A New Canadian Lake Database: Estimates of Carbon Accumulation in Canadian Boreal Lakes and New Thematic Products

by

Jamie Alexander MacGregor
B.Sc., University of Victoria, 2007

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the School of Earth and Ocean Science

©Jamie Alexander MacGregor, 2011
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

Supervisory Committee

A New Canadian Lake Database: Estimates of Carbon Accumulation in Canadian Boreal Lakes and New Thematic Products

by

Jamie Alexander MacGregor
B.Sc., University of Victoria, 2007

Supervisory Committee

Dr. Kevin Telmer, Department of Earth and Ocean Science
Supervisor

Dr. Katrin Meissner, Department of Earth and Ocean Science
Departmental Member

Dr. Terri Lacourse, Department of Biology
Outside member

Abstract

Supervisory Committee

Dr. Kevin Telmer, Department of Earth and Ocean Science
Supervisor

Dr. Katrin Meissner, Department of Earth and Ocean Science
Departmental Member

Dr. Terri Lacourse, Department of Biology
Outside member

Lake size is a strong control on lake function and on how lakes interact with the environment. For example, lake size is related to carbon burial rates in lake sediments. Lake size distribution (the number of small, medium, and large lakes per unit area) can be used to extrapolate lake function to landscapes at local, regional and global scales. This research examined the utility of using radar satellite imagery (ALOS PALSAR) and existing spatial data (CanVec) for the construction of a new Canadian lake database, which was then used to estimate carbon accumulation in Canadian boreal lake sediments.

The capability of ALOS PALSAR images for classifying lakes from eight pilot regions across Canada was assessed by direct comparison to existing CanVec data. The PALSAR lake classification differed between -1.8% to 18.0% for overall lake area and -56.0% to 196.0% for overall lake count compared to CanVec. The wide range in difference can be explained by limitations in resolution, classification method, and how a lake was defined. While the temporal resolution of PALSAR was superior, it did not provide better spatial resolution and accuracy than existing datasets. PALSAR's utility therefore is in short term change determination. Consequently, CanVec was used to construct the final database describing lake distribution in Canada, resulting in over 13.2 million features with a total area of almost 1.2 million km^2. Lake database results suggest that the scaling rules used in previous studies to estimate the number of very small lakes regionally and globally have limits. The use of real lake data allowed for a better understanding of regional differences in lake distribution across Canada that was not possible with scaling rule approaches.

Estimates of carbon accumulation in boreal Canada lake sediments based on the new CanVec lake distribution and literature-based accumulation rates ranged from 1.65 and 2.34 Mt C yr^{-1}, or roughly equal to the carbon emissions of 300,000-450,000 cars per year. Similarly, it would require only 36 years for Canada's total annual emissions to account for all the carbon accumulation in Canadian boreal lakes over the Holocene (last 10,000 years). Thematic products derived from the lake database suggest that number of lakes is more important than the distribution of small, medium and large lakes when estimating carbon accumulation in the lake sediments of boreal Canada.

Table of Contents

List of Tables

List of Figures

Acknowledgments

This thesis would not have been possible without the support of many people. I am grateful to my fellow graduate students and Aqueous Geochemistry Lab members: Pete Crockford, Ricardo Rossin, and Daniel Stapper. Throughout the highs and lows of graduate studies, you guys have been around to share a laugh, and to lend a supporting hand.

I also have to thank my supervisor Kevin Telmer. Kevin is responsible for many of the original ideas and their implementation found in this work. I am grateful for his support, feedback, and great stories that have all helped to make this thesis successful. I would also like to thank Maycira Costa. Maycira was one of my favourite undergraduate professors and it was only by her recommendation that I had the opportunity to work with Kevin and to undertake graduate studies. I would also like to thank JAXA for their support through the K&C Initiative, who provided the satellite imagery necessary for much of my analysis.

Finally, I would like to thank my parents Donald and Colleen. They have provided unconditional support over the last 3 years and without them, none of this would have been possible.

CHAPTER ONE

1.0 Introduction

Much effort has been made into understanding the components of the carbon cycle and where potential sinks exist. The carbon cycle is composed on four major reservoirs: the terrestrial biosphere, hydrosphere, atmosphere and lithosphere, connected through a series of pathways and exchange processes (Holmén 1992). The role of inland aquatic systems in the carbon cycle is commonly ignored because they represent a relatively small fraction of the planet. When considered, inland waters were traditionally thought of as simple sealed pipes for transporting terrestrial derived carbon to the oceans. More recently this view has been modified to include their role as active sites of CO_2 evasion to the atmosphere and carbon accumulation in sediments (Cole et al. 2007).

Carbon accumulation in lake sediments has been identified as an important secondary sink to terrestrially derived carbon (Mulholland and Elwood 1982, Dean and Gorham 1998, Kortelainen et al. 2004, Squires et al. 2006). Most studies which estimate carbon accumulation in lake sediments are usually restricted to a small number of lakes in a single geographic area because of the large amount of sampling effort required. As such, in order to scale estimates up to regional and global scales, studies depend on lake census data to extrapolate rates. Lake census data comes from a variety of sources and can vary greatly in both spatial and temporal resolution. While these datasets sufficiently represent large lakes, they tend to underestimate the number of small lakes and invariably rely on scaling rules to estimate small lake distribution (Maybeck 1995, Kalff 2002, Downing et al. 2006). This is a critical shortfall considering that smaller lakes have disproportionately high carbon accumulation rates, while also being the most susceptible to changes in climate (Kortelainen et al. 2004, Smith et al. 2005, Riordan et al. 2006, Benoy et al. 2007).

In Canada, CanVec produced by Natural Resources Canada (NRCan) is the most up-to-date and freely-available spatial data source representing hydrographic features, including lakes. The CanVec dataset has good spatial accuracy for most areas. However, its incomplete coverage in the north (such as the high arctic) and coarse temporal resolution (spanning 50+ years for some areas) potentially limit its application in areas where changes are occurring at fine spatial and temporal scales (Stow et al. 2004, Smith et al. 2005, Riordan et al. 2006). A digital spatial database describing lake area, location and size would help to quantify the current and future state of lakes across Canada. This would not only improve the accuracy of studies that rely upon lake data for extrapolation- such as carbon accumulation in lake sediments, but would also help improve our understanding of what controls lake distribution.

The goal of this research was to quantify the distribution of lakes regionally across Canada, and to use this to estimate carbon accumulation in Canadian boreal lake sediments using literature-based accumulation rates. An additional goal was to explore the utility of new satellite-based radar remote sensing data for constructing a lake database. However, challenges with image classification and delays in receiving the appropriate data in a timely manner resulted in the use a superior product (CanVec) for constructing the final lake database. This may change in the future with the introduction of higher resolution remote sensing products. The new CanVec derived lake database provided an accurate assessment of lake distribution in Canada along with new estimates of total carbon stock and annual accumulation rates and allowed for a better understanding of how Canadian lakes contribute to the carbon cycle and how this may change with future shifts in climate and human activity.

1.1 Objectives

The four main objectives of this study were to:

1. Quantify the utility of ALOS PALSAR and CanVec data for the construction of a lake database.
2. Construct a lake database describing the size, location, frequency, and density of lakes across Canada using CanVec.
3. Calculate new estimates of carbon accumulation in lakes for boreal Canada using the lake database and literature-based accumulation rates.
4. Explore the utility of spatial thematic products derived from the database for applications such as expressing regional changes in lake distribution and carbon accumulation over time.

These are presented in two chapters. The first (Chapter 2) examines the utility of PALSAR and CanVec for the construction of a lake database for Canada. The second (Chapter 3) focuses on the construction of a lake database from CanVec, using this database to estimate carbon accumulation in Canadian boreal lake sediments, and deriving new spatial thematic products based on lake distribution.

CHAPTER TWO

2.0 Estimates of Lake Size Distribution for Select Pilot Sites Across Canada Derived from Synthetic Aperture Radar

2.1 Abstract

Studies examining lake functions generally rely on lake census data to scale up processes to regional and global scales. Most historical lake distribution estimates have relied upon a synthesis of both aspatial (tabular) and spatial data sources collected by different methods over wide ranging time periods. Similarly, these studies have relied upon scaling rules to estimate the distribution of the very small lakes, which in many cases are the most important and susceptible to change over time. This chapter examines the utility of ALOS PALSAR data for classifying lakes and constructing a new spatially and temporally constrained database describing lake distribution across Canada. The PALSAR lake classification differed by -1.8% to -18.0% for overall lake area and -56.0% to 196.0% for overall lake count compared to the existing CanVec data across the eight pilot sites. This wide range in accuracy can be explained by limitations in resolution, water-land classification method, and how a lake is defined. While the temporal accuracy of PALSAR is much higher than existing datasets, it is unlikely that it will be able to provide a more accurate estimate of lake distribution in Canada than the CanVec dataset. The utility of PALSAR for classifying lakes therefore will likely be restricted to regions where CanVec coverage is lacking and in detecting changes over short time scales in localized areas.

2.2 Introduction

The global distribution of lakes has long been an important question. One of the first comprehensive lake studies at a global scale was completed by Herdendorf (1982) for lakes exceeding 500 km^2. Maybeck (1995) extended these estimates to include all lakes greater than 0.01 km^2 by incorporating additional local and regional census data and scaling rules. Lake density tends to increase tenfold when going from large lakes to small lakes on a logarithmic scale. Maybeck (1995) used such scaling rules to extrapolate values for small lakes. This can be expressed mathematically as a power function in the form of:

$$y = bx^m \tag{1}$$

where y is the number of lakes for a given size x, b is the intercept and m is the slope. When expressed as log units, the function becomes linear.

$$\log(y) = m \log(x) + \log(b) \tag{2}$$

Average lake area within each log size class is roughly 2.5 times the lower class bound (e.g. 1-10 ha, average lake size is 2.5 ha) (Maybeck 1995, Lehner and Döll 2004, Downing et al. 2006). If scaling rules hold, a slope value (m) of -1 indicates a tenfold increase in lake count from large lakes to the next lower size class and an equal amount of lake area across all size class. A more negative slope (<-1) indicates that small lakes account for more surface area than large lakes, whereas a less negative (>-1) slope indicates that large lakes account for more surface area than small lakes.

Maybeck (1995) was one of the first to extend global lake estimates to include small lakes (0.01 km^2) using scaling rules, but by his own estimation, was likely < 50% accurate for the smallest lakes because of limited data. Furthermore, Maybeck's (1995) data was tabular or aspatial in nature, and did not show how lake distribution varied spatially from region to region. More recently, Lehner and Döll (2004) attempted to improve global lake estimates by incorporating spatial datasets from geographic information systems (GIS) and remote sensing sources into their Global Lakes and Wetlands Database (GLWD), producing a spatially explicit database describing lakes > 0.1 km^2 as well as estimates of lake count and area down to 0.01 km^2 and 0.001 km^2 using scaling rules. The latest estimate of global lake distribution was produced by Downing et al. (2006). Using data from Lehner and Döll (2004) and new GIS sources, Downing et al. (2006) modified scaling rules to extend estimates for lakes down to 0.001 km^2 in size.

A common feature of all of these studies is that they all to some extent rely on scaling to estimate the smallest lakes, and are generally limited in their spatial and/or temporal resolution. This is because these studies depend on incomplete data from multiple sources spanning multiple time frames. In Canada, the CanVec dataset produced by NRCan constitutes one of the most comprehensive and freely available empirical datasets describing Canadian lakes (CanVec Hydrographic Dataset 2007). CanVec displays excellent spatial resolution and includes features less than 0.001km^2 in size, negating the need for power law based lake estimates. However, similar to global studies, CanVec includes some regions with relatively coarse temporal resolution, spanning up to

50 years or more. This may limit its capacity to detect more recent changes in lake distribution that are occurring on annual or decadal scales.

A new approach to quantifying lake distribution was recently explored by Telmer and Costa (2007). They used radar remote sensing data to extract estimates of lake count and areal extent. Using the same methods here, the goal of this chapter was to examine the utility of new L-band Synthetic Aperture Radar (SAR) data for classifying lakes over various regions of Canada. This was done as a pilot study in anticipation of using the same lake classification method on a Canada-wide 50 m resolution mosaic produced by the Jet Propulsion Laboratory (JPL) under the ALOS Kyoto and Carbon (K&C) Initiative global monitoring project supported by the Japanese Aerospace Exploration Agency (JAXA) (Telmer et al. 2008). As of this writing, this mosaic has yet to be completed, but as results show, it would not have produced a database superior to CanVec.

2.3 Study Area and Data

2.3.1 Study Area

The Canadian landmass accounts for an area of over 9.9 million km^2 and covers a wide variety of physiographic regions from the arctic in the north, to boreal forest, prairies, and maritime regions in the south. In Canada, fifteen distinct terrestrial ecozones have been used to describe broad differences in landforms, ecology, geology, and climate (Wiken 1986). In the north, the arctic zones are dominated by low precipitation, sparse vegetation, permafrost, and post-glacial tundra ponds and lakes. South of the arctic, the boreal region is Canada's largest ecozone, and is loosely defined by the extent of its forests, spanning from coast to coast, bounded by tree line to the north, and the prairies and Montane/Cordillera regions to the south and west. This region accounts for 77% of Canada's forested land and is composed of a variety of conifer and deciduous trees including multiple species of pine and spruce (Wiken 1986).

2.3.2 ALOS PALSAR

The Phased Array L-band Synthetic Aperture Radar (PALSAR) sensor is part of the Advanced Land Observing Satellite (ALOS) platform launched in 2006 by JAXA and consists of a fully Polarimetric (HH, HV, VH, VV) L-band (23.5 cm/1270MHz) radar imaging sensor operating at 14 and 28 MHz bandwidths (Rosenqvist et al. 2007). The PALSAR sensor has five main observation modes: fine beam single polarization (FBS), fine beam dual polarization (FBD), polarimetric mode (POL), ScanSAR mode, and Direct Transmission (DT). These observation modes allow for a wide range of applications such as monitoring wetlands and forests to mapping subsurface geology (Paillou et al. 2010, Almeida-Filho et al. 2009). In this study, imagery consisted of FBS standard georeferenced images. Each 16-bit image covered an area of approximately 70x70 km in a single polarization (HH) channel with a nominal spatial resolution of 12.5 m. All images for this study were acquired during the summer months of June through September of 2006, 2007 and 2008 directly from JAXA under the ALOS K&C Initiative global monitoring project (ALOS K&C Initiative 2007).

Pixel values in radar images represent the magnitude of microwave energy returned to the sensor, commonly referred to as backscatter. Typically, backscatter strength is a function of target characteristics, the wavelength in use, and viewing geometry (Jenson 2006). Textured or rough surfaces such as vegetation or soil generally scatter incoming radar waves equally in all directions resulting in low to moderate signal return (Raney 1998). Generally, as surface roughness increases, backscatter brightness also increases. In contrast, smooth surfaces such as calm water results in specular or forward scattering. Combined with the off nadir acquisition angle of radar systems, this result in a very low backscatter return to the sensor from water bodies. This inherent characteristic of low signal return from water surfaces forms the basis of this study. Environmental factors such as wind, emergent vegetation, and ice cover can all add a level of roughness to an otherwise smooth surface (Raney 1998). This can be minimized using longer wavelength microwave energy. Typically a surface will appear smoother as wavelength increases. For a given surface, L band (λ = 23.5 cm) radar is less sensitive to minor height variations compared to other radar bands such as C and X bands (λ = 6 cm and λ = 3 cm respectively). Limitations of radar include slant range scale distortion and foreshortening. This occurs when the horizontal scale is compressed as a function of topography and off nadir image acquisition. Furthermore, layovers can occur in regions with steep topography where the signal return from the top of an object returns to the senor before the signal from the base. Both of these scenarios can result in shadows or areas of no data in the image.

2.3.3 CanVec, Landsat, SPOT 4/5, and AVNIR-2 Data

CanVec is a relatively new and free cartographic reference product produced by NRCan. CanVec aims to accurately represent topographic entities across the Canadian landmass including hydrographic data. CanVec is produced from the best available data sources, derived mainly the National Topographic Database (NTDB), the Geobase initiative, and Landsat imagery (Figure 2.1A.) (CanVec Hydrographic Dataset 2007). Spatial accuracy complies with georeferencing standards used by NRCan for Landsat orthoimages, ranging from 15 to 30 m for most areas. Temporal resolution varies greatly, ranging from 1946 to 2009 depending on the location (Figure 2.1 B). CanVec's coverage is good for most of the Canadian landmass with gaps only in the most northern latitudes.

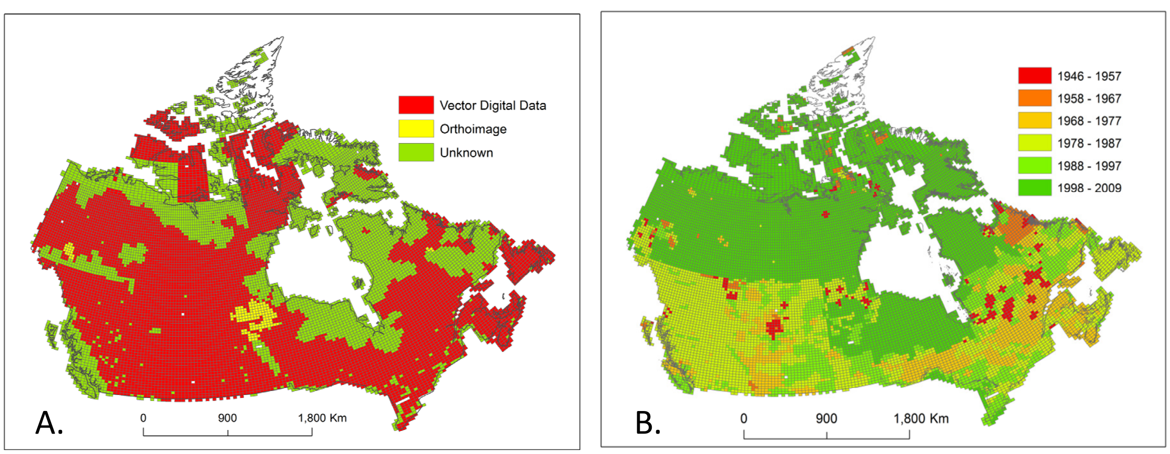

Figure 2.1. CanVec coverage is nearly complete for all of Canada with exception of the most northern latitudes. (A) Shows the source of the data when known. Vector digital data refers to digitized map features, orthoimage refers to georeferenced satellites images and unknown refers to unknown/unclassified data source. (B) Shows the date range associated with the data used to create the features in CanVec.

The Landsat 7 satellite launched by the National Aeronautics and Space Administration (NASA) in 1999 consists of an 8 band sensor with a nominal resolution of 15, 30 and 60 m covering panchromatic, multispectral and thermal ranges with image sizes of 183x170 km (Jenson 2005). SPOT 4 and 5 satellites were launched in 1998 and 2002, respectively, by the French Space Agency and consist of 60x60 km panchromatic, multispectral and thermal bands at 10 and 20 m nominal resolutions. Both Landsat and SPOT images were acquired free from the GeoBase portal under the Canadian Council on Geomatics (CCOG) and NRCan. Acquisition dates for Landsat and SPOT images ranged between 1999 and 2003 and 2005 and 2010, respectively, and were available for most of the Canadian landmass. Planimetric accuracy ranged from 20 m in the south to 30 m in the north.

The Advanced Visible and Near Infrared Radiometer type 2 (AVNIR-2) sensor on board the ALOS remote sensing platform was launched by JAXA in 2006 and consists of a multispectral radiometer with a 10 m nominal resolution and 70 km swath width. Having AVNIR-2 and PALSAR sensors on the same remote sensing platform allowed for the near simultaneous acquisition of images and better direct comparisons of PALSAR lake classifications to time synchronous AVNIR-2 images. All AVNIR-2 images where acquired directly from JAXA under the ALOS K&C Initiative global monitoring project (ALOS K&C Initiative 2007).

