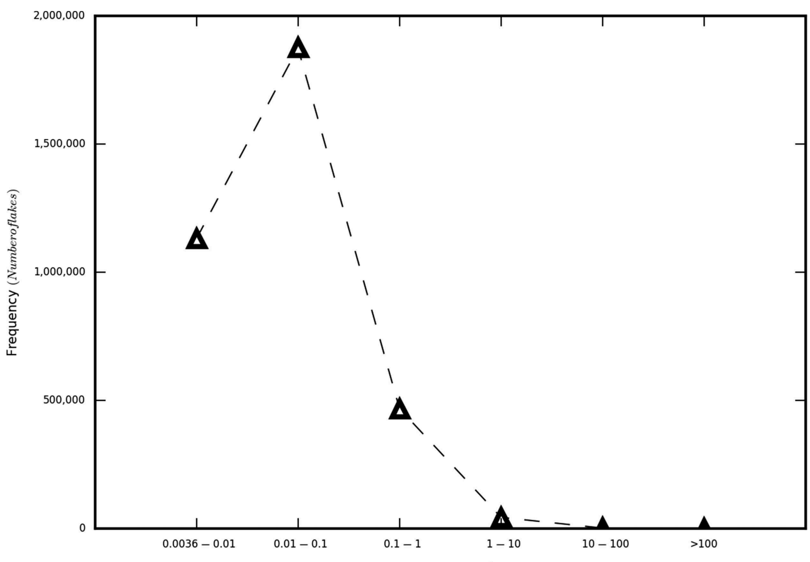

Figure 2. Relationship between lake surface area and lake frequency.

least regionally, as the size and number of lakes varies strongly spatially, depending upon features such as topography, surficial geology, and permafrost.

In our data set the majority of lakes in the Arctic are of less than 0.1 km^2 (10 ha) in extent. This matches the observations of Grosse et al. (2008) for an area of the NE Siberian coast and McDonald et al. (2012) for Alaska. Clearly, the inclusion of small size categories in the database increases the lake count considerably, and thus also lake density, over lower-resolution estimates. For the Arctic, the GLWD area estimate is 35% lower than ours. However, it defines only 70,000 lakes, none with an area <0.5 km^2 (50 ha), a 50-fold difference in lake number compared with NALGDB. Other formulated scaling laws used to estimate total freshwater area in the Arctic (including both lakes and rivers) provide an estimate of ~300,000 km^2 (Bastviken et al. 2004; Downing and Duarte 2009), whereas our estimate for lakes alone equates to 350,000 km^2. Thus, previous values obtained for lake cover in the Arctic are considerable underestimates, either through use of the previously best available data set (GLWD) or because they are derived from local studies that do not lend themselves well to upscaling. On the other hand, the estimate for all of Alaska by McDonald et al. (2012), based on incomplete data, is 1.0–2.5 million lakes, with a cover of 50,000 km^2. This compares to our estimates for Alaska (north of 65°N) of 135,000 lakes covering about 14,000 km^2. In this case, there are considerable regional differences in lake cover across Alaska, so direct comparisons are not possible.

The importance of small lakes in the Arctic

While the size distribution of Arctic lakes does not deviate from that described by a power function in either Verpoorter et al.'s (2014) database or in ours, the prevalence of small lakes in the Arctic likely reflects the peculiarities of permafrost hydrology. Lake density is correlated with permafrost distribution, and small water bodies are known to form via a range of permafrost-related processes: anastomosing polygonal ponds, thermokarst, and degradation of partially frozen peat (Schneider, Grosse, and Wagner 2009; Hinzman et al. 1991). Moreover, such lakes (here all classified as thermokarst lakes) can form rapidly after landscape disturbance, and their processes of deepening and expansion are unidirectional. They may eventually be susceptible to drainage through local thaw action or erosion (Grosse, Jones, and Arp 2013), and they have been reported as disappearing where continuous permafrost is becoming fragmented, allowing sub-surface drainage (Hinzman et al. 1991). The abundance of dynamic lakes with small circumference–area ratios implies that geochemical processes linked to bank collapse and other lake-edge processes may be particularly important to quantify more accurately (for example, methane release from thermokarst lakes; Walter et al. 2006). Furthermore, the density of these small lakes may or may not be in equilibrium with climate, and, as implied above, trends in climate-driven effects on lake dynamics may be opposite in different regions, depending, for example, on the status of permafrost. To monitor this dynamism, a flexible lake database that includes small lakes and which can be both updated with contemporary information and explored for previous decades is required. For example, in the course of this study we were able to distinguish a drained lake basin from a full lake. Thus, the NALGDB and the method used to construct it can provide this functionality.