2.4 Methods

2.4.1 PALSAR Classification Strategy

High resolution PALSAR FBS HH 12.5 m images were acquired from JAXA under the K&C agreement for eight pilot sites across Canada and were selected to cover a variety of physiographic conditions (Figure 2.2).

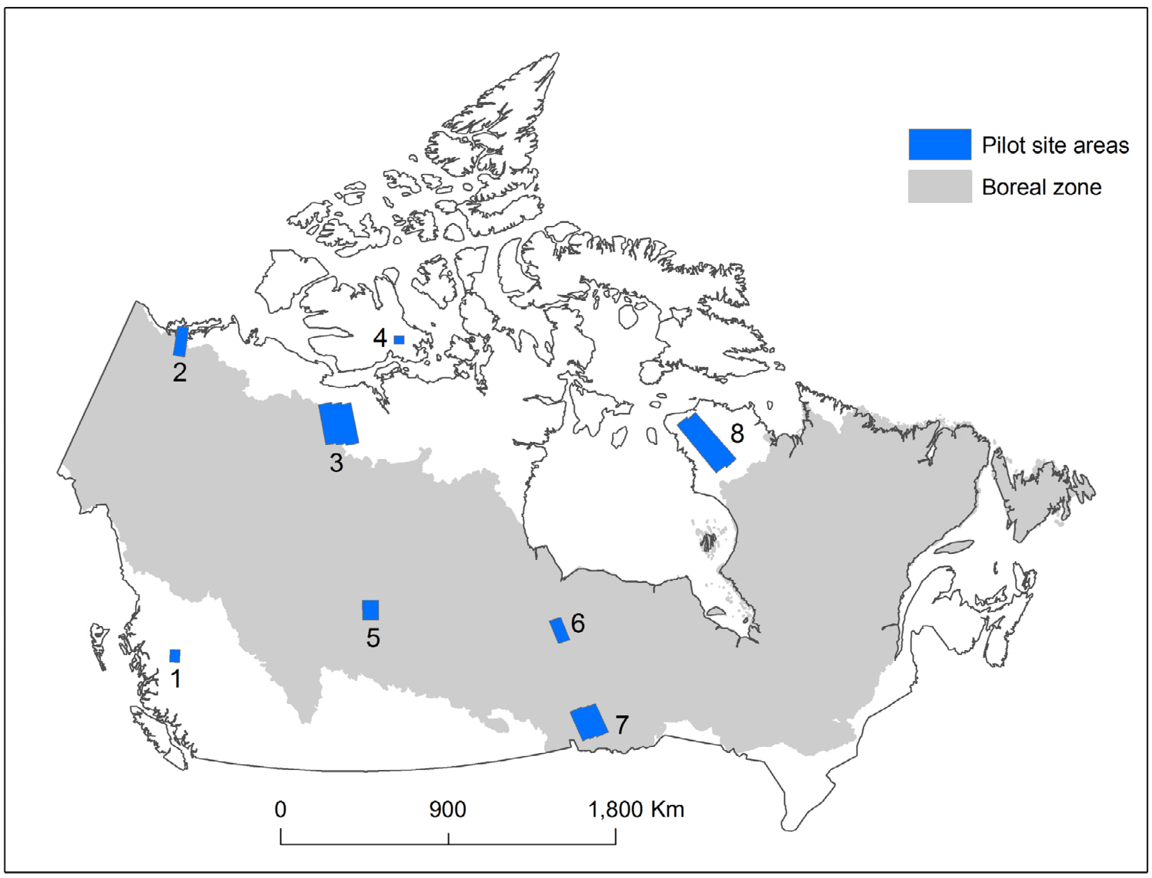

Figure 2.2. The location of eight pilot sites throughout Canada where PALSAR images were acquired. (1) central British Columbia (BC), (2) Mackenzie Delta, (3) central Northwest Territories (NWT), (4) Victoria Island, (5) northern Alberta, (6) north-eastern Manitoba, (7) Experimental Lakes Area (ELA) Ontario, and (8) north-eastern Quebec.

Data was processed using the Alaska Satellite Facility processing tools to output 16-bit raw amplitude images. Each pilot site consisted of 2 to 16 images that were mosaiced together and filtered using a 3x3 pixel Enhanced Lee speckle filter in PCI Geomatica® remote sensing software to minimize inherent noise and strengthen contrast between water and land. Before a threshold classification was applied, a preliminary exploration of histograms revealed a characteristic bi-modal distribution. Digital number (DN) values within the first mode were assumed to represent water and DN values in the second mode were assumed to represent land. This was verified by querying pixels of known cover within each image (Figure 2.3).

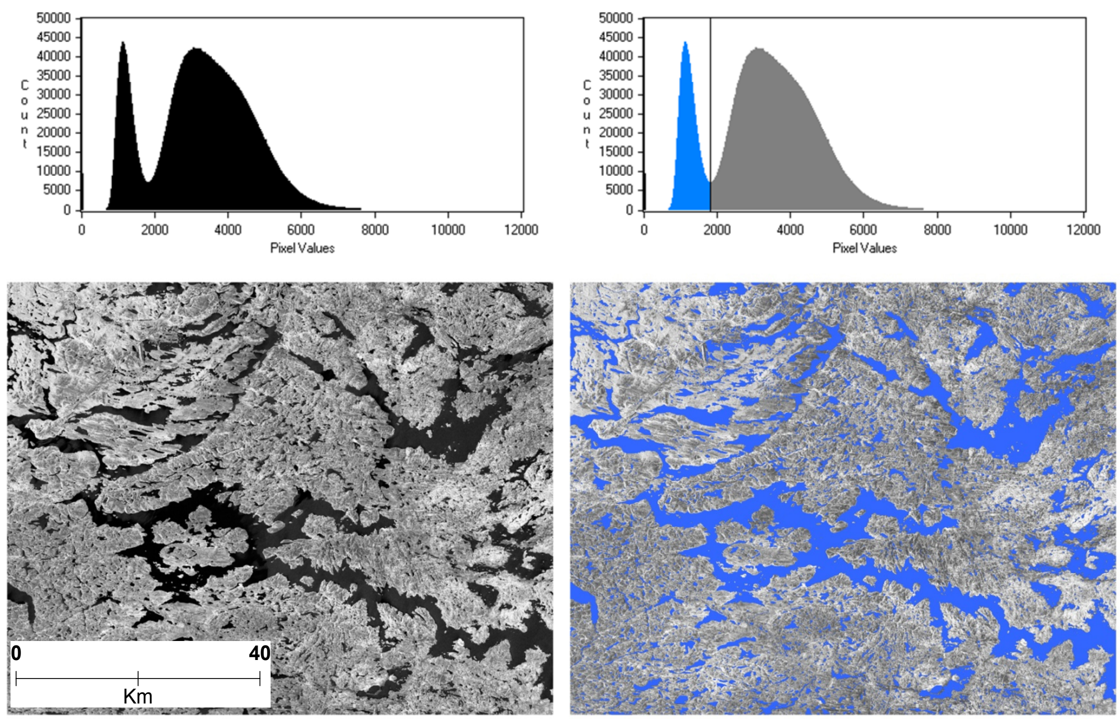

Figure 2.3. Left: A subset of a FBS HH PALSAR image and histogram for the central NWT site. Right: Applying a threshold between the two modes results in the classification of water bodies (blue).

Histograms showed considerable differences in where and how quickly the distributions transitioned between modes. In regions such as northern Alberta and central BC where relatively less surface water was present, transitions between modes were broad and less pronounced (Figure 2.4). Consequently, a single threshold did not produce acceptable results for all pilot sites. In an effort to empirically derive where the cut-off between lakes and land should be placed, derivatives were calculated for each distribution. Derivative values of zero represented inflection points between modes and provided a logical starting point for subsequent tuning.

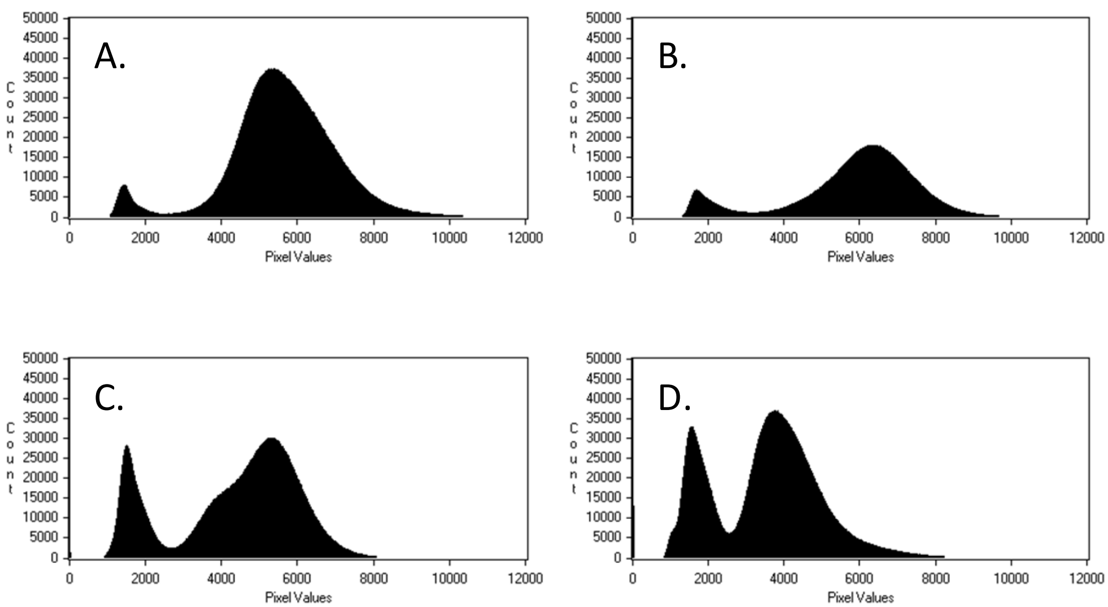

Figure 2.4. Examples of bimodal histograms from four of the eight pilot sites. (A) northern Alberta, (B) central BC, (C) north-eastern Manitoba and (D) central NWT. Sites A and B have relatively low surface water counts and consequently have very broad transitions between modes. In contrast, sites C and D with high counts of surface water have relatively narrow transitions between modes and required little to no tuning of the threshold classification.

Preliminary threshold values were applied using the EASI Modeller in PCI Geomatica® to produce a binary classification in a new 8-bit channel. The results were then visually inspected and compared with CanVec, Landsat, SPOT, and AVNIR-2 imagery to determine how well water bodies were classified. Although the derivative method provided a good starting point for classification, several sites had poor results and had to be manually tuned. Once acceptable thresholds were established, the classifications were converted into polygon Shapefiles and exported for further analysis in ESRI ArcGIS® software.

2.4.2 Accuracy Assessment

In ArcGIS®, PALSAR lake size classes were compared with CanVec vector data to understand classification accuracy. Initial assessment revealed a considerable overestimation by PALSAR of very small lakes in the range of 1 to 16 pixels (0.0156 to 0.25 ha) largely attributable to residual noise and speckle inherent to radar data. Consequently, a sieve filter had to be applied with a cut off of 16 pixels (0.25 ha), which removed holes within features and erroneous small polygons.

2.5 Results

2.5.1 PALSAR Vs. CanVec

The PALSAR lake classification underestimated total lake area for all eight pilot sites when compared to the CanVec (Table 2.1). Between sites, difference in total lake area varied from -1.8% in BC to -18.0% on Victoria Island. Similarly, within each region, estimates varied between overestimation and underestimation across size classes. A closer look at some of these discrepancies revealed that in some cases, adjacent size classes account for roughly the same amount of difference but in opposite directions. For example, in northern Alberta size classes 1-10 ha and 10-100 ha have differences of -147.0 ha and 145.1 ha respectively. Similarly, in the Mackenzie Delta the size classes 1,000-10,000 ha and 10,000-100,000 ha show differences of -29,876.1 ha and 25,868.9 ha, respectively.

Table 2.1. A comparison of lake area between the PALSAR classification and CanVec data for eight pilot sites across Canada

		Lake size range (ha)							
Site	Lake area (ha)	<1	1 - 10	10 - 100	100 - 1,000	1,000 - 10,000	10,000 - 100,000	Total	Limnic Ratio*
Central BC	CanVec	120.9	1,070.6	3,690.6	6,352.4	2,738.9	22,821.7	36,795.2	10.4
	PALSAR	265.1	1,832.1	4,783.1	5,316.6	2,420.5	21,502.8	36,120.2	10.2
	Difference	144.2	761.5	1,092.5	-1,035.9	-318.4	-1,318.9	-675.0	
	Difference (%)	119.2	71.1	29.6	-16.3	-11.6	-5.8	-1.8	
Mackenzie Delta	CanVec	2,603.2	25,540.0	65,451.5	54,353.2	54,211.2	24,833.8	226,993.0	23.6
	PALSAR	2,511.5	20,366.2	58,728.8	47,887.6	24,335.1	50,702.8	204,532.1	21.2
	Difference	-91.6	-5,173.7	-6,722.7	-6,465.6	-29,876.1	25,868.9	-22,460.9	
	Difference (%)	-3.5	-20.3	-10.3	-11.9	-55.1	104.2	-9.9	
Central NWT	CanVec	11,564.9	76,364.4	165,112.5	174,880.1	126,245.0	286,141.5	840,308.4	22.6
	PALSAR	9,639.7	54,800.5	131,289.4	168,776.3	117,891.0	245,380.0	727,776.9	19.6
	Difference	-1,925.2	-21,563.9	-33,823.2	-6,103.8	-8,354.0	-40,761.5	-112,531.6	
	Difference (%)	-16.6	-28.2	-20.5	-3.5	-6.6	-14.2	-13.4	
Victoria Island	CanVec	2,205.5	9,151.7	13,963.2	8,284.4	16,422.8	0.0	50,027.5	21.2
	PALSAR	1,226.3	6,405.6	11,620.2	7,878.5	13,873.8	0.0	41,004.3	17.4
	Difference	-979.1	-2,746.1	-2,343.0	-405.9	-2,549.0	0.0	-9,023.1	
	Difference (%)	-44.4	-30.0	-16.8	-4.9	-15.5	0.0	-18.0	
Northern Alberta	CanVec	462.6	1,860.7	5,750.5	12,498.1	17,320.4	20,654.1	58,546.4	7.2
	PALSAR	328.0	1,713.7	5,895.6	11,841.4	15,182.2	20,551.9	55,512.7	6.8
	Difference	-134.6	-147.0	145.1	-656.7	-2,138.2	-102.2	-3,033.7	
	Difference (%)	-29.1	-7.9	2.5	-5.3	-12.3	-0.5	-5.2	
North-eastern Manitoba	CanVec	372.1	5,498.2	28,816.8	30,075.8	44,531.3	99,738.3	209,032.4	24.3
	PALSAR	362.9	5,181.3	27,606.2	32,600.0	37,468.9	94,566.4	197,785.7	23.0
	Difference	-9.1	-316.9	-1,210.5	2,524.2	-7,062.4	-5,171.9	-11,246.7	
	Difference (%)	-2.4	-5.8	-4.2	8.4	-15.9	-5.2	-5.4	
ELA	CanVec	1,060.6	15,093.3	70,909.2	107,389.1	125,931.1	207,679.4	528,062.7	21.7
	PALSAR	2,957.9	16,883.9	79,290.0	110,908.6	132,632.8	157,639.6	500,312.7	20.6
	Difference	1,897.3	1,790.6	8,380.8	3,519.5	6,701.6	-50,039.8	-27,750.0	
	Difference (%)	178.9	11.9	11.8	3.3	5.3	-24.1	-5.3	
North-western Quebec	CanVec	30,339.3	103,024.3	202,936.2	258,082.4	220,759.3	120,676.2	935,817.6	22.9
	PALSAR	18,440.1	79,031.7	182,415.6	231,562.8	178,134.5	98,966.1	788,550.7	19.3
	Difference	-11,899.2	-23,992.5	-20,520.6	-26,519.6	-42,624.8	-21,710.1	-147,266.9	
	Difference (%)	-39.2	-23.3	-10.1	-10.3	-19.3	-18.0	-15.7	

*Limnic ratio (expressed in percent) is the ratio between total lake area and total surface area of the site

Similar to area estimates, the lake classification showed a wide range of differences in lake count compared to CanVec (Table 2.2). The largest overall underestimate was Victoria Island at -56.0% while the largest overestimate was in the ELA at 196.0%. Within each region differences in lake count varied greatly. Some of the largest underestimates occurred for the smallest lakes in northern locations such as Victoria Island, Mackenzie Delta, and north-western Quebec with count differences ranging from -1,901 to -55,003.

Table 2.2. A comparison of lake count between the PALSAR classification and CanVec data for eight pilot sites across Canada

		Lake size range (ha)						
Site	Lake count	<1	1-10	10-100	100-1,000	1,000-10,000	10,000-100,000	Total
Central BC	CanVec	380	272	131	18	1	1	803
	PALSAR	529	508	188	16	1	1	1,243
	Difference	149	236	57	-2	0	0	440
	Difference (%)	39.2	86.8	43.5	-11.1	0.0	0.0	54.8
Mackenzie Delta	CanVec	6,826	7,325	2,402	250	12	1	16,816
	PALSAR	4,925	5,875	2,086	225	7	2	13,120
	Difference	-1,901	-1,450	-316	-25	-5	1	-3696
	Difference (%)	-27.9	-19.8	-13.2	-10.0	-41.7	100.0	-22.0
Central NWT	CanVec	37,224	22,627	6,076	713	60	7	66,707
	PALSAR	39,686	16,265	4,732	639	56	7	61,385
	Difference	2,462	-6,362	-1,344	-74	-4	0	-5322
	Difference (%)	6.6	-28.1	-22.1	-10.4	-6.7	0.0	-8.0
Victoria Island	CanVec	7,937	2,859	563	36	7	0	11,402
	PALSAR	2,562	1,972	440	32	6	0	5,012
	Difference	-5,375	-887	-123	-4	-1	0	-6390
	Difference (%)	-67.7	-31.0	-21.9	-11.1	-14.3	0.0	-56.0
Northern Alberta	CanVec	1,556	565	198	46	5	1	2,371
	PALSAR	700	502	209	47	4	1	1,463
	Difference	-856	-63	11	1	-1	0	-908
	Difference (%)	-55.0	-11.2	5.6	2.2	-20.0	0.0	-38.3
North-eastern Manitoba	CanVec	1,139	1,304	956	119	12	2	3,532
	PALSAR	708	1,276	914	126	9	2	3,035
	Difference	-431	-28	-42	7	-3	0	-497
	Difference (%)	-37.8	-2.2	-4.4	5.9	-25.0	0.0	-14.1
ELA	CanVec	2,784	3,784	2,330	426	45	5	9,374
	PALSAR	20,304	4,412	2,562	413	53	5	27,749
	Difference	17,520	628	232	-13	8	0	18375
	Difference (%)	629.3	16.6	10.0	-3.1	17.8	0.0	196.0
North-western Quebec	CanVec	130,231	33,451	7,278	969	91	5	172,025
	PALSAR	75,228	24,656	6,360	884	74	5	107,207
	Difference	-55,003	-8,795	-918	-85	-17	0	-64,818
	Difference (%)	-42.2	-26.3	-12.6	-8.8	-18.7	0.00	-37.7

A closer look at acquisition dates for PALSAR and CanVec shows that differences ranged from 1 to 54 years (Table 2.3). The largest differences occurred in central BC, northern Alberta, north-eastern Manitoba and the ELA. In contrast, the Mackenzie Delta, central NWT, Victoria Island and north-western Quebec all showed differences less than 10 years.

Table 2.3 A comparison of temporal resolution and acquisition dates between PALSAR and CanVec. The CanVec valid date range is based on the "VALDATE" attribute found in the CanVec data and represents the date of the data source used to create, revise or confirm an object (CanVec Hydrographic Dataset 2007).

Site	PALSAR acquisition dates	CanVec valid date range	Difference between PALSAR and CanVec (years)
Central BC	2006	1980-1988	18-26
Mackenzie Delta	2006	2000-2002	4-6
Central NWT	2006-2008	1999-2001	5-9
Victoria Island	2006	1999-2000	6-7
Northern Alberta	2006	1952-1989	17-54
North-eastern Manitoba	2006	1989-2006	0-17
ELA	2007- 2008	1969-1992	15-39
North-western Quebec	2007	2000-2006	1-7

2.6 Discussion

Taken as a whole, the PALSAR lake classification showed mixed results in estimating both lake area and count. The PALSAR lake classification underestimated overall lake area between -1.8% to -18.0% across the eight pilot sites compared to CanVec. The classification also showed a wide range of accuracy for total lake count, ranging from an underestimate of -56.0% to a large overestimate of 196.0%. Some of the differences between the PALSAR classification and CanVec can be explained by simple class shifts between adjacent size classes. This is because of differences in the way the PALSAR classification and CanVec define lakes.

In many regions of Canada lakes can have very sinuous, complex shapes with high perimeter to area ratios (Kalff 2002). Lakes with complex and narrow junction points can be broken into multiple features based on the limits of resolution in the PALSAR data and the classification technique. Although the threshold classification has some clear advantages in computational efficiency (Kozlenko and Jeffries 2000), it does a poor job of dealing with mixed pixel effects found along the edges of water bodies and in narrow spans where water meets shoreline and surrounding vegetation. As such, mixed pixels can occur along lake shores with high degrees of sinuosity and narrow sections on the same order of magnitude as the nominal resolution of the PALSAR data (12.5 m). This pixel mixing inflates the DN values causing them to be excluded from the threshold classification. This can result in a loss of lake area as well as an increase in lake count where single lakes are broken down into multiple smaller lakes (Figure 2.5). The end result is an increase in the overall lake count with a decrease in overall lake area. This not only gives different overall estimates but also changes how lakes are distributed in the size classification. Interestingly, in some cases the PALSAR classification may be a better representation of large lakes with high degrees of sinuosity because these lakes may function more like multiple small lakes with respect to processes such as carbon accumulation.

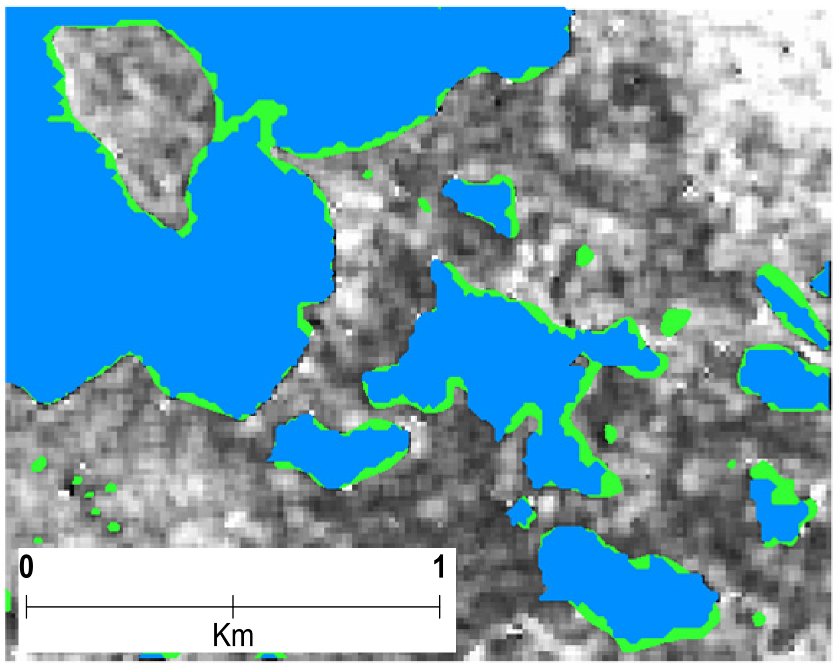

Figure 2.5. Differences between the PALSAR lake classification and the CanVec dataset from the NWT pilot site. The PALSAR classification (blue) has lost area along the edges of lakes and in some cases has split single water bodies into multiple features when compared to the CanVec lakes (green). Similarly, very small features in CanVec are completely missed by PALSAR classification as a consequence of resolution, mixed pixels, and filtering.

Another possible reason for differences between the PALSAR classification results and CanVec is radar-target interaction properties. The premise of the threshold classification is based on the smooth surface of a water body inducing specular reflection of the incoming radar waves and directing them at an opposite and equal angle away from the sensor, resulting in a low signal return or DN. Whether a surface appears smooth or rough to a sensor depends on three parameters: incident angle, wavelength, and surface irregularities (Raney 1998). In the case of PALSAR, both incident angle and wavelength are fixed, leaving surface irregularities as the lone variable. Generally, a surface is considered smooth and results in specular reflection if height variations are considerably smaller than the incident wavelength (Raney 1998). The PALSAR sensor is as an L-band sensor and uses a relatively long wavelength of 23.5 cm. This means that any features with surface roughness on the order of 23.5 cm or smaller are likely to result in varying degrees of specular reflection and low signal returns back to the sensor. In urban areas anthropogenic features such as farmer's fields, paved areas (roads and parking lots), and clear cuts tend to have low levels of surface roughness for an L-band sensor. Consequently, this can result in threshold classification errors in these areas and inflate both lake count and area estimates (Figure 2.6).

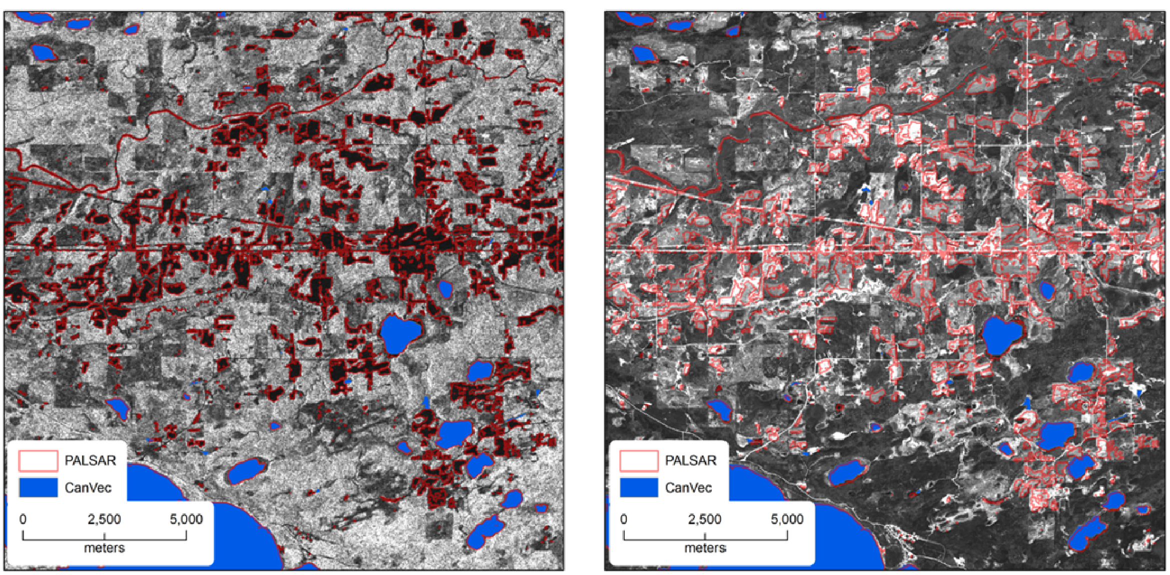

Figure 2.6. A sub-section of the ELA region showing the difference between CanVec and the PALSAR classification. The base image on the left is the PALSAR image for which the lake classification was based on. The base image on the right is a corresponding SPOT panchromatic image. Many features identified by the PALSAR classification (left image, dark areas outlined in red) turn out to be anthropogenic features such as roads, farmer's fields and new clear cuts when examined closely in alternate high resolution imagery.

When determining the accuracy of the PALSAR classification one also has to consider the accuracy of the CanVec data. While every effort was made to validate the accuracy of CanVec, there are some important limitations that need to be considered. CanVec generally displays excellent spatial resolution. However, its biggest shortfall is its temporal resolution, ranging up to 50+ years for some areas (Figure 2.1 B, Table 2.3). In contrast, PALSAR data represents a more current and fine temporal resolution on the order of several summer months (June through September) in 2006, 2007and 2008. In theory, this advantage should allow PALSAR to more accurately represent the current extent and number of lakes regionally and provide a better baseline for future changes. Most existing lake datasets including CanVec represent a synthesis of multiple sources collected using various methods from different time periods which allows for the risk of inter-study bias. The PALSAR data represents a single source and acquisition method which removes any inter-study bias and strengthens conclusions on lake extent and count. However, ultimately the temporal resolution advantages of PALSAR are countered by its limitations in spatial accuracy compared to CanVec. The majority of lakes are unlikely to change over a temporal range of 50 years, making CanVec the superior product. For those that do, and in areas where CanVec coverage is limited, PALSAR may be a viable alternative for monitoring lakes.

Kozlenko and Jefferies (2000) successfully used the threshold technique to determine shallow water bathymetry and lake extent in northern thaw lakes. Although their focus was on determining bathymetry, the same classification principles were used here, and previously by Telmer and Costa (2007) to map lakes. Similar to this study, Telmer and Costa (2007) applied threshold classifications to 100 m resolution Japanese

Earth Resources Satellite (JERS-1) SAR data for study sites in Canada and Brazil. The FBS PALSAR data used in this study was 16 and 64 times higher spatial resolution compared to the proposed 50 m PALSAR mosaic and 100 m JERS-1 data. Even with a resolution advantage (12.5 m vs. 50 m and 100 m), the threshold classification still underestimates overall lake count and area when compared to existing CanVec data, especially for the smallest lakes. Given that these methods were to be used for lake classification Canada-wide using a 50 m PALSAR mosaic, the results found here indicate that the threshold classification technique is unlikely to provide better estimates of lake size distribution than what is already available in CanVec.

2.7 Conclusions

The classification of lakes in eight pilot sites across Canada achieved mixed results. The main advantages of using satellite imagery such as PALSAR for constructing a lake database are its single acquisition method and fine temporal resolution. However, these advantages were offset by limitations in spatial accuracy and image classification method. The PALSAR data used here had a nominal spatial resolution of 12.5 m, or 16 times higher than the proposed 50 m Canada-wide PALSAR mosaic for which these methods were ultimately to be applied to. Considering the results of the 12.5 m pilot areas, it is unlikely that a coarser resolution PALSAR mosaic would produce favourable results for a Canada-wide lake classification. A more advanced classification method may have provided better results; however, it is certain these methods would still have not overcome significant resolution limitations for very small lakes. 12.5 m and 50 m resolution data is simply not fine enough to reproduce a CanVec database. It is possible that a higher resolution sensor would accomplish better results, but to do so at a national scale would require a massive investment in both time and computational resources not feasible in this study.

CHAPTER THREE

3.0 Construction of a Canadian Lake Database: Carbon Estimates and Thematic Products

3.1 Abstract

Most studies examining carbon accumulation in lake sediments rely on upon lake census data to extrapolate estimates to regional and global scales. Here, CanVec hydrographic data produced by NRCan was used to construct a database describing lake distribution in Canada and to estimate carbon accumulation in Canadian boreal lake sediments. This resulted in a database with over 13.2 million lakes with a total area of 1.2 million km^2 ranging between 0.1 and 1,000,000+ ha in size. The CanVec lake database constructed here is the only lake database based entirely on empirical data and that does not rely upon scaling rules to estimate lake distribution. Compared with previous lake census studies results suggest that scaling rules commonly used to estimate the number of very small lakes within Canada and globally have limits and are likely inaccurate. Based on literature derived carbon accumulation rates and this new CanVec lake database, Canadian boreal lake sediments accumulate between 1.65 and 2.34 Mt C yr^{-1} or roughly equal to the carbon emissions of 300,000-450,000 cars per year. Similarly, it would require only 36 years for Canada's per capita emissions to account for all the carbon accumulation in Canadian boreal lakes over the Holocene (last 10,000 years). Improved understanding of lake distribution in boreal Canada suggests that carbon accumulation is higher in boreal lakes than previously thought. Results from the database show that in

boreal Canada, the total number of lakes is a more important determinant in estimating carbon accumulation than the distribution of small medium and large lakes.

3.2 Introduction

Historically, inland aquatic environments have not been adequately represented in the global carbon cycle. The traditional view of the carbon cycle consists of four major reservoirs: the terrestrial biosphere, hydrosphere (oceanic), atmosphere, and lithosphere, connected through a series of conduits and exchange processes (Holmén 1992). Inland aquatic systems such as lakes and rivers were originally thought of as simple conduits or sealed pipes, transporting carbon from land to the oceans. More recently, this sealed pipe approach has been modified to account for *in situ* carbon storage in sediments and flux to the atmosphere (Cole et al. 2007, Tranvik et al. 2009). Quantifying the role of inland waters in the storage, transport, and flux of carbon has become an important question in refining the global carbon cycle. Unlike terrestrial carbon stores such as soils and living biomass, lake sediment carbon storage is considered a long term sink holding carbon for tens of thousands of years (Cole et al. 2007). The carbon in lake sediments is primarily in the form of organic carbon (OC). The contribution of inorganic carbon (IC) is generally low and usually only occurs from weathering in areas of carbonate substrates or hydrologically closed silicate drainage basins with high rates of evaporation (Einsele et al. 2001). OC in lake sediments is generally derived from two sources, autochthonous carbon, which is a product of within-lake primary productivity, and allochthonous carbon, which is derived from primary productivity in the surrounding watershed (trees and vegetation), making lakes an important secondary sink to terrestrially derived carbon (Kalff 2002). Means of OC sedimentation is usually in the form of particulate organic carbon (POC) derived from within lake productivity (phytoplankton) and the surrounding watershed (detritus/leaf litter). Light-mediated flocculation of dissolved organic carbon (DOC) to POC also plays a role, especially in boreal lakes (von Wachenfeldt et al. 2008, von Wachenfeldt and Tranvik 2008).

Compared to other carbon stores in the boreal zone such as forest soils, peat lands, and wetlands, lake sediments are less vulnerable to climate change in the short term (Bhatti et al. 2003, Benoy et al. 2007). Evaluating how carbon accumulation and storage in boreal lakes will be affected by climate change is difficult. According to data derived from the Coupled Global Climate Model (CGCM3.1/T47) produced by the Canadian Center for Climate Modelling and Analysis (CCCma), mean annual temperature will increase ~ 2 to 4 ˚C and annual precipitation will increase ~ 0 to 50% by mid-century in boreal Canada under an A1B emission scenario (2040-2069 relative to 1961-1990) (Figure 3.1).

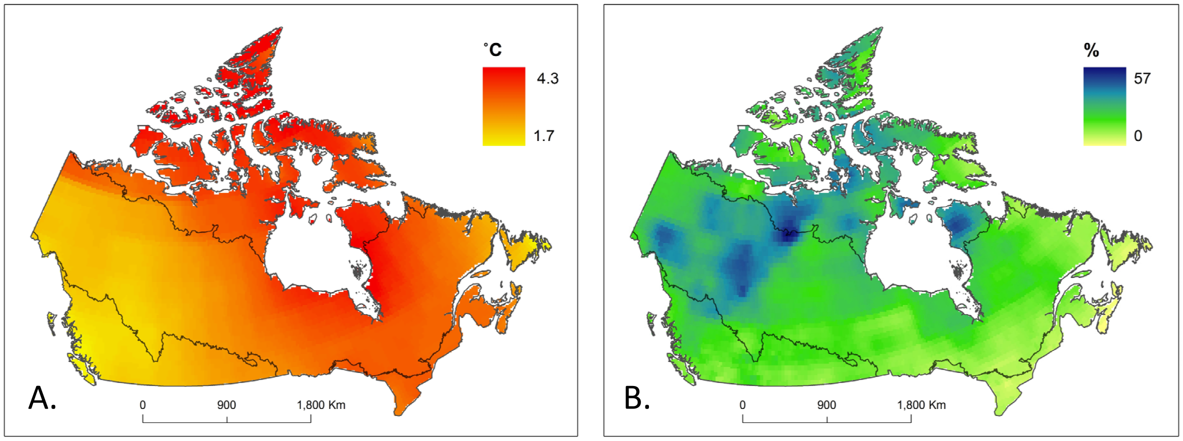

Figure 3.1. Change in (A) mean annual temperature and (B) precipitation according to the Coupled Global Climate Model (CGCM3/T47) developed by the Canadian Center for Climate Modelling and Analysis for the time period 2040-2069 relative to 1961-1990 under an A1B emission scenario. The boreal zone is outlined in black. Temperature and precipitation data used to produce these maps was retrieved from the Climate Wizard website (www.climatewizard.org).

Fine scale regional and spatiotemporal changes in temperature and precipitation are likely to occur throughout the boreal zone and result in a range of impact severities affecting hydrologic systems (Benoy et al. 2007, Christensen et al. 2007). DOC export is expected to increase throughout Canada, and in the boreal zone this increase will likely occur in the late winter/spring as a product of earlier spring melt and discharge (Clair et al. 1999).

Peat lands, wetlands, and lakes are important sites and conduits of the transport and storage of terrestrially derived biogeochemical compounds such as DIC, DOC and POC to aquatic environments. Changes in hydrology occurring with changes in climate could result in the isolation/decoupling of these systems, reducing connectivity between sites, and limiting exchange and transport. Reduction in wetland distribution and the disappearance of small arctic lakes have both been linked to permafrost melting and degradation (Smith et al. 2005, Riordan et al. 2006, Avis et al. 2011). Furthermore, changes in runoff could result in the exposure of shallow littoral sediments, possibly resulting in an increased flux of CH_4 and CO_2 to the atmosphere and reduced sediment carbon storage (Benoy et al. 2007).

Current global estimates of annual carbon accumulation in lakes range from 30 to 70 Tg C yr^{-1} (Mulholland and Elwood 1982, Dean and Gorham 1998, Stallard 1998, Einsele et al. 2001, Squires et al. 2006, Cole et al. 2007, Tranvik et al. 2009). In the boreal zone, rates are more conservative as a product of seasonal limitations in productivity and hydrology, and range from 2 to 31 Tg C yr^{-1} (Molot and Dillon 1996, Campbell et al. 2000, Algesten et al. 2004, Kortelainen et al. 2004, Benoy et al. 2007). Methods for determining carbon accumulation in these studies vary from direct measures such as lake coring and sediment analysis to indirect measures such as carbon budgets and flux estimates. However, a common theme in all is the use of lake distribution data to extrapolate carbon accumulation estimates to regional and global scales.

Most global lake distribution studies have relied upon a top down approaches for lake estimates, starting with canonical data describing large lakes and incorporating scaling rules to estimate the number of small lakes (Herdendorf 1982, Maybeck 1995). This was necessary because of a lack of data describing all but the largest lakes. Similarly, these early studies were strictly aspatial (tabular), giving little insight into how lake distribution varied at local and regional spatial scales. Newer studies have incorporated new spatial datasets describing small lake distribution while reducing dependency on scaling rules (Lehner and Döll 2004). However, even the most recent estimates of global lake distribution still rely heavily on scaling rules to estimate the number of very small lakes (Downing et al. 2006).