Heterogeneity in lake distribution across the Arctic

Lake distribution exhibits clear heterogeneity both within regions and across the whole Arctic. Distributional patterns and lake size in the Arctic co-vary spatially with key landscape properties that are, to all intents and purposes, fixed, such as topography and surficial geology, but, importantly, they also co-vary with climate-sensitive features such as permafrost status.

An understanding of lake distribution patterns within landscapes is crucial for understanding Arctic hydrology. For instance, on the Canadian Shield, the geometry and spatial distribution of lakes in drainages are known to play a major role in securing the continuity of stream flow connections (Woo et al. 2008; Spence and Woo 2006). Moreover, shape attributes such as area–circumference ratios are crucial in the quantification of gas-exchange properties (Bastviken et al. 2004; Walter et al. 2006). Lakes in yedoma regions and in the boreal forest zone (not mutually exclusive) typically have high rates of carbon processing and may emit considerable amounts of methane and carbon dioxide (Walter et al. 2006; Kortelainen et al. 2006). The NALGDB has the potential to underpin an Arctic-wide lake classification system incorporating these critical features, which can be used, for example, to upscale key processes such as carbon sequestration and atmospheric gas flux.

Conclusion

We have demonstrated that Landsat imagery can successfully be used to detect water bodies in the Arctic at a continental and regional scale via a simple density-slicing technique applied to its shortwave infrared band. The NALGDB records higher overall lake cover than currently available estimates specifically focusing on the circum-Arctic region, and a lake size distribution that corresponds to a recent global estimate. The technique, with some further testing, can be transferable to other past and current Landsat imagery.

The role of Arctic freshwater ecosystems in landscape and biogeochemical processes is still poorly understood. The NALGDB provides an important research tool in regard to Arctic lakes as a component of the Earth's system. This comprehensive description of lake distribution and size for the Arctic can lead to a better quantification of the response of lakes to climatic processes: for example, their function as carbon sinks via their sediment deposition and the balance between carbon sequestration and emission via greenhouse gas release. The development of an historical archive using this technique would advance the study of decadal trends in features such as the location, shape, and abundance of lakes, and of ice phenology, and it would provide a useful basis for upscaling a range of lake-related hydrological processes.

Acknowledgements

The Landsat programme and USGS have made the data set with which this study was developed publically available. All data discussed in this article are freely available by contacting the corresponding author. We greatly appreciate Jed Kaplan's provision of an unpublished, updated version of an Arctic vegetation map. We thank the two anonymous reviewers who helped improve the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Supplemental data

Supplemental data for this article can be accessed here.

References

Bastviken, D., J. Cole, M. Pace, and L. Tranvik. 2004. "Methane Emissions from Lakes: Dependence of Lake Characteristics, Two Regional Assessments, and a Global Estimate." *Global Biogeochemical Cycles* 18: 1–12. doi:10.1029/2004GB002238.

Brown, J., O. J. Ferrians Jr, J. A. Heginbottom, and E. S. Melnikov. 2011. *Circum-Arctic Map of Permafrost and Ground Ice Conditions*. Boulder, CO: National Snow and Ice Data Center. http://nsidc.org/data/docs/fgdc/ggd318_map_circumarctic/.

Carroll, M. L., J. R. G. Townshend, C. M. DiMiceli, T. Loboda, and R. A. Sohlberg. 2011. "Shrinking Lakes of the Arctic: Spatial Relationships and Trajectory of Change." *Geophysical Research Letters* 38 (20). doi:10.1029/2011GL049427.

Cole, J. J., Y. T. Prairie, N. F. Caraco, W. H. McDowell, L. J. Tranvik, R. G. Striegl, and C. M. Duarte. 2007. "Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget." *Ecosystems* 10: 171–184. doi:10.1007/s10021-006-9013-8.