In Canada, hydrographic data is relatively comprehensive, freely available, and covers most of the landmass with very good spatial resolution. However, there is no single database describing lake distribution in Canada. The goal of this work was threefold. First, to construct a new database describing the size, location, and number of lakes across Canada derived from existing CanVec spatial data. Second, to use the new lake database to scale up literature based rates of carbon burial in lake sediments to estimate annual carbon burial and total storage in Canadian boreal lakes. And finally, third, to explore new thematic spatial products derived from the lake database that describe how lake size distribution and carbon accumulation vary across the landscape.

3.3 Methods

3.3.1 *Database Construction*

CanVec is distributed by NRCan under 11 distribution themes representing topographic entities. The data is available online free of charge from Geogratis in several geographic file formats including ESRI® Shapefiles based on the 1:50,000 National Topographic Survey (NTS) naming system (CanVec Hydrographic Dataset 2007). Data was downloaded from NTS directories and consisted of up to 256 individual files amounting to a total of 13,680 files covering most of the Canadian landmass (Figure 3.2).

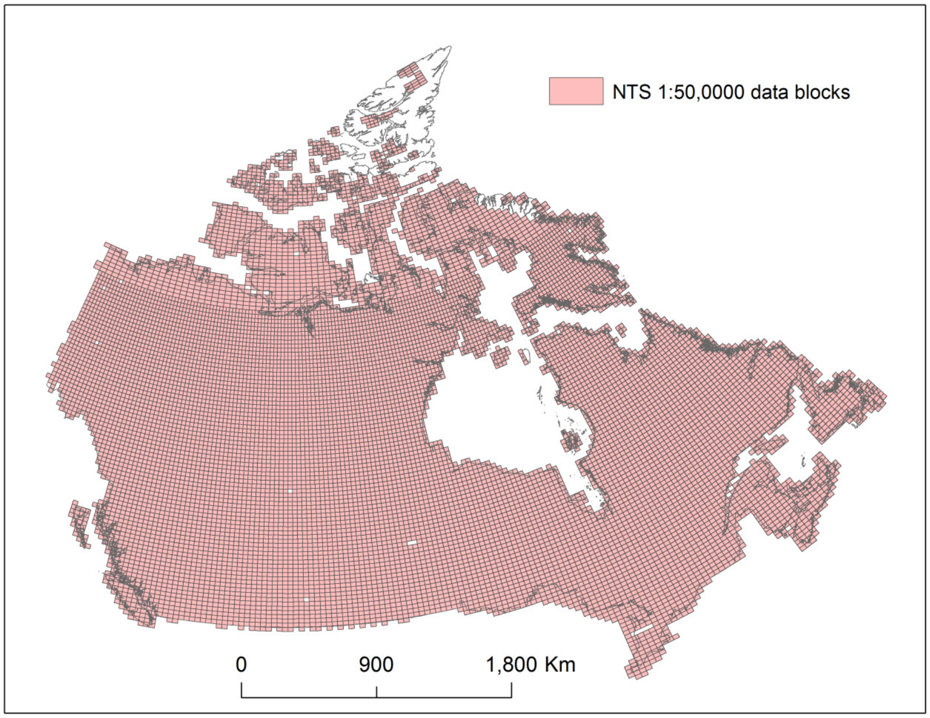

Figure 3.2. NTS 1:50,000 map sheets showing the distribution of CanVec hydrographic data.

File names containing HD_1480009 represented generic water bodies and were selectively extracted and sorted into NTS block folders. The hydrographic layer in CanVec includes both natural and artificial features such as lakes, reservoirs, canals, and liquid waste sites for a total of 13,785,933 features. In order to remove non-lake features, an SQL expression in ArcGIS® was used to select features based on attributes such as permanency on the landscape and water definition (Table 3.1). Only features with permanency values of -1 or 1 were selected as they were assumed to be persistent in nature. A preliminary exploration of the data showed that the majority of lakes were classified with water definition codes of 0, -1 or -2 (None, Unknown, Indefinite). These values are used by NRCan when insufficient data is available to classify features. As such, in order to not remove real lake features, the SQL expression had to be expanded to include these values.

Furthermore, many features such as ocean water body polygons had to be removed by hand because they were unclassified. The removal of ocean and non-lake polygons left a total of 13,637,602 features in the database.

Table 3.1. CanVec parameters used to exclude hydrographic records that did not represent lakes. Parameters in bold were used to preliminarily select features that may have been lakes.

Water Definition		
Code	Label	Definition
-2	**Indefinite**	**Not determined.**
-1	**Unknown**	**Impossible to determine.**
0	**None**	**None of the other values.**
1	Canal	An artificial body of water serving as a navigable waterway or to channel water.
3	Ditch	Small, open artificial channel constructed through earth or rock for the purpose of conveying water.
4	**Lake**	**A natural and usually flat body of water.**
5	**Reservoir**	**A wholly or partially artificial body of water for storing and/ or regulating and controlling water.**
6	Watercourse	A natural body of water through which water may flow.
7	Tidal river	A natural body of water in which flow and water surface elevation are affected by the tide.
8	Liquid waste	Liquid waste or discharge from an industrial complex.
9	**Pond**	**A body of standing water, usually smaller than a lake.**
10	Side channel	A channel providing an alternative water way within a flowing body of water.
100	Ocean	Coastal water body.

Permanency		
Code	Label	Definition
-1	**Unknown**	**Impossible to determine.**
1	**Permanent**	**Intended to exist or function for a long, indefinite period.**
2	Intermittent	Coming and going at intervals.

Similar to datasets used by Lehner and Döll (2004) for the construction of the GLWD, the CanVec dataset did a poor job of distinguishing between lakes and rivers, commonly representing them as one seamless polygon. In order to separate these features, the data was visually inspected and cut lines were added at probable lake inflow and outlet locations. In many cases, rivers were already distinct features but were left unclassified by attributes. As such, following Lehner and Döll (2004), a surface development equation was employed to distinguish rivers from lakes. By comparing the ratio of a water body's surface area to the area of the smallest bounding circle, the surface development equation allowed for the morphological shape of features to be described by a single value (Eq. 3, Figure 3.3). Lakes tend to have a more circular and compact shape whereas rivers are more linear and extended. Surface development values were explored to determine a reasonable division between rivers and lakes. Following Lehner and Döll (2004), a value of 3% was used as a threshold between rivers and lakes. The application of the surface development equation removed an additional 111,367 river polygons leaving 13,526,235 features.

$$\text{Surface development} = \frac{\text{Lake surface area}}{\text{Area of smallest enclosing circle}} \times 100 \tag{3}$$

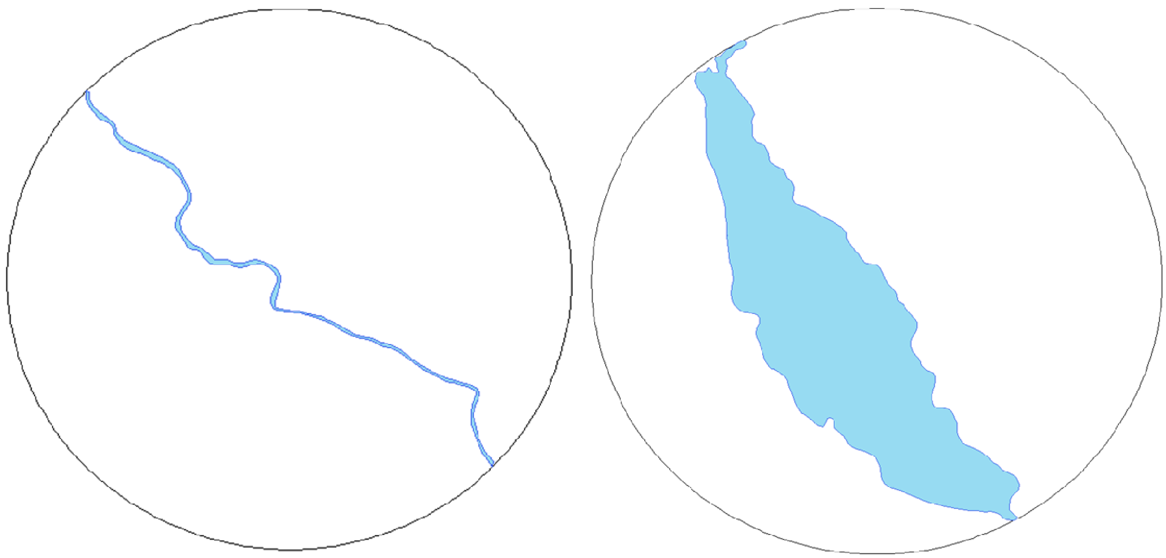

Figure 3.3. An illustration of how the surface development equation describes water body shape. A low surface development value reflects a low water body area to bounding circle area ratio (Lehner and Döll 2004). Typically rivers have values < 3% whereas lakes have values > 3%.

With the preliminary removal of rivers and non-lake features, the next step was to start merging the individual 1:50,000 NTS blocks. One consequence of working with 13,680 spatially distinct blocks was that features crossing the boundaries between neighbouring data blocks were cut into 2 or more polygons. Unfortunately, no unique identifier was included in the attribute data that would allow contiguous multipart features to be merged back into single polygons. This meant that data blocks had to be merged and then subsequently dissolved using the Merge and Dissolve tools in ArcGIS®. Because of the extremely high number of lakes, this was done incrementally in a hierarchical fashion in order to minimize processing errors and system crashes: 1:50,000 data blocks were first merged and dissolved into NTS map sheets, then into large regional blocks, and finally into a single contiguous feature class in a File Geodatabase representing a total of

13,275,731 lakes (Figure 3.4). The Great Lakes along the southern border of Ontario were bisected by the international border with the United States. In order to represent their surface area accurately, polygons from the GLWD where imported into the CanVec database. Lake Michigan was excluded because it lies completely within the United States. Finally, surface area for each feature was re-calculated and then classified based on respective log size classes for comparison to other datasets (Herdendorf 1982, Maybeck 1995, Lehner and Döll 2004, Downing et al. 2006).

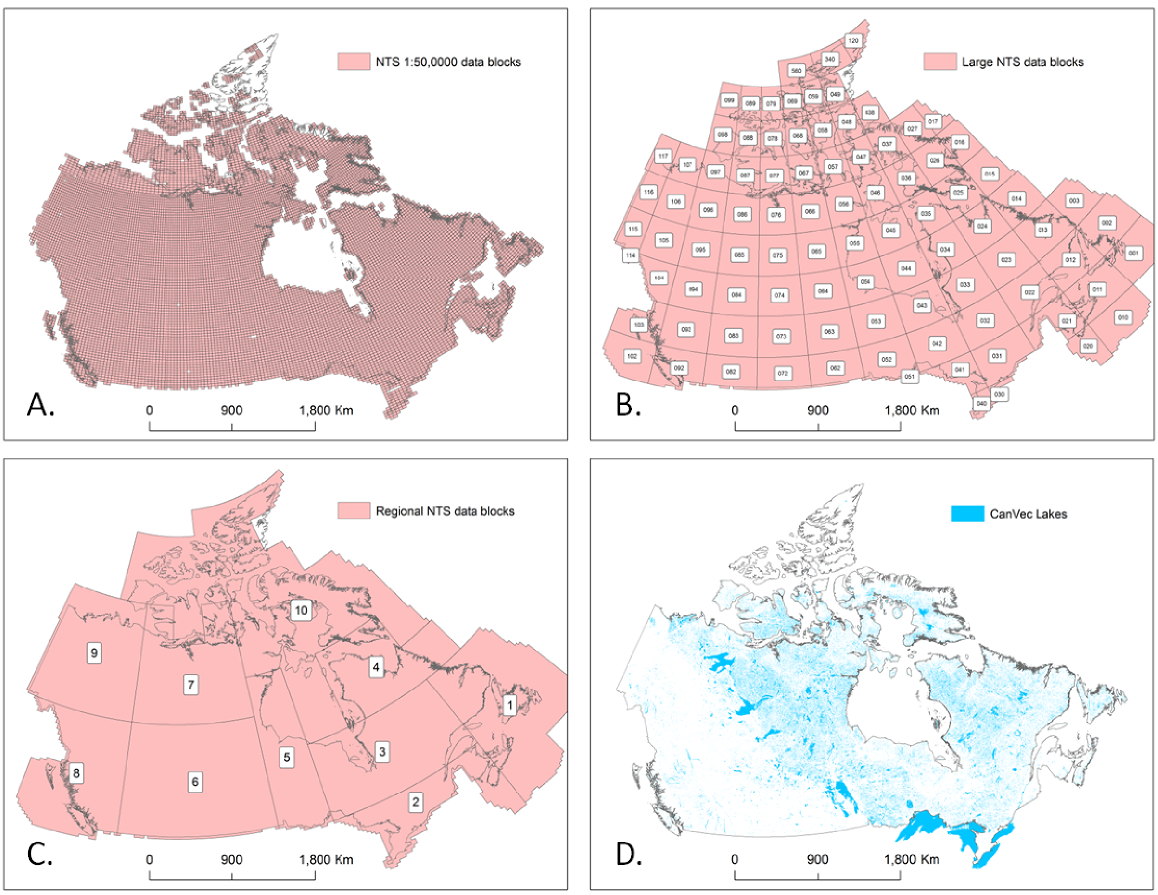

Figure 3.4. Data aggregation from (A) 1:50,000 NTS blocks, to (B) NTS map sheets, to (C) regional data blocks, and finally to (D) a single feature class representing lakes in Canada.

3.3.2 Carbon Accumulation Estimates

Once construction of the lake database was complete, lake sediment carbon accumulation rates were applied. Many of the literature based estimates of lake sediment carbon accumulation in Canada and Europe have come from within the boreal zone (Campbell et al. 2000, Algesten et al. 2004, Kortelainen et al. 2004, Jonsson et al. 2007). The landscape of Canada can be divided into ecozones that reflect underlying biotic and abiotic factors. In order to simplify the application of carbon accumulation data, a modified version of the terrestrial ecozones was used to identify the boreal region in Canada. Similar to Benoy et al. (2007), the boreal zone comprised 7 of the 15 terrestrial ecozones of Canada, covering ~ 5.8 million km^2 and accounting for 77% of Canada's forested land (Wiken 1986, Kalff 2002) (Figure 3.5).

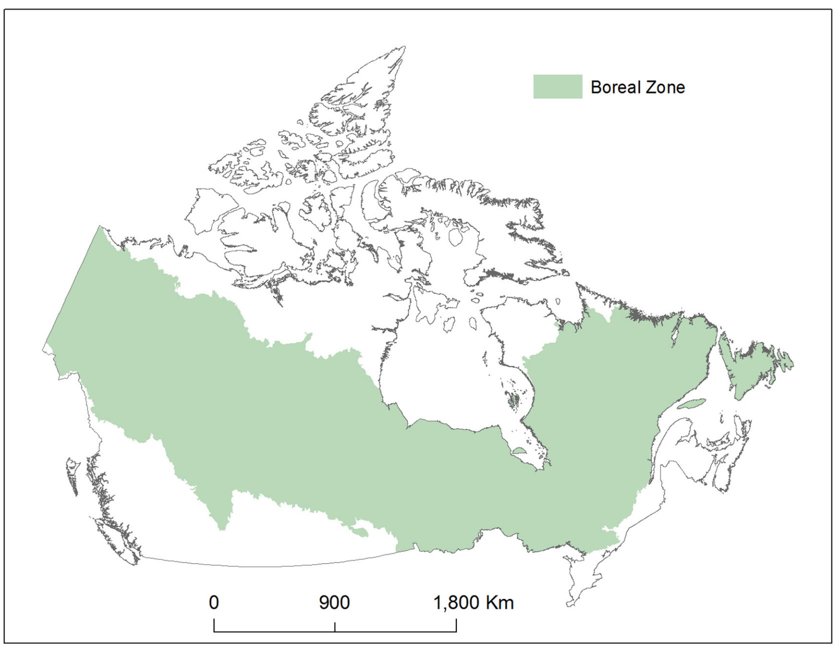

Figure 3.5. The boreal zone derived from the Canadian ecozone classification scheme (Wiken 1986).

The accumulation rates used here were derived from a collection of Finnish boreal lakes following Telmer and Costa (2007) (Table 3.2). Finnish accumulation rates were used because boreal Finland and boreal Canada have similar physiographic conditions. Moreover, the Finnish data were based on a large lake census and rates varied as a function of lake size across five size classes. This allowed for estimates to be directly applied to the CanVec database, which describes lakes as a function of their size. Carbon accumulation rates were applied to the entire lake database; however, only lakes located within the boreal region are discussed.

Table 3.2. Lake sediment carbon accumulation rates from Pajunen (2000, 2004) for Finnish lakes selected from the Northern Lake Survey Database.

Lake size class (ha)	C accumulation rate (t C ha^{-1} yr^{-1}) (Pajunen 2000) n = 31	C accumulation rate (t C ha^{-1} yr^{-1}) (Pajunen 2004) n = 140
<10	0.0600*	0.0240
10-100	0.0572	0.0400
100-1,000	0.0444	0.0310
1,000-10,000	0.0231	0.0180
>10,000	0.0100	0.0096

* Extrapolated (Telmer and Costa 2007)

3.3.3 Thematic Products: Lake Size Distribution and Carbon Accumulation Maps

Local variations in lake size distribution can be expressed thematically by the slope value (m) in Eq. 3 and Eq. 4. A higher proportion of small lakes relative to large lakes results in a more negative slope value, whereas a higher proportion of large lakes relative to small lakes results in a less negative slope. Empty polygon grids covering Canada were constructed in ArcGIS® at four resolutions - 625 km^2 (25x25 km), 2500 km^2 (50x50 km), 10,000 km^2 (100x100 km) and 40,000 km^2 (200x200 km) - to test the effect of cell size on lake size distribution. Grid polygons were then spatially joined to intersecting lake features from the database using the Spatial Join tool in ArcGIS®. This allowed each grid cell to retain the attributes of intersecting features and for lake sums within each size class to be calculated. The final step in constructing a slope distribution map was to export the distribution to Microsoft Excel® where the data was log-linearized and fitted with a power function in the form of:

$$y = bx^m \tag{3}$$

OR

$$\log(y) = m\log(x) + \log(b) \tag{4}$$

This allowed for variables such as slope, intercept, coefficient of determination (R^2), and class membership (the number of lake size classes present) to be calculated for each cell.

Once imported back into ArcGIS®, slope values were expressed thematically to explore how values varied spatially.

Similarly, carbon accumulation rate was expressed thematically for each cell as the sum of the product of the number of lakes (y) (see Eq. 3 and Eq.4), the average lake size, and accumulation rate for each of the five lake size classes:

$$Grid\ Cell\ Carbon\ Accumulation\ Rate = \sum_{i=1}^{5}(y \times Avg.lake\ size\ \times\ Acc.Rate) \qquad (5)$$

This allowed for tabular rates to be expressed spatially by incorporating per grid cell differences in local lake distribution. Two sets of values were used for average lake size: the true mean of each bin (taken from the CanVec lake database), and following previous studies, 2.5 times the lower size class limit (e.g. 25 ha for the 10-100 ha class) (Maybeck 1995, Lehner and Döll 2004, Downing et al. 2006) (Table 3.3).

Table 3.3. The average lake size used in calculating carbon accumulation via Eq. 5 for each of the five size classes based on the CanVec database and literature values.

Lake size class (ha)	Average lake size per size class (ha)	
	CanVec Database	Literature values[a]
<10	1.03	2.5
10 - 100	27.02	25
100 – 1,000	243.38	250
1,000 – 10,000	2,417.78	2,500
>10,000	97,047.96	25,000

[a] Based on class average lake sizes from previous studies (Maybeck 1995, Lehner and Döll 2004, Downing et al. 2006).

3.4 Results

3.4.1 Lake Database

The CanVec lake database contained a total of 13,273,745 features and covered 1,192,486 km^2 across nine log size classes (Table 3.4). The distribution of the database was complete for much of the Canadian landmass with the exception of the far north were data were either unavailable or incomplete (Figure 3.6).

Table 3.4. The number and size of lakes in the CanVec database distributed across log size classes.

Lake size class		Lake area		Lake count
hectares	**km^2**	**hectares**	**km^2**	
< 0.1	< 0.001	162,559	1,626	2,674,371
0.1 - 1	0.001 - 0.01	2,410,085	24,101	6,533,713
1 - 10	0.01 - 0.1	10,277,577	102,776	3,235,853
10 - 100	0.1 - 1	20,107,648	201,076	744,190
100 – 1,000	1 - 10	19,347,835	193,478	79,496
1,000 – 10,000	10 - 100	13,469,465	134,695	5,571
10,000 – 100,000	100 – 1,000	12,790,683	127,907	489
100,000 – 1,000,000	1,000 – 10,000	13,694,275	136,943	55
> 1,000,000	> 10,000	26,988,466	269,885	7
	Sum:	119,248,594	1,192,486	13,273,745

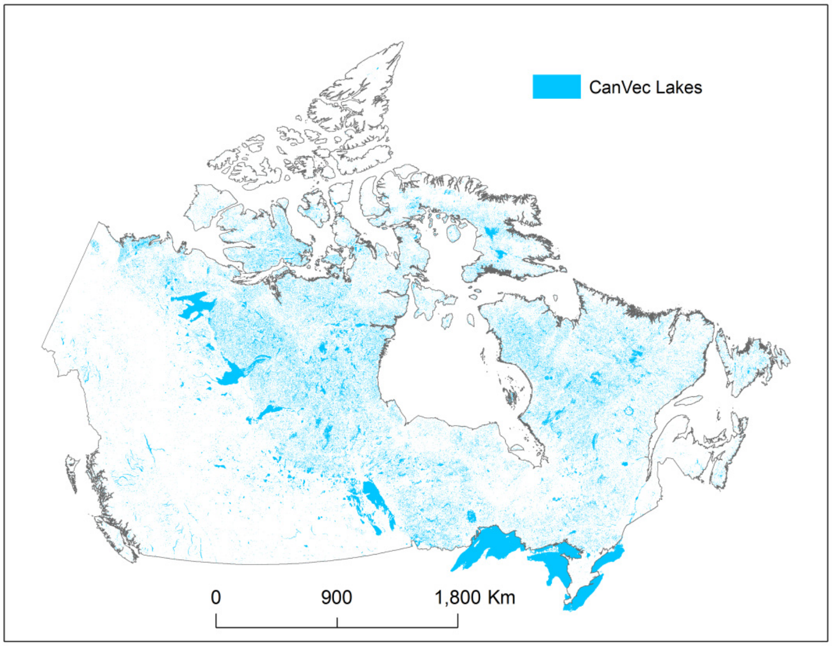

Figure 3.6. Final lake database coverage derived from CanVec and supplemented with GLWD for the great lakes. Missing data is mostly restricted to 70˚ N latitude and above.

When viewed graphically, the distribution showed a consistent increase in the number of lakes going from largest to smallest with the exception of lakes <0.1 ha. Lake area was relatively evenly distributed across the middle size classes between 1 ha and 1,000,000 ha. The exceptions were lakes less than 1 ha in size and lakes larger than 1,000,000 ha in size. The smallest lakes were numerically dominant yet had the least areal extent. In contrast, the largest lakes, while only few in number, accounted for over 22% of total lake area (Figure 3.7 , Table 3.4).

Across nine size classes, the power function from Eq. 3 fit the data reasonably well (Figure 3.7, R^2 = 0.9359). When the lower size classes (< 0.1 and 0.1 – 1 ha) were excluded, the fit of the power function improved (Figure 3.8, R^2 = 0.9848 and Figure 3.9, R^2 = 0.9964).

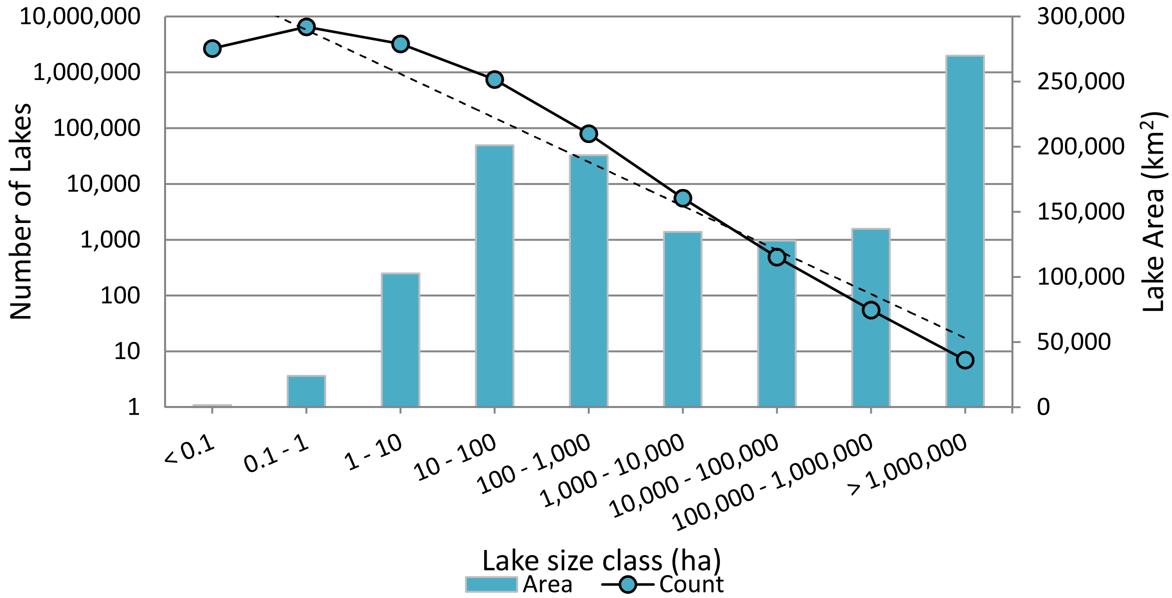

Figure 3.7. CanVec lake count and area distributed across nine log size classes ranging from < 0.1 to > 1,000,000 ha and fitted with a power function (dashed line: $y = 935{,}908.7258x^{-0.7887}$ $R^2 = 0.9359$).

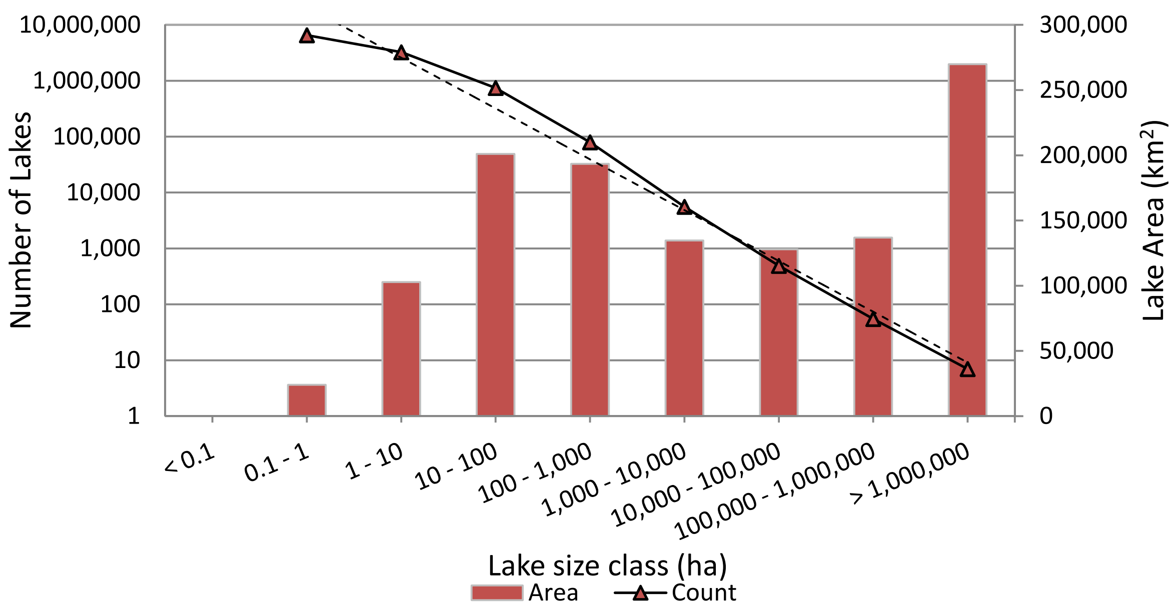

Figure 3.8. CanVec lake count and area distributed across eight log size classes ranging from 0.1 to > 1,000,000 ha. Here lakes less < 0.1 ha were excluded resulting in a power function with a steeper slope and a better fit (dashed line: $y = 2{,}580{,}844.9836x^{-0.9088}$ $R^2 = 0.9848$).

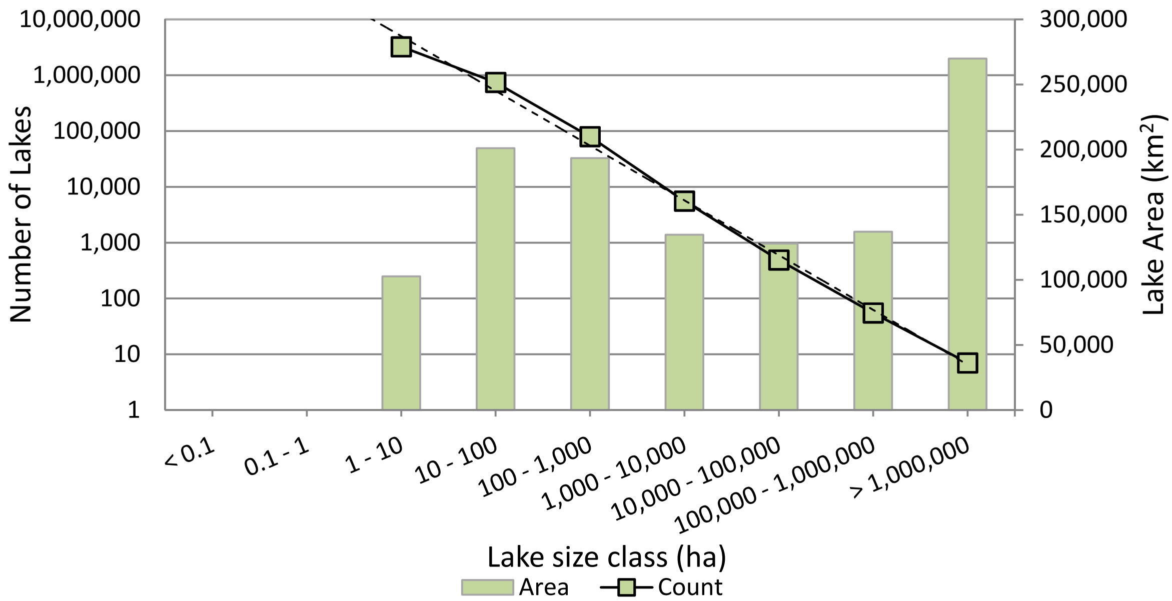

Figure 3.9. The CanVec lake count and area distributed across seven logarithmic size classes ranging from 1 to > 1,000,000 ha. Here lakes less < 1 ha were excluded resulting in a steeper slope and a better fitting power function (dashed line: $y = 5{,}018{,}444.4288x^{-0.9810}$ $R^2 = 0.9964$).

3.4.2 Carbon Accumulation Estimates

Carbon estimates were calculated for the boreal zone using carbon accumulation rates derived from a series of lakes in boreal Finland (Pajunen 2000, 2004). When applied to the CanVec lake database, total carbon accumulation was 1.65 and 2.34 Mt C·yr^{-1}. The 0.68 Mt difference between the two estimates is a product of the difference in accumulation rates between Pajunen (2000) and Pajunen (2004), particularly in the smaller lake size classes (Table 3.2). When these rates were applied to the lake distribution, small to moderate sized lakes (10 – 100 ha) and very large lakes (>10,000 ha) accounted for the most carbon accumulation (Table 3.5, Figure 3.10).

Table 3.5. Carbon accumulation rates for Canadian lakes in the boreal zone based on the CanVec lake database and Finnish accumulation rates.

Lake size class (ha)	Lake count	Lake area (km^2)	Carbon accumulation ($t\ C \cdot yr^{-1}$) Pajunen (2000)	Carbon accumulation ($t\ C \cdot yr^{-1}$) Pajunen (2004)
<10	6,886,926	69,029.88	414,179.29*	165,671.72
10 - 100	436,122	119,683.41	684,589.10	478,733.63
100 – 1,000	49,908	123,372.69	547,774.76	382,455.35
1,000 – 10,000	3,778	92,461.85	213,586.86	166,431.32
> 10,000	412	477,278.77	477,278.77	458,187.62
Sum:	**7,377,146**	**881,826.60**	**2,337,409**	**1,651,480**
Stock (10,000 yrs) (Pg):			**23.37**	**16.51**

* Based on an extrapolated accumulation rate by Telmer and Costa (2007) (see Table 3.2)

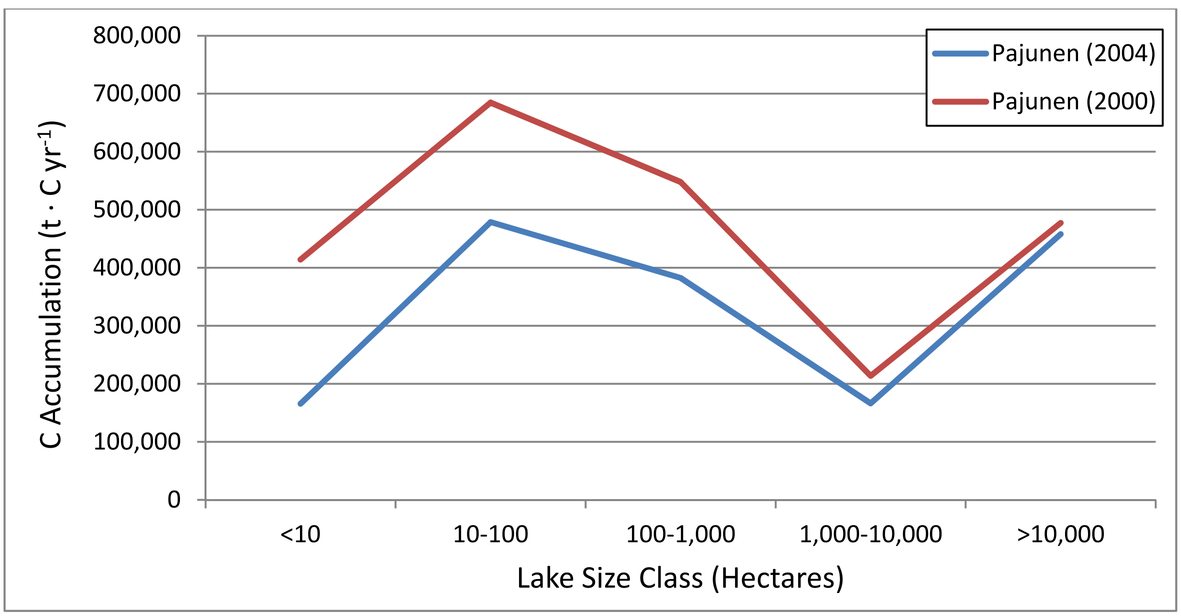

Figure 3.10. Carbon accumulation in Canadian boreal lakes based on rates from Pajunen (2000) and Pajunen (2004) applied to the CanVec lake database. Data values are listed in Table 3.5.

3.4.3 Thematic Products

3.4.3.1 Lake Size Distribution Maps

The power function shown in Eq. 3 and Eq. 4 allowed for the calculation of the slope values for each grid cell as an expression of the relative frequency of small, medium, and large lakes in the underlying lake distribution. A more negative slope indicated a higher proportion of small lakes relative to large lakes whereas a less negative slope indicated a lower proportion of small lakes relative to large lakes (Figure 3.11). Slope values for all four resolutions ranged from 0.301 to -3.564 (Table 3.6). As cell size increased, slope value range and standard deviation decreased and mean R^2 values increased. Another measure of fit was class membership count, which represented the number of lake size classes present within each grid cell. Similar to R^2, average class membership increased with increasing grid cell size while variation decreased (Table 3.6).

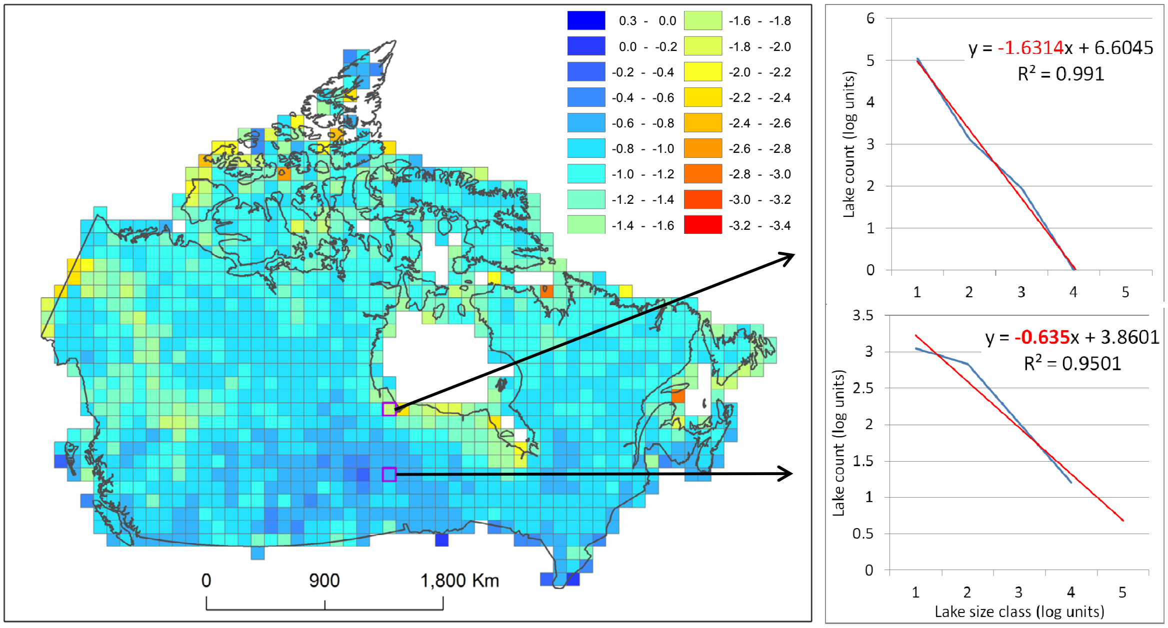

Figure 3.11. A schematic of how lake size distribution (slope) values are calculated for each cell. The best fit lines show that the yellow cell has a steeper slope than the blue cell as a product of different underlying lake distributions.

Table 3.6. Descriptive statistics for slope, R^2 and Class membership counts for the four grid cell resolutions.