Collins, M., J. R. Knutti, J.-L. Arblaster, T. Dufresne, P. Fichefet, X. Friedlingstein, and W. J. Gao. 2013. "Long-Term Climate Change: Projections, Commitments and Irreversibility." In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Joussaume, A. Mokssit, K. Taylor, and S. Tett. Cambridge, UK: Cambridge University Press.

Downing, J. A., and C. M. Duarte. 2009. "Abundance and Size Distribution of Lakes, Ponds, and Impoundments." In *Encyclopedia of Inland Waters*, edited by G. E. Likens. Boston, MA: Elsevier.

Downing, J. A., Y. T. Prairie, J. J. Cole, C. M. Duarte, L. J. Tranvik, R. G. Striegl, and W. H. McDowell. 2006. "The Global Abundance and Size Distribution of Lakes, Ponds, and Impoundments." *Limnology and Oceanography* 51 (5): 2388–2397. doi:10.4319/lo.2006.51.5.2388.

Foody, G. M. 2002. "Status of Land Cover Classification Accuracy Assessment." *Remote Sensing of Environment* 80 (1): 185–201. doi:10.1016/S0034-4257(01)00295-4.

Frazier, P. S., and K. J. Page. 2000. "Water Body Detection and Delineation with Landsat TM Data." *Photogrammetric Engineering & Remote Sensing* 66 (12): 1461–1467. doi:0099-1112I0OI6612-1461$3.00/0.

Grosse, G., B. Jones, and C. Arp. 2013. *Treatise on Geomorphology*. Elsevier. doi:10.1016/B978-0-12-374739-6.00216-5.

Grosse, G., K. Romanovsky, A. M. Walter, H. Lantuit, and S. Zimov. 2008. "Distribution of Thermokarst Lakes and Ponds at Three Yedoma Sites in Siberia." In *Ninth International Conference on Permafrost*, 551–556. http://epic.awi.de/19086/1/Gro2008g.pdf

Hinzman, L. D., D. L. Kane, R. E. Gieck, and K. R. Everett. 1991. "Hydrologic and Thermal-Properties of the Active Layer in the Alaskan Arctic." *Cold Regions Science and Technology* 19: 95–110. doi:10.1016/0165-232X(91)90001-W.

Kanevskiy, M., Y. Shur, D. Fortier, M. T. Jorgenson, and E. Stephani. 2011. "Cryostratigraphy of Late Pleistocene Syngenetic Permafrost (yedoma) in Northern Alaska, Itkillik River Exposure." *Quaternary Research* 75 (3): 584–596. doi:10.1016/j.yqres.2010.12.003.

Kaplan, J. O., N. H. Bigelow, I. C. Prentice, S. P. Harrison, P. J. Bartlein, T. R. Christensen, and W. Cramer, et al. 2003. "Climate Change and Arctic Ecosystems: 2. Modeling, Paleodata-Model Comparisons, and Future Projections." *Journal of Geophysical Research* 108 (D19). doi:10.1029/2002JD002559.

Kortelainen, P., M. Rantakari, J. Huttunen, T. Mattsson, J. Alm, S. Juutinen, T. Larmola, J. Silvola, and P. Martikainen. 2006. "Sediment Respiration and Lake Trophic State Are Important Predictors of Large CO2 Evasion from Small Boreal Lakes." *Global Change Biology* 12 (8): 1554–1567. doi:10.1111/j.1365-2486.2006.01167.x.

Lehner, B., and P. Döll. 2004. "Development and Validation of a Global Database of Lakes, Reservoirs and Wetlands." *Journal of Hydrology* 296: 1–22. doi:10.1016/j.jhydrol.2004.03.028.

McDonald, C. P., J. A. Rover, E. G. Stets, and R. G. Striegl. 2012. "The Regional Abundance and Size Distribution of Lakes and Reservoirs in the United States and Implications for Estimates of Global Lake Extent." *Limnology and Oceanography* 57 (2): 597–606. doi:10.4319/lo.2012.57.2.0597.

Peltier, W. R. 2004. "Global Glacial Isostasy and the Surface of the Ice-Age Earth: The ICE-5G (VM2) Model and GRACE." *Annual Review of Earth and Planetary Sciences* 32: 111–149. doi:10.1146/annurev.earth.32.082503.144359.