		Cell size (km²)			
		625	**2,500**	**10,000**	**40,000**
Lake size distribution (slope)	Cell count:	15,755	4,440	1,239	349
	Minimum:	-3.564	-3.035	-2.898	-2.330
	Maximum:	0.301	0.301	0	0
	Mean:	-0.970	-1.022	-1.055	-1.069
	SD:	0.408	0.342	0.294	0.245
R^2	Cell Count:	15,732	4,299	1,237	349
	Minimum:	0	0.383	0.481	0
	Maximum:	1	1	1	1
	Mean:	0.950	0.980	0.972	0.978
	SD:	0.086	0.042	0.050	0.065
Class membership count	Cell Count:	18,273	4,610	1,268	353
	Minimum:	0	1	1	1
	Maximum:	5	5	5	5
	Mean:	3.012	3.682	4.048	4.363
	SD:	1.303	1.088	1.017	0.887

Variation in slope values can be shown spatially at the four grid cell resolutions (Figure 3.12). Approximately 33% of Canada had slope values ≤ -1 at the finest resolution of 625 km^2. This increased to 43%, 51%, and 63% for 2,500 km^2, 10,000 km^2 and 40,000 km^2, respectively. The smallest grid cell size had the highest proportion of empty cells (Figure 3.12 A). These cells corresponded with regions that had low class membership counts (Figure 3.14 A). The number of empty grid cells was reduced considerably with an increase in cell size to 2,500 km^2, and completely removed at 10,000 km^2 and 40,000 km^2 resolutions (Figure 3.12 B, C, and D). Class membership also increased with increasing cell size (Figure 3.14 B, C, and D). The highest slope values occurred in regions along Canada's coast, geographic edges of the database, and in regions with class membership values of 2 or 3 (Figure 3.12, Figure 3.13, and Figure 3.14).

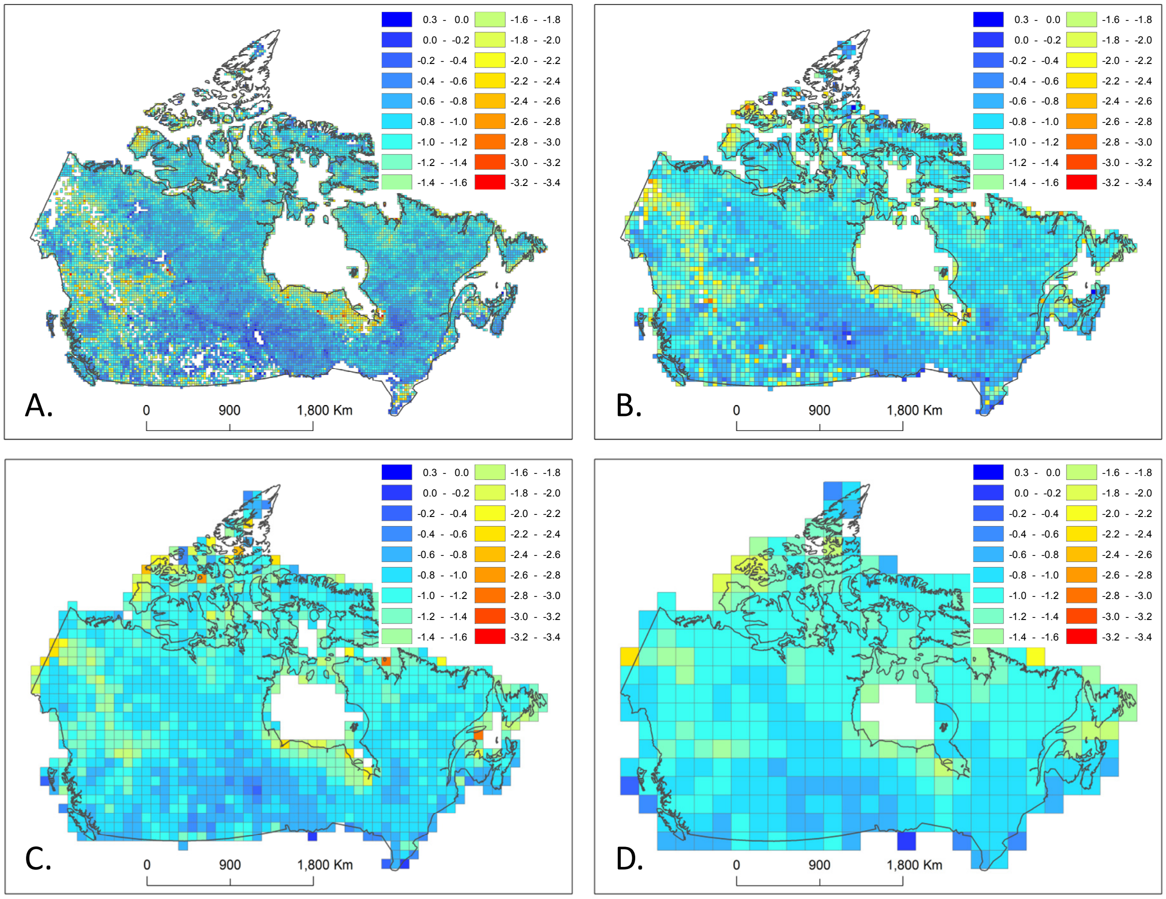

Figure 3.12. Lake size distribution (slope) maps at four grid cell resolutions. (A) 625 km^2 (25x25 km), (B) 2,500 km^2 (50x50 km), (C) 10,000 km^2 (100x100 km), (D) 40,000 km^2 (200x200 km). Color scale is based on 0.2 slope intervals with the exception of the positive range (0.0 – 0.3). Empty cells indicate insufficient data to calculate slope.

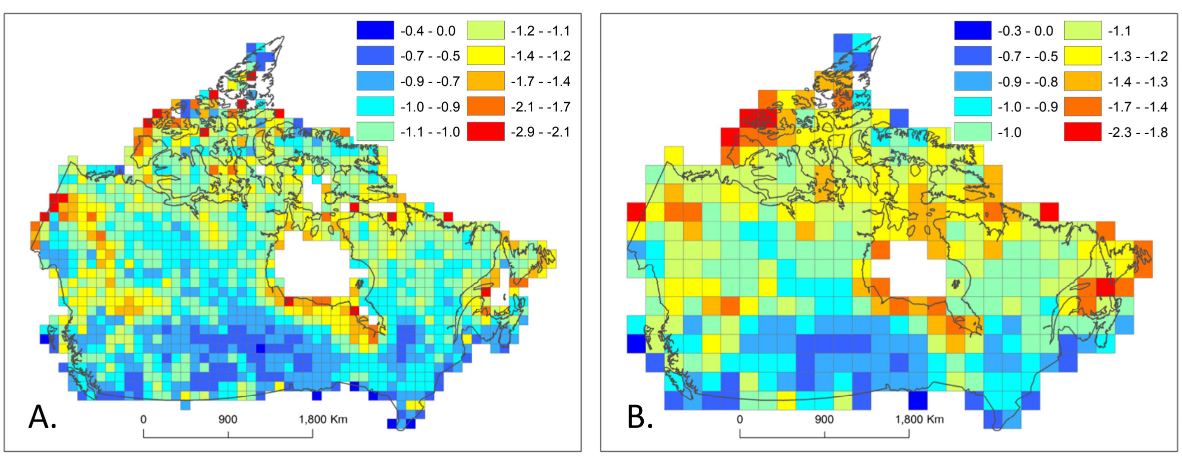

Figure 3.13. Lake size distribution (slope) values for (A) 10,000 km^2 and (B) 40,000 km^2 using a color scale based on natural breaks within each respective distribution to enhance contrast. Warmer colors reflect a more negative slope, which indicates a higher proportion of small lakes relative to large lakes. If we assume that average lake size per class is 2.5 times the lower bound, then slopes <-1 indicate that small lakes account for more area than large lakes.

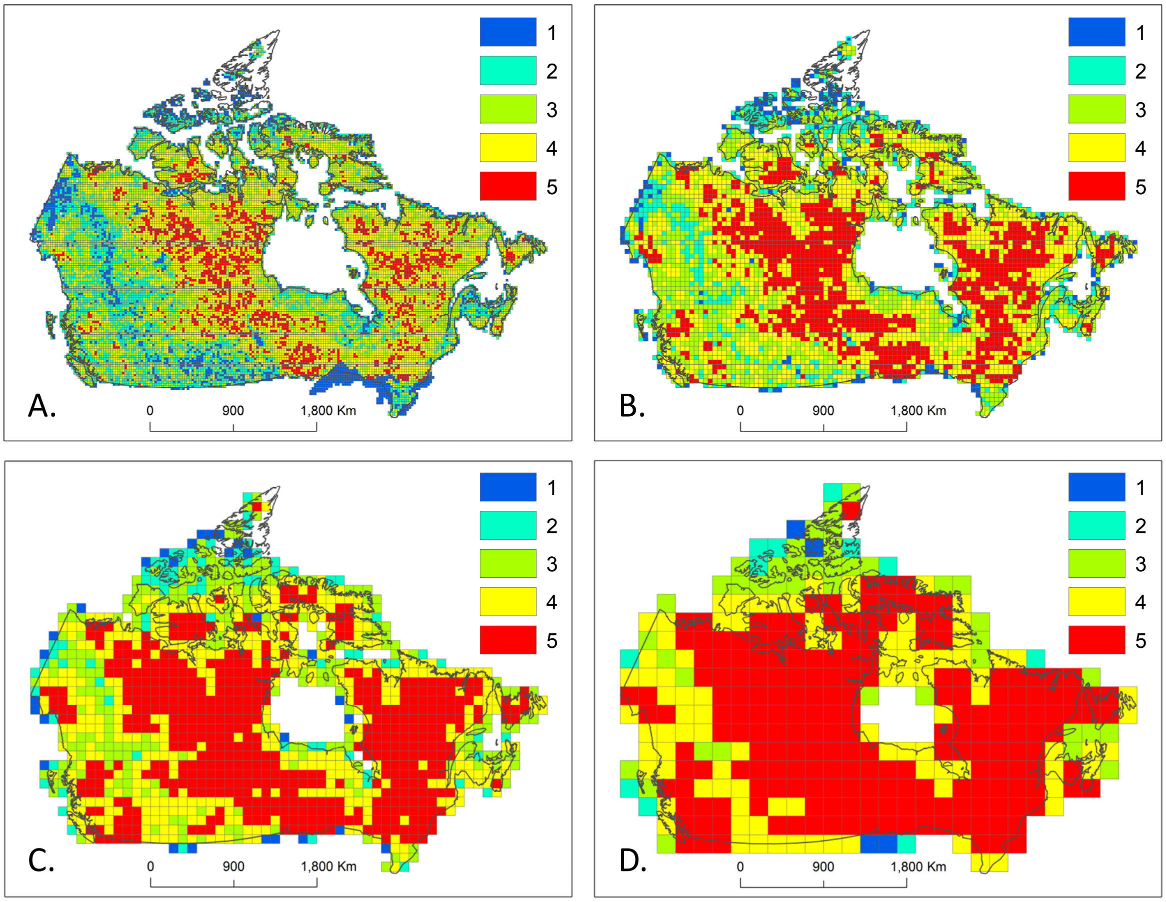

Figure 3.14. Class membership values at the four grid cell resolutions. (A) 625 km^2 (25x25 km), (B) 2,500 km^2 (50x50 km), (C) 10,000 km^2 (100x100 km), (D) 40,000 km^2 (200x200 km). Values are based on the number of lake size classes represented within each cell.

The spatial distribution of overall lake frequency showed very high lake densities in much of the central and northern parts of Canada at all four grid cell resolutions (Figure 3.15). Low frequency values occurred throughout much of the prairies and in the far north. There appeared to be no consistent relationship between lake frequency and lake size distribution (slope value). In some regions such as the Hudson plains to the south of Hudson Bay, high lake frequencies correlated with more negative/steeper slope values.

In contrast, areas such as parts of the Yukon in north-west Canada showed an opposite trend, where low frequencies correlated with more negative/steeper slope values (Figure 3.13 and Figure 3.15).

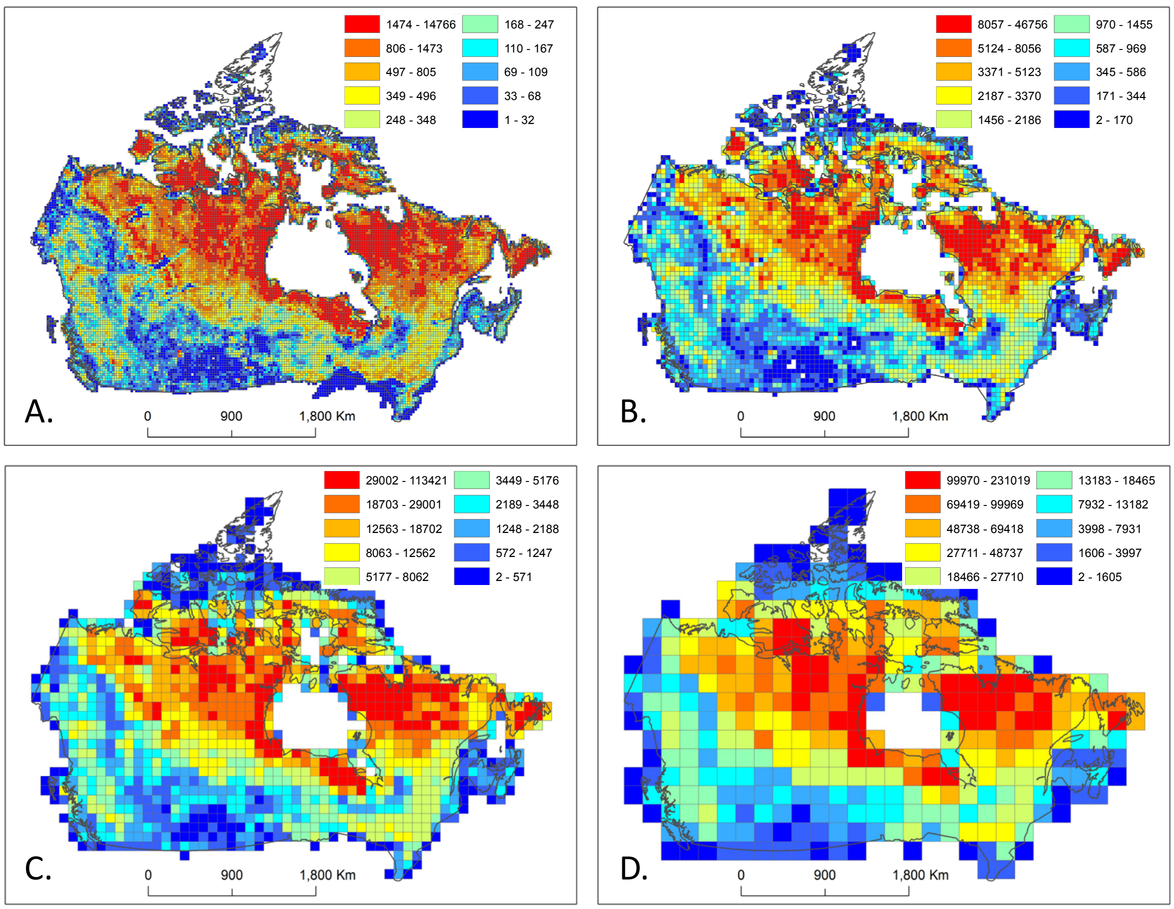

Figure 3.15. Total lake frequency within each grid cell at four resolutions. (A) 625 km^2 (25x25 km), (B) 2,500 km^2 (50x50 km), (C) 10,000 km^2 (100x100 km), (D) 40,000 km^2 (200x200 km). Note the difference in color scales between the four resolutions. Quantiles from each distribution were used for color scales because of the wide range of frequency values between cell sizes. Regions of very high lake frequency such as the Hudson Bay Lowlands are commonly dominated by small lakes, which account for most of the lake area.

3.4.3.2 Carbon Accumulation Maps

The carbon accumulation maps produced from Eq.5 relied on slope and intercept values from the best fit power function. Only maps using the two largest grid cells sizes (10,000 km^2 and 40,000 km^2) were used because of better power function fit (higher R^2 and class membership values), which helped to minimize estimation errors (Table 3.6). Similar distributions resulted from the two measures of average lake size per class (true mean vs. literature values, cf. Table 3.3). Overall patterns of carbon accumulation were similar but varied in magnitude within and between the two grid cell resolutions (Figure 3.16 and Figure 3.17). Areas of high carbon accumulation were focused in the northern boreal region and southern arctic to the west and east of Hudson Bay. Total carbon accumulation for the boreal region varied between 1,610,884 and 2,808,750 t C yr^{-1} depending on cell resolution, average lake size per class used, and carbon accumulation rate (Table 3.7).

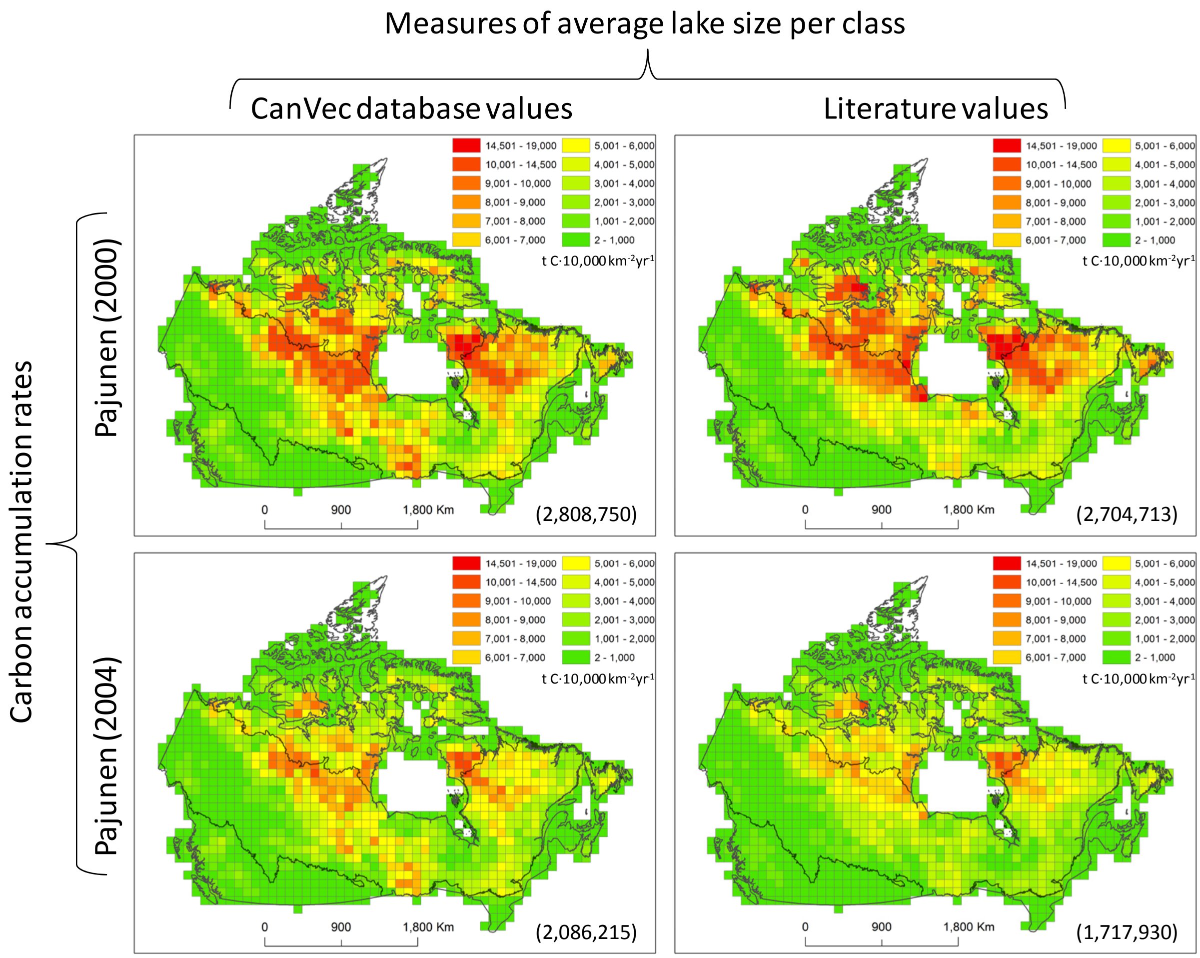

Figure 3.16. Carbon accumulation maps at 10,000 km^2 resolution using two measures of average lake size per size class and two rates of carbon accumulation. The boreal zone is outlined in black. Values in parentheses are the total sum of carbon accumulation (t C yr^{-1}) for grid cells within the boreal zone.

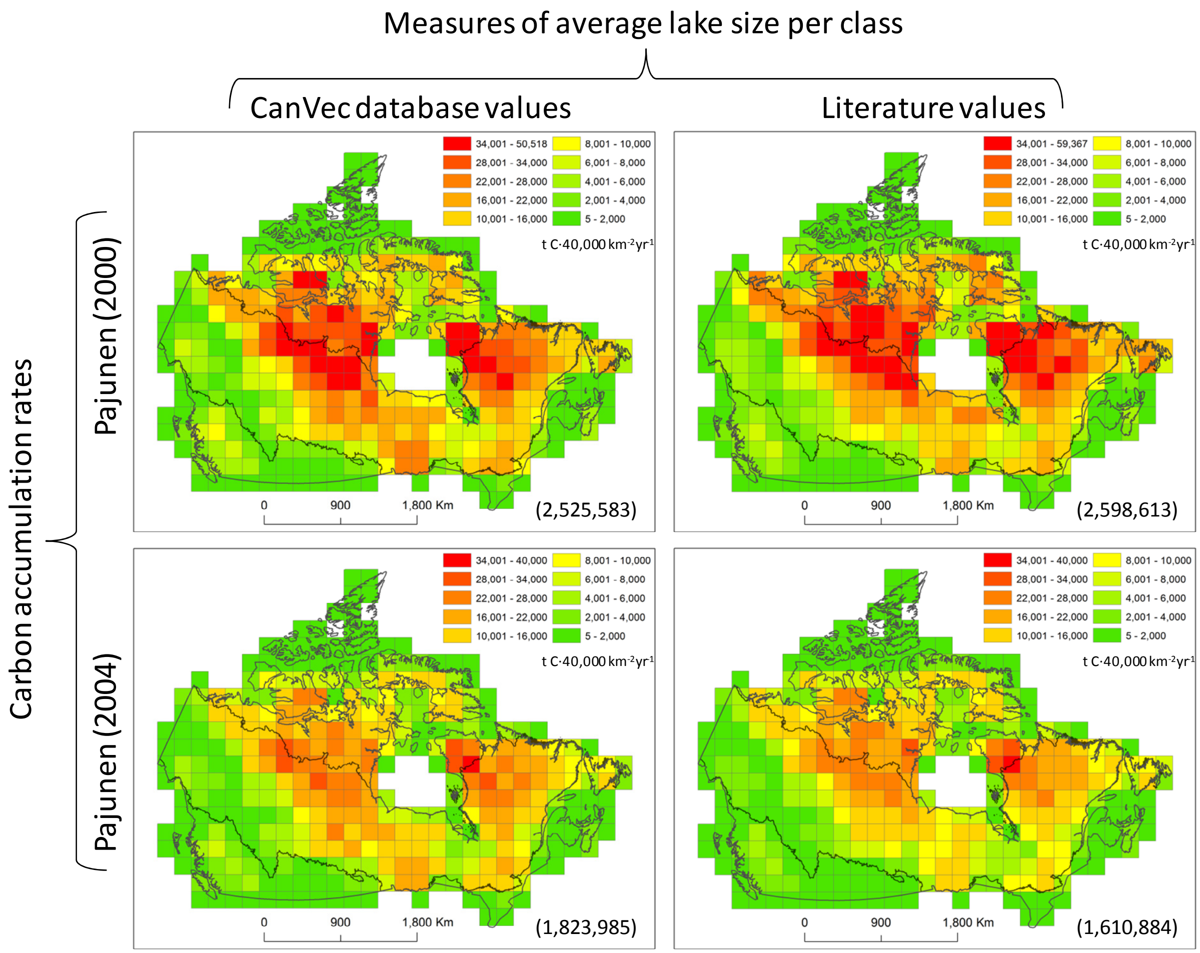

Figure 3.17. Carbon accumulation maps at 40,000 km^2 resolution using two measures of average lake size per size class and two rates of carbon accumulation. The boreal zone is outlined in black. Values in parentheses are the total sum of carbon accumulation (t C yr^{-1}) for grid cells within the boreal zone.

Table 3.7. Comparing the boreal zone at the two cell resolutions (10,000 km^2 and 40,000 km^2) using the two accumulation rates and measures of average lake size within each size class.

		Carbon accumulation (t C yr^{-1})	
Cell size (km^2)	Accumulation rate	CanVec database[a]*	Literature values[b]*
10,000	Pajunen (2000)	2,808,750	2,704,713
	Pajunen (2004)	2,086,215	1,717,930
40,000	Pajunen (2000)	2,525,583	2,598,613
	Pajunen (2004)	1,823,985	1,610,884

a. Average lake size per class from the CanVec database (see Table 3.3)
b. Average lake size per class from previous studies (i.e. 2.5 times the lower bound: see Table 3.3)
*Lake count data used to calculate carbon accumulation in both cases was from the CanVec database

3.5 Discussion

3.5.1 Lake Database

The CanVec database resulted in 13,273,745 lakes with a total area of 1,192,486 km^2 (Table 3.4). In comparison, the GLWD produced by Lehner and Döll (2004) contained just 119,736 lakes and reservoirs covering a total area of 933,651 km^2 for Canada. Similarly, Maybeck (1995) estimated 489,061 lakes with a total area of 833,000 km^2. A closer look at how these previous studies compare to CanVec shows that there is good agreement in both the number and area for the largest lakes (Table 3.8). This is expected as large lakes are the easiest to account for and the least likely to change over time. The main difference between these estimates and the CanVec database is in the frequency of smaller lakes (< 100 ha). This is a product of both how a lake is defined and the availability of small lake data. In the case of Maybeck (1995), all lakes less than 1,000 km^2 (100,000 ha) were estimated from a power function derived from scaling rules. Lehner and Döll (2004) improved on Maybeck (1995) by incorporating new spatial data that removed the need to estimate lakes < 1,000 km^2. However, they limited lakes to > 10 ha because of a lack of data and the difficulty in defining the transition from lakes to wetlands. At the smallest size class included (10-100 ha), their dataset unrepresented the number of lakes and area by 96% and 89% when compared to CanVec and by 93% and 80% compared to Maybeck (1995), respectively. Not only were the smallest lakes (< 10 ha) not represented, this suggests that small to moderate sized lakes (< 100 ha) are also underrepresented.

Although not included in the published GLWD, Lehner and Döll (2004) also used a power function based on scaling rules. They estimated the global number of lakes > 1 ha to be ~15 million with a total area of 3.2 million km^2. Although they did not include region specific estimates, the CanVec database (representing Canada alone) still accounts for over 88% of their global lake count and 37% of their total global lake area. This suggests that even with the inclusion of scaling rules, both the number of lakes and area were underrepresented both regionally (in Canada) and globally.

Table 3.8. Comparison of CanVec, Maybeck (1995), and GLWD datasets (Lehner and Döll 2004)

ha	< 10	10 - 10^2	10^2 - 10^3	10^3 - 10^4	10^4 - 10^5	10^5 - 10^6	10^7 +	Sum	Limnic ratio (%)
km^2	< 0.1	0.1 - 1	1 - 10	10 - 100	100 - 1000	1,000 – 10,000	10,000 +		
CanVec									
A_o (km^2)	128,502	201,076	193,478	134,695	127,907	136,943	269,885	1,192,486	11.0
n (count)	12,443,937	744,190	79,496	5,571	489	55	7	13,273,745	
Maybeck									
A_o (km^2)		114,000 [a]	114,000 [a]	117,000 [a]	136,000 [a]	88,000	270,000	839,000	8.4
n (count)		440,000 [a]	44,000	4,500 [a]	523 [a]	31	7	489,061	
GLWD									
A_o (km^2)		22,696	217,768	174,709	144,526	104,177	269,776	933,651	8.6
n (count)		31,896	79,934	7,274	579	46	7	119,736	

a. Extrapolated based on scaling rules

A common theme in global lake distribution studies has been the incorporation at some level of lake estimates derived from scaling rules (Herdendorf 1982, Maybeck 1995, Kalff 2002, Lehner and Döll 2004, Downing et al. 2006). This characteristic exponential increase in lake count with decreasing lake size can also be seen clearly in the CanVec database (Figure 3.7, Figure 3.8 and Figure 3.9). More recently, Downing et al. (2006) used scaling rules to describe global lake size distribution for lakes as small as 0.1 ha in size for a global estimate of 304 million lakes covering 4.2 million km^2. Downing et al. (2006) assumed that regional distribution patterns for lakes as small as 0.1 ha could be extrapolated to global scales. Although CanVec (regional) and Downing et al. (2006) (global) cannot be directly compared, the shape of the distributions and the lake density values (d_L, lakes per million km^2) can be. The CanVec database shows higher lake densities and a similar slope to Downing et al. (2006) for lakes > 10 ha. Below 10 ha, the CanVec slope become less negative and begins to flatten (Figure 3.18). By the smallest lake size (0.1 – 1 ha), Downing et al. (2006) estimates a lake density over three times what is observed in the CanVec dataset.

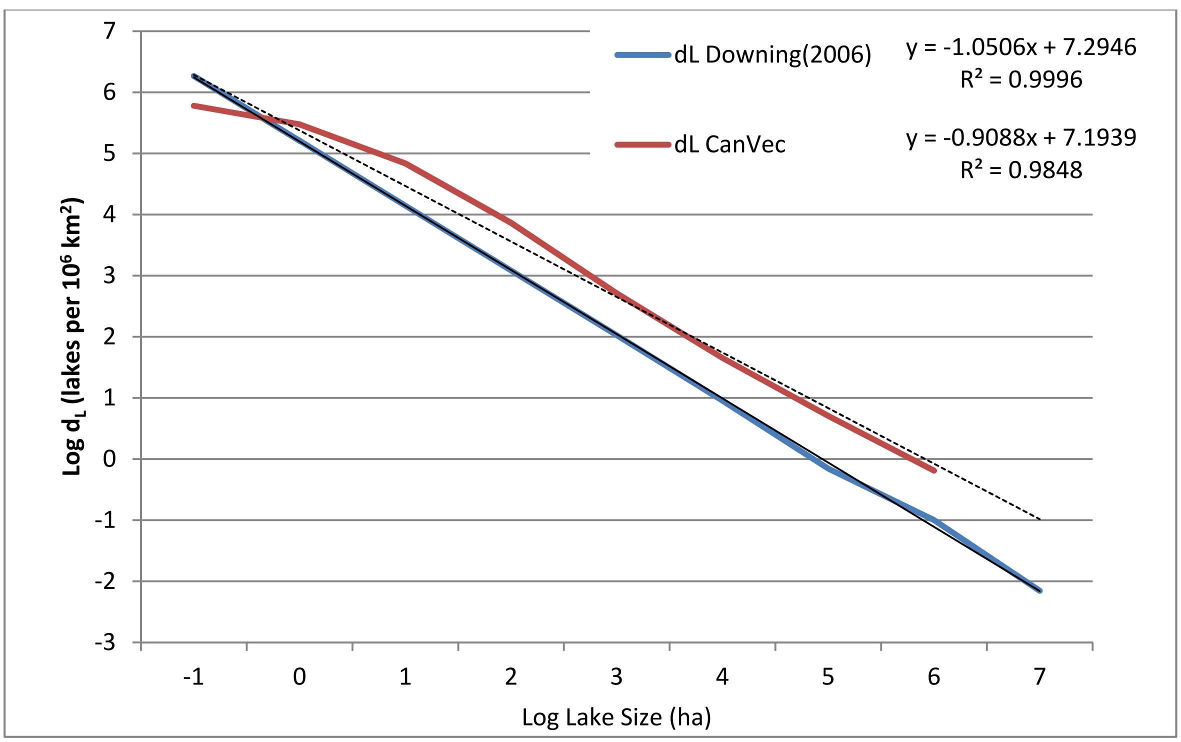

Figure 3.18. Comparing slope between Downing et al. (2006) and the CanVec lake database based on lake density values (per 10^{6}km^2). The solid and dashed black lines represent the best fit lines for each distribution.

This indicates that either CanVec is incomplete and is under-representing the smallest lakes, or the application of scaling rules by Downing et al. (2006) to estimate lakes as small as 0.1 ha may not be valid. If we assume that the CanVec database accurately represents the true lake size distribution, this would demonstrate, at least in Canada, that scaling rules may only be accurate when estimating lake distribution above a certain size class. If the scaling rules do accurately represent the smallest lakes in Canada, they may not apply well elsewhere. This could mean the Downing et al. (2006) global estimate of 304 million lakes, of which 277 million are less than 1 ha in size is inaccurate.

Validating the completeness of the CanVec database, especially for the smaller lakes, is difficult. However, there is good agreement between the CanVec database and other existing datasets for lakes > 10 ha in size and based on visual inspection with limited but very high resolution imagery, CanVec appears to do a good job of representing even smaller lakes. Similarly, CanVec is the only database that is based entirely on real empirical data and does not rely on scaling rules to estimate any portion of the lake distribution. At some point lake counts will decline given that a distinction is made between lakes and intermittent water bodies such as wetlands. The new CanVec lake database suggests that at least in Canada, this transition and the lower bound of small lakes occurs at a larger size class and at a lower frequency than previously proposed by Downing et al. (2006). At some point scaling rules have to break down and the results presented here suggest that this breakdown occurs at a larger lake size than assumed by Downing et al. (2006). Ultimately, this becomes a question of how one defines the difference between lakes and wetlands and whether this is easily determined on the ground and in the data.

3.5.2 Carbon Accumulation Estimates

Carbon accumulation for lakes in boreal Canada totalled between 1.65 and 2.34 Mt C yr^{-1}, representing 2-8% of the estimated 0.03–0.07 Pg of annual accumulation in lake sediments globally (Cole et al. 2007, Tranvik et al. 2009). When put into context, this represents the CO_2 emissions of ~ 300,000-450,000 thousand cars per year, or roughly 0.2-0.5% of Canada's 2007 per capita emissions and equates to each Canadian requiring a ~530 ha lake to offset personal emissions (Environmental Protection Agency 2000, World Bank 2011). If we assume a total carbon pool of 20 Pg (2 Mt x 10,000 years), it would only take 36 years for emissions in Canada to account for the total carbon stored in Canadian boreal lakes over the last 10,000 years.

These carbon estimates are a product of accumulation rates and lake distribution (size and count). The accumulation rates used here are based on Finnish lake data (Pajunen 2000, 2004) and are relatively conservative when compared with other published values (Mulholland and Elwood 1982, Molot and Dillon 1996, Dean and Gorham 1998, Einsele et al. 2001, Benoy et al. 2007, Tranvik et al. 2009). The Finnish data was used for several reasons:

1.) Published accumulation rates specifically for Canadian lakes are extremely limited and are generally restricted to small geographic areas, are based on a small number of lakes, do not necessarily represent the long term rate of accumulation, and in some cases rely on indirect measures (Ritchie 1989, Molot and Dillon 1996, Campbell et al. 2000, DesJardins et al. 2004, Squires et al. 2006, Benoy et al. 2007). The Finnish data is derived from the Northern Lake Survey (NLS) database, which consists of over 870 lakes and is considered to be statistically representative of lakes in Finland based on characteristics such as water chemistry, lake size, watershed size, land type, and surficial geology (Kortelainen et al. 2004). Of the two studies, Pajunen (2004) rates are more conservative, especially for the smallest lakes. The smallest lakes tend to have thinner sediment blankets and therefore reduced areal mass accumulation rates compared to their larger counterparts. This is offset to a certain extent by higher OC content. The thinner sediment blanket primarily is a product of reduced particulate inputs. Most very small lakes lack direct inputs such as streams or rivers, which act as conduits for particulate organic matter to enter the system. Moreover, small lakes in the boreal zone are commonly surrounded by peat and mire that effectively trap particulates from the watersheds that may otherwise end up in the lakes (Pajunen 2004).

2.) Boreal Canada and boreal Finland represent similar physiographic and climate conditions. Although Finland lies between 60˚ and 70˚ N latitude, north of most of boreal Canada, its climate is moderated based on proximity to the Baltic Sea and warm air flows from the Gulf Stream. As such, it shares the same K□ppen-climate classification as boreal Canada (Dfc- Continental Subarctic/Boreal) indicating similar vegetation type, precipitation and air temperatures (McKnight and Hess 2005).

3.) Unlike many studies, these accumulation rates represent the long term average since lake inception/deglaciation and not just surface sediments, which can be subject to further degradation/respiration. Furthermore, the Finnish rates were averaged over the entire lake surface area based on geophysical sonar surveys and multiple sediment cores. This accounts for the tendency of sediment focusing in the deepest parts of lakes, which is commonly neglected in other studies.

4.) Several other studies have observed a relationship between lake size and carbon accumulation rate (Mulholland and Elwood 1982, Dean and Gorham 1998). Generally, small to medium sized lakes (10-1,000 ha) accumulate carbon faster than their larger counterparts, likely a product of a higher catchment area to lake area ratio, concentrating allochthonous inputs into a smaller sediment area. The Finnish data was expressed in logarithmic size classes ranging from <10 to >10,000 hectares, allowing it to be directly applied to the size classified CanVec lake database.

Using the same Finnish data, Kortelainen et al. (2004) predicted an annual rate of carbon accumulation of 1.9-2.7 Mt C yr^{-1} for all boreal lakes globally, and a total carbon pool of 19-27 Pg over the past 10,000 years of which 11 Pg was in North America. When CanVec data was used to extrapolate these same rates, lakes in boreal Canada alone accounted for 16-23 Pg or 60-100% of the Kortelainen et al.'s (2004) estimated total carbon pool for all of the boreal zone. The difference between these estimates is largely a product of the difference in lake size distribution, illustrating the importance of accurate lake estimates when scaling lake processes such as carbon accumulation and storage to regional and global scales.

Compared to other studies, the estimates found here and by Kortelainen et al. (2004) for both carbon pool and annual accumulation are the most conservative. Molot and Dillon (1996) estimated a total carbon pool of 120 Pg over 5,000 years with an annual accumulation rate of 18-31 Mt C yr^{-1} for lakes throughout the boreal zone. These estimates were later modified by Benoy et al. (2007) to reflect only boreal Canada for a total carbon pool of 33 Pg. Globally, estimates of annual accumulation range from 30-70 Mt C yr^{-1} (Mulholland and Elwood 1982, Dean and Gorham 1998, Stallard 1998, Cole et al. 2007, Tranvik et al. 2009). If we assume lakes cover 3.2 million km^2 globally (the average of Downing et al. (2006), Lehner and Döll (2004) and Maybeck (1995) estimates), lakes in boreal Canada account for ~26% of global lake surface area while only accounting for 2-8% of the global carbon accumulation of lakes. This is a small amount when considering the large areal extent of lakes in Canada, and is mainly a product of the low accumulation rates used. Most studies have accumulation rates ranging from 5-100 g C m^{-2} yr^{-1} (Mulholland and Elwood 1982, Molot and Dillon 1996, Dean and Gorham 1998, Campbell et al. 2000). If the estimates produced here are accurate, lakes in boreal Canada accumulate and store moderate amounts of carbon when compared with other terrestrial carbon pools such as peatlands (113 Pg) and forest soils (65 Pg), and is about double that of living plant biomass (8 Pg) (Bhatti et al. 2003).

It is important to understand the limitations of extrapolating carbon accumulation rates based solely on lake size. Size dependence is only one of many factors which have been found to influence carbon accumulation rates. Catchment area to lake area ratio, water depth, DOC, nutrient levels, agricultural development, and watershed geology have all be found at some level to influence accumulation rates (Molot and Dillon 1996, Dean and Gorham 1998, Kortelainen et al. 2004, Squires et al. 2006, Benoy et al. 2007).