Roach, J., B. Griffith, and D. Verbyla. 2012. "Comparison of Three Methods for Long-Term Monitoring of Boreal Lake Area Using Landsat TM and ETM+ Imagery." *Canadian Journal of Remote Sensing* 38 (4): 427–440. doi:10.5589/m12-035.

Rouse, W. R., M. S. V. Douglas, R. E. Hecky, A. E. Hershey, G. W. Kling, L. Lesack, and P. Marsh, et al. 1997. "Effects of Climate Change on the Freshwaters of Arctic and Subarctic North America." *Hydrological Processes* 11 (8): 873–902. doi:10.1002/(SICI)1099-1085(19970630)11:8<873::AID-HYP510>3.0.CO;2-6.

Schneider, J., G. Grosse, and D. Wagner. 2009. "Land Cover Classification of Tundra Environments in the Arctic Lena Delta Based on Landsat 7 ETM+ Data and Its Application for Upscaling of Methane Emissions." *Remote Sensing of Environment* 113 (2): 380–391. doi:10.1016/j.rse.2008.10.013.

Smith, L. C., Y. Sheng, and G. M. MacDonald. 2007. "A First Pan-Arctic Assessment of the Influence of Glaciation, Permafrost, Topography and Peatlands on Northern Hemisphere Lake Distribution." *Permafrost and Periglacial Processes* 18 (2): 201–208. doi:10.1002/ppp.581.

Smith, L. C., Y. Sheng, G. M. MacDonald, and L. D. Hinzman. 2005. "Disappearing Arctic Lakes." *Science* 308 (June): 1429. doi:10.1126/science.1108142.

Smol, J. P., A. P. Wolfe, H. J. B. Birks, M. S. V. Douglas, V. J. Jones, A. Korhola, and S. Sorvari, et al. 2005. "Climate-Driven Regime Shifts in the Biological Communities of Arctic Lakes." *Proceedings of the National Academy of Sciences of the United States of America* 102 (12): 4397–4402.

Spence, C., and M.-K. Woo. 2006. "Hydrology of Subarctic Canadian Shield: Heterogeneous Headwater Basins." *Journal of Hydrology* 317 (1–2): 138–154. doi:10.1016/j.jhydrol.2005.05.014.

Tranvik, L. J., J. A. Downing, J. B. Cotner, S. A. Loiselle, R. G. Striegl, T. J. Ballatore, and P. Dillon, et al. 2009. "Lakes and Reservoirs as Regulators of Carbon Cycling and Climate." *Limnology and Oceanography* 54 (6): 2298–2314. doi:10.4319/lo.2009.54.6_part_2.2298.

Verpoorter, C., T. Kutser, D. A. Seekell, and L. J. Tranvik. 2014. "A Global Inventory of Lakes Based on High-Resolution Satellite Imagery." *Geophysical Research Letters* 41 (18): 2014GL060641. 6396–6402. doi:10.1002/2014GL060641.

Walter, K. M., M. E. Edwards, G. Grosse, S. Zimov, and F. S. Chapin. 2007. "Thermokarst Lakes as a Source of Atmospheric CH4 during the Last Deglaciation." *Science* 318: 633–636. doi:10.1126/science.1142924.

Walter, K. M., L. C. Smith, and F. S. Chapin. 2007. "Methane Bubbling from Northern Lakes: Present and Future Contributions to the Global Methane Budget." *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 365 (May): 1657–1676. doi:10.1098/rsta.2007.2036.

Walter, K. M., S. A. Zimov, J. P. Chanton, D. Verbyla, and F. S. Chapin. 2006. "Methane Bubbling from Siberian Thaw Lakes as a Positive Feedback to Climate Warming." *Nature* 443 (September): 71–75. doi:10.1038/nature05040.

Woo, M., D. L. Kane, S. K. Carey, and D. Yang. 2008. "Progress in Permafrost Hydrology in the New Millennium." *Permafrost and Periglacial Processes* 19 (2): 237–254. doi:10.1002/ppp.613.

Zhang, T. 2005. "Spatial and Temporal Variability in Active Layer Thickness over the Russian Arctic Drainage Basin." *Journal of Geophysical Research* 110 (D16): D16101. doi:10.1029/2004JD005642.