Ultimately, the CanVec database has allowed for a new look at carbon estimates in Canadian boreal lakes, especially at finer scales. In its current form, the CanVec database provides only two lake attributes (size and location). Future estimates of carbon accumulation could be refined with the inclusion of new datasets specific to Canadian lakes. Lake properties such as lake trophic status, water residence time, nutrient levels (DOC, Fe, P, and N-NO_3) and pH, as well as watershed properties such as watershed area, land use, and surficial geology could be added relatively easily and would all help to refine carbon accumulation estimates.

The utility of the CanVec database is not restricted to examining only carbon accumulation and could be used to examine other lake functions. Although boreal lakes serve as sinks by sequestering carbon in sediments, they are also commonly supersaturated in dissolved CO_2, and have been found to be large emitters of CO_2 to the atmosphere (Telmer and Veizer 1999, Algesten et al. 2004, Jonsson et al. 2007, Roehm et al. 2009). Furthermore, CO_2 evasion rates have been shown to reflect lake size dependence, suggesting that the CanVec lake database may also be useful in estimating other carbon processes such as regional evasion rates of CO_2 from Canadian boreal lakes (Kortelainen et al. 2006).

3.5.3 Thematic Products

3.5.3.1 Lake Size Distribution Maps

Slope values (*m*) from Eq. 4 are a reflection of the underlying lake distribution and measure the relative frequency of lakes across the five size classes. In effect, slope allows for a quick determination of the relative proportion of small lakes versus large depending on the steepness of the slope. Slope differs from other measures of lake distribution such as absolute frequency in that it expresses the relative proportion of lakes depending on their size and not simply the total number of lakes. The principle of using slope as an accurate measure of lake distribution is dependent on the observed linear increase in lake count going from large to small lakes on a logarithmic scale (Maybeck 1995, Downing et al. 2006). This log-linear relationship has been found for most regional lake distributions (excluding very small lakes) of the world to be ~ -1 for all but the most arid regions (Maybeck 1995). When slope is expressed spatially, differences in local lake distribution can be visualized, allowing for a better understanding of how lakes vary across the landscape.

A comparison of the four resolutions examined here (625 km^2, 2,500 km^2, 10,000 km^2 and 40,000 km^2) indicates that as cell size increases, slope values tend to better reflect the underlying lake distribution. This is supported collectively by an increase in mean R^2 and class membership counts, as well as a decrease in variation in both slope and class membership count with increasing cell size (Table 3.6). As cell size increases, the sampling area increases, improving the likelihood of capturing lakes within all size classes and helping to ensure that the slope value is an accurate reflection of the underlying distribution.

Slope value is least accurate when the cell size is either small and/or along geographic and coastal edges. This can be seen clearly at the smallest resolution with the presence of empty cell values and low class membership counts (Figure 3.12 A and B, Figure 3.14 A and B). Similarly, the most negative (extreme) slope values at all four resolutions generally occurred along edges such as the Canadian border and coastlines. In many of these cases the grid cell area covers regions where either lakes cannot exist (in the ocean) or there is insufficient/incomplete data (geographic borders of Canada). These cells generally correlate with low class membership counts further indicating that in these cases slope value alone is a poor representation of lake size distribution because of a poor fit. This "edge effect" highlights an important limitation to consider when comparing slope values from different regions and at different resolutions.

3.5.3.2 Carbon Accumulation Maps

Fitting a best fit line to the lake distribution allowed for more than just slope maps to be produced. Annual carbon accumulation was estimated based on the predicted number of lakes from the best fit line multiplied by the average lake size and accumulation rate for each size class (Eq. 5). Maps of carbon accumulation were produced for only the two largest cell resolutions because of the better fit with higher R^2 and class membership values (Table 3.6 and Figure 3.14). Estimates of carbon accumulation are a reflection of both the absolute frequency (total number) and relative frequency (slope) of lakes within each class of the underlying lake distribution. A closer look at how carbon accumulation varies with lake distribution shows that absolute frequency appears to influence accumulation more than relative frequency and at both 10,000 km^2 and 40,000 km^2 resolutions (Figure 3.13, Figure 3.15 Figure 3.16 and Figure 3.17). Total carbon accumulation is a product of the size dependent accumulation rate and lake frequency within each class. In areas with steeper (more negative) slopes, there will be a higher proportion of small lakes. Given that small lakes tend to accumulate more carbon per unit area, it might be expected that areas with a steeper slope would account for more accumulation than areas with shallower slopes. The carbon accumulation maps produced here show that the distribution of lakes within each size class alone does not drive carbon accumulation. Slope values indicate nothing of the total number of lakes present, only their relative proportions to each other. This indicates that while slope may provide unique and important insight into lake distribution, total number of lakes is the dominant driver of carbon accumulation. Another reason why slope by itself may be a poor indicator of carbon accumulation can be explained by size dependent accumulation rates. Larger lakes tend to accumulate much less carbon per unit area than small lakes. However, this low rate can be offset by a much larger overall areal extent (the Great lakes are a good example of this).

Total carbon accumulation in the boreal zone ranged from 1,610,884 to 2,808,750 t C yr^{-1} depending on cell resolution, average lake size per class, and accumulation rate (Figure 3.16, Figure 3.17). When the same parameters are compared between the two resolutions, estimates were 4 to 14% higher at the finer resolution than the coarser resolution. Considering the areal extent at both resolutions was the same, the difference indicates that resolution has an effect on carbon accumulation. A smaller cell size allows for fine scale differences in lake distribution and carbon accumulation to be resolved. However, as a consequence, the likelihood of lakes occurring across more than one cell increases. This can lead to lakes being counted more than once, inflating both frequency and accumulation estimates. Conversely, using a coarser cell resolution will decrease the likelihood of double counts occurring, improving both frequency and accumulation estimates at the cost of resolution at fine spatial scales. This is supported by the fact that at 40,000 km^2 accumulation estimates are closer to the tabular estimates derived directly from the lake database than at 10,000 km^2, but regional differences in accumulation are more difficult to recognize.

Correcting for double counts is difficult but possible in future work. One possibility is to fraction lake area proportionally across grid cells they overlap. This would allow for areal estimates to remain consistent but would not address the issue of double counts. Another possible option is to use the geometric center of lakes to determine a single grid cell for a lake to exist in. This would eliminate double counts but would also have the undesirable effect of misrepresenting the spatial footprint of larger lakes that occur over multiple cells. Regardless, slope and carbon accumulation maps presented here provide a unique perspective on how lake distribution varies across Canada and where potential carbon accumulation hotspots exist in the boreal zone. At the very least, these maps provide a new thematic representation of Canada's boreal lakes and demonstrate how spatial data can provide new insight into lake distribution and function and how these may change in the future.

3.6 Conclusions

The construction of a new lake database using CanVec data has provided a new perspective on lake distribution in Canada. This database has allowed for direct comparisons to previous estimates of Canadian lake distribution while also allowing the validity of scaling rules used in global estimates to be scrutinized. The utility of this database has been illustrated by the production of new estimates of carbon accumulation in Canadian boreal lake sediments and the production of new thematic products describing lake size distribution and carbon accumulation. Lake size distribution maps provide a new perspective on how lakes vary across the landscape with respect to their size.

The utility of the CanVec lake database is not restricted to carbon accumulation. This database and its derived products represent a new parameterization of surface water in Canada. Until now, most CGCM's have had little to no representation of terrestrial surface waters as a feedback. With the expected changes in precipitation and air temperature for the boreal zone and the rest of Canada, the CanVec lake database could provide new and important feedbacks to CGCM systems at regional scales in Canada. This would not only help to refine the current role of Canadian lakes as sites of carbon storage, transport, and flux between the terrestrial biosphere, hydrosphere, and atmosphere, but also how this role may change in the future.

References

Algesten, G., S. Sobek, A. K. Bergstrom, A. Agren, L. J. Tranvik, and M. Jansson. 2004. Role of lakes for organic carbon cycling in the boreal zone. Global Change Biology 10:141-147.

Almeida-Filho, R., Y. E. Shimabukuro, A. Rosenqvist, and G. A. Sanchez. 2009. Using dual-polarized ALOS PALSAR data for detecting new fronts of deforestation in the Brazilian Amazonia. International Journal of Remote Sensing 30:3735-3743.

ALOS K&C Initiative. 2007. The ALOS Kyoto & Carbon Initiative Science Plan v. 3.0.

Avis, C. A., A. J. Weaver, and K. J. Meissner. 2011. Reduction in areal extent of high-latitude wetlands in response to permafrost thaw. Nature Geoscience 4:444-448.

Benoy, G., K. Cash, E. McCauley, and F. Wrona. 2007. Carbon dynamics in lakes of the boreal forest under a changing climate. Environmental Reviews 15:175-189.

Bhatti, J. S., G. C. Van Kooten, M. J. Apps, L. D. Laird, I. D. Campbell, C. Campbell, M. R. Turetsky, Z. Yu, and E. Banfield. 2003. Carbon balance and climate change in boreal forests. Chapter 20. *In* Towards Sustainable Managment of the Boreal Forest. *Edited by* P. J. Burton, C. Messier, D. W. Smith, and W. L. Adamowicz. NRC Research Press, Ottawa, Ontario, Canada. pp. 799-855.

Campbell, I. D., C. Campbell, D. H. Vitt, D. Kelker, L. D. Laird, D. Trew, B. Kotak, D. LeClair, and S. Bayley. 2000. A first estimate of organic carbon storage in Holocene lake sediments in Alberta, Canada. Journal of Paleolimnology 24:395-400.

CanVec Hydrographic Dataset. 2007. CanVec Hydrographic dataset. Produced by Natural Resources Canada. (ftp2.cits.rncan.gc.ca/pub/canvec/).

Christensen, J.H., B. Hewitson, A. Busuioc, A. Chen, X. Gao, I. Held, R. Jones, R.K. Kolli, W.-T. Kwon, R. Laprise, V. Magaña Rueda, L. Mearns, C.G. Menéndez, J. Räisänen, A. Rinke, A. Sarr and P. Whetton, 2007. Regional Climate Projections. *In* Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. *Edited by* Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. pp 847-940.

Clair, T. A., J. M. Ehrman, and K. Higuchi. 1999. Changes in freshwater carbon exports from Canadian terrestrial basins to lakes and estuaries under a 2xCO_2 atmospheric scenario. Global Biogeochemical Cycles 13:1091-1097.

Cole, J. J., Y. T. Prairie, N. F. Caraco, W. H. McDowell, L. J. Tranvik, R. G. Striegl, C. M. Duarte, P. Kortelainen, J. A. Downing, J. J. Middelburg, and J. Melack. 2007. Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. Ecosystems 10:171-184.

Dean, W. E., and E. Gorham. 1998. Magnitude and significance of carbon burial in lakes, reservoirs, and peatlands. Geology 26:535-538.

DesJardins, M. J., K. Telmer, and S. Beauchamp. 2004. Apportioning atmospheric pollution to Canadian and American sources in Kejimkujik National Park, Nova Scotia, using Pb isotopes in precipitation. Atmospheric Environment 38:6875-6881.

Downing, J. A., Y. T. Prairie, J. J. Cole, C. M. Duarte, L. J. Tranvik, R. G. Striegl, W. H. McDowell, P. Kortelainen, N. F. Caraco, J. M. Melack, and J. J. Middelburg. 2006. The global abundance and size distribution of lakes, ponds, and impoundments. Limnology and Oceanography 51:2388-2397.

Einsele, G., J. P. Yan, and M. Hinderer. 2001. Atmospheric carbon burial in modern lake basins and its significance for the global carbon budget. Global and Planetary Change 30:167-195.

Environmental Protection Agency. 2000. Emission Facts. United States Environmental Protection Agency- Office of Transportation and Air Quality. EPA420-F-00-013. (www.epa.gov/oms/consumer/f00013.pdf).

Herdendorf, C. E. 1982. Large lakes of the world. Journal of Great Lakes Research 8:379-412.

Holmén, K. 1992. The Global Carbon Cycle. Chapter 11. *In* Global Biogeochemical Cycles. *Edited by* Butcher, S. S., R. J. Charlson, G. H. Orians, and G. V. Wolfe. Academic Press, London. pp. 239-261.

Jenson, J. R. 2005. Introductory Digital Remote Image Processing: A Remote Sensing Perspective. 3rd edition. Prentice Hall, Upper Saddle River, NJ, USA. pp. 526.

Jenson, J. R. 2006. Remote Sensing of the Environment: An Earth Resource Perspective. 2nd edition. Prentice Hall, Upper Saddle River, NJ, USA. pp. 608.

Jonsson, A., G. Algesten, A. K. Bergstrom, K. Bishop, S. Sobek, L. J. Tranvik, and M. Jansson. 2007. Integrating aquatic carbon fluxes in a boreal catchment carbon budget. Journal of Hydrology 334:141-150.

Kalff, J. 2002. Limnology: Inland Water Ecosystems. Prentice Hall, Upper Saddle River, NJ, USA. pp. 592.

Kortelainen, P., H. Pajunen, M. Rantakari, and M. Saarnisto. 2004. A large carbon pool and small sink in boreal Holocene lake sediments. Global Change Biology 10:1648-1653.

Kortelainen, P., M. Rantakari, J. T. Huttunen, T. Mattsson, J. Alm, S. Juutinen, T. Larmola, J. Silvola, and P. J. Martikainen. 2006. Sediment respiration and lake trophic state are important predictors of large CO_2 evasion from small boreal lakes. Global Change Biology 12:1554-1567.

Kozlenko, N., and M. O. Jeffries. 2000. Bathymetric mapping of shallow water in thaw lakes on the North Slope of Alaska with spaceborne imaging radar. Arctic 53:306-316.

Lehner, B., and P. Döll. 2004. Development and validation of a global database of lakes, reservoirs and wetlands. Journal of Hydrology 296:1-22.

Maybeck, M. 1995. Global distribution of lakes. In Physics and Chemistry of lakes. Edited by A. Lerman, D. M. Imboden, J. R. Gat. 2nd edition. Spinger-Verlag, Berlin Heidelberg. pp. 1-32.

McKnight, T. L., and D. Hess. 2005. Climate Zones and Types: The Köppen System. *In* Physical Geography: A Landscape Appreciation. 8th Edition. Prentice Hall, Upper Saddle River, NJ, USA. pp. 640.

Molot, L. A., and P. J. Dillon. 1996. Storage of terrestrial carbon in boreal lake sediments and evasion to the atmosphere. Global Biogeochemical Cycles 10:483-492.

Mulholland, P., and J. W. Elwood. 1982. The role of lake and reservoir sediments as sinks in the perturbed global carbon cycle. Tellus 34:490-499.

Paillou, P., S. Lopez, T. Farr, and A. Rosenqvist. 2010. Mapping Subsurface Geology in Sahara Using L-Band SAR: First Results from the ALOS/PALSAR Imaging Radar. Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing 3:632-636.

Pajunen H. 2000. Lake sediments; their carbon store and related accumulation rates. In Carbon in Finnish Lake Sediments, Pajunen H (ed.). Special Paper no. 29, Geological Survey of Finland. pp. 39–69.

Pajunen, H. 2004. Lake sediments as a store of dry matter and carbon. Report of Investigation 160 (In Finnish with English Summary). Geological Survey of Finland, Espoo. pp. 291-308.

Raney, K. 1998. Radar Fundamentals: Technical Perspective. Chapter 2. *In* Principles and Applications of Imaging Radar, Volume 3, Manual of Remote Sensing. *Edited by* F. M. Henderson, A. J. Lewis. John Wiley & Sons Inc., New York. pp. 9-130.

Riordan, B., D. Verbyla, and A. D. McGuire. 2006. Shrinking ponds in subarctic Alaska based on 1950-2002 remotely sensed images. Journal of Geophysical Research-Biogeosciences 111:11.

Ritchie, J. C. 1989. Carbon Content of Sediments of Small Reservoirs. Water Resources Bulletin 25:301-308.

Roehm, C. L., Y. T. Prairie, and P. A. del Giorgio. 2009. The pCO_2 dynamics in lakes in the boreal region of northern Quebec, Canada. Global Biogeochemical Cycles 23, GB3013, doi:10.1029/2008GB003297.

Rosenqvist, A., M. Shimada, N. Ito, and M. Watanabe. 2007. ALOS PALSAR: A Pathfinder mission for global-scale monitoring of the environment. Ieee Transactions on Geoscience and Remote Sensing 45:3307-3316.

Smith, L. C., Y. Sheng, G. M. MacDonald, and L. D. Hinzman. 2005. Disappearing Arctic lakes. Science 308:1429-1429.

Squires, M. M., D. Mazzucchi, and K. J. Devito. 2006. Carbon burial and infill rates in small Western Boreal lakes: physical factors affecting carbon storage. Canadian Journal of Fisheries and Aquatic Sciences 63:711-720.

Stallard, R. F. 1998. Terrestrial sedimentation and the carbon cycle: Coupling weathering and erosion to carbon burial. Global Biogeochemical Cycles 12:231-257.

Stow, D. A., A. Hope, D. McGuire, D. Verbyla, J. Gamon, F. Huemmrich, S. Houston, C. Racine, M. Sturm, K. Tape, L. Hinzman, K. Yoshikawa, C. Tweedie, B. Noyle, C. Silapaswan, D. Douglas, B. Griffith, G. Jia, H. Epstein, D. Walker, S. Daeschner, A. Petersen, L. M. Zhou, and R. Myneni. 2004. Remote sensing of vegetation and land-cover change in Arctic Tundra Ecosystems. Remote Sensing of Environment 89:281-308.

Telmer, K. H., and M. P. F. Costa. 2007. SAR-based estimates of the size distribution of lakes in Brazil and Canada: a tool for investigating carbon in lakes. Aquatic Conservation-Marine and Freshwater Ecosystems 17:289-304.

Telmer, K., M. Costa, J. MacGregor. 2008. K&C Science Report- Phase 1, Global Lakes Census. (www.eorc.jaxa.jp/ALOS/en/kyoto/phase_1/KC-Phase1-report_Telmer.pdf).

Telmer, K., and J. Veizer. 1999. Carbon fluxes, pCO_2 and substrate weathering in a large northern river basin, Canada: carbon isotope perspectives. Chemical Geology 159:61-86.

Tranvik, L. J., J. A. Downing, J. B. Cotner, S. A. Loiselle, R. G. Striegl, T. J. Ballatore, P. Dillon, K. Finlay, K. Fortino, L. B. Knoll, P. L. Kortelainen, T. Kutser, S. Larsen, I. Laurion, D. M. Leech, S. L. McCallister, D. M. McKnight, J. M. Melack, E. Overholt, J. A. Porter, Y. Prairie, W. H. Renwick, F. Roland, B. S. Sherman, D. W. Schindler, S. Sobek, A. Tremblay, M. J. Vanni, A. M. Verschoor, E. von Wachenfeldt, and G. A. Weyhenmeyer. 2009. Lakes and reservoirs as regulators of carbon cycling and climate. Limnology and Oceanography 54:2298-2314.

von Wachenfeldt, E., S. Sobek, D. Bastviken, and L. J. Tranvik. 2008. Linking allochthonous dissolved organic matter and boreal lake sediment carbon sequestration: The role of light-mediated flocculation. Limnology and Oceanography 53:2416-2426.

von Wachenfeldt, E., and L. J. Tranvik. 2008. Sedimentation in boreal lakes - The role of flocculation of allochthonous dissolved organic matter in the water column. Ecosystems 11:803-814.

Wiken, E. 1986. Terrestrial ecozones of Canada. Lands directorate, Environment Canada, Ottawa.

World Bank. 2011. World Development Indicators. World Bank. (www.data.worldbank.org/country/canada).