K&C Science Report – Phase 1
Global Lake Census

Kevin Telmer[1] and Maycira Costa[2] and Jamie MacGregor[1]
1.) School of Earth and Ocean Sciences, 2.) Department of Geography, University of Victoria, 3800 Finnerty Rd., Victoria, BC Canada V8P 5C2
Email: ktelmer@uvic.ca

Abstract— **Lake size is a strong control on lake function and on how lakes interact with the environment. For example, lake size is strongly related to carbon burial rates in lake sediments. Lake size distribution (how many small, medium, and large lakes, occur per unit area) can be used to extrapolate lake function to landscapes at local, regional, and global scales. Lake size distribution can be parameterized as:**

m = log(number of lakes/unit area) / log(lake size) – b

Where m less than -1 (m < -1) indicates that small lakes occupy greater area than large lakes and vice versa. A digital database describing lake size distribution and lake shape can therefore be used to investigate and estimate lake-landscape interactions at various scales but this database does not currently exist.

Here the capability of high resolution ALOS PALSAR data is examined for mapping lakes across multiple study sites covering over 100,000 km^2 of the Canadian landscape. Classification of lakes versus land was done by expert systems based threshold analysis and also by using derivatives of the frequency histograms of the SAR image digital numbers. The expert system approach produced better accuracy but cannot be applied automatically. Accuracy varied between sample regions and across the lake size ranges, however; overall, the PALSAR lake classification differed from existing hydrographic spatial datasets for total lake area by up to 18%. Comparisons between datasets cannot be perfect because they were not created over the same time period. Due to its short time of collection and singular data source, the ALOS data is superb for capturing current conditions, establishing a baseline, and comparing to future conditions.

Estimates of carbon burial for the sample regions were made based on the lake size distribution determined from the PALSAR classification and on literature estimates of carbon accumulation for various lake sizes under different climatic conditions. Carbon accumulation in lake sediments in the Canadian boreal region alone could account for as much 13.4 Mt of CO_2 equivalent per 100,000 km^2 per year.

Index Terms—**ALOS PALSAR, K&C Initiative, wetlands, carbon, lakes.**

I. INTRODUCTION

Carbon accumulation in lake sediments has long been recognized as an important component of the global carbon cycle [1-6]. Of the many parameters which influence carbon accumulation rates, lake size has been identified as a simple predictor of carbon burial rates in lake sediments [3, 5, 7]. A database describing lake size distribution can therefore be used to investigate and estimate lake-landscape interactions.

Accurate and reliable lake census data is a fundamental first step in extrapolating carbon accumulation rates to regional and global scales. The majority of available hydrographic data is generally limited in both spatial and temporal resolution (e.g. Global Lakes and Wetlands Database [8]). This is because the majority of these datasets do not come from a single data source produced by a single method nor are they acquired from a single time period.

In Canada the CanVec hydrographic dataset produced by Natural Resources Canada (NRCan) constitutes one of the most comprehensive and freely available hydrographic datasets [9]. The CanVec dataset demonstrates very good spatial accuracy. However, its temporal resolution is limited; especially in remote areas and can span up to 50 years or more. This can limit direct application to studies examining regions undergoing potential hydrological change over short time scales such as the northern boreal forest of Canada.

The Phased Array L-band Synthetic Aperture Radar (PALSAR) sensor on board the Advanced Land Observing Satellite (ALOS) launched in early 2006 by the Japanese Aerospace Exploration Agency (JAXA) has provided a unique opportunity to construct a global lake database. PALSAR's all weather night and day acquisition capability affords continuous global coverage and allows for the construction a database from a single source over a single period of time (24 months). One of the inherent benefits of L-band radar systems such PALSAR is its capacity to detect water bodies. L-band radar can penetrate through sparse vegetation while also being less sensitive to water surface roughness. This makes it ideal for differentiating water from land under variable conditions.

A. Relevance to the K&C drivers

The objective of the global lake census is to establish a baseline map of lake size distribution and how this is controlled at regional and global levels. An up to date and temporally constrained lake database paired with existing and new carbon accumulation rates will allow for an improved first order estimate of the role lake sediments play in the storage of carbon.

Under the framework of JAXA's K&C Initiative and the three C's: Conventions, Carbon and Conservation- the global lake census will help to determine the role lakes play in the terrestrial carbon cycle now, and with changes that may occur under future climate conditions. This information in turn can be used to guide policy on carbon trading, wetland conservation, and creating and enforcing conventions needed to maximize carbon uptake from the atmosphere as well as ecological protection in general.

B. *Work approach*

The methodological approach in developing the global lakes census has progressed in a series of steps:

i.) Lake classification

We utilized high resolution PALSAR FBS and FBD 12.5 meter images to produce a binary classification of lakes and land for 9 pilot regions across Canada (Fig. 1). PALSAR images were filtered and reduced to 8-bit to minimize speckle and data depth. A single threshold digital number (DN) was selected for each region as cut-off value between lakes and land. This simple classification method has been applied successfully in previous studies identifying water bodies in RADAR imagery [7, 10]. The main advantage of the threshold approach over more complex classification methods is computational efficiency. This is especially important for very large datasets such as the one proposed here.

Threshold selection was first attempted quantitatively based on a first derivatives approach. This method relies on the bimodal nature of the RADAR imagery where the first mode represents water bodies and the second mode represents land (Fig. 2). The first derivative quickly identifies the inflection point between the two modes and provides a logical starting point for a threshold DN. Histograms from each pilot site demonstrated the same characteristic bimodal shape. However, each differed in how quickly and at what DN value the transition between the lakes and land occurred. This made it difficult to rely solely on first derivatives to produce an accurate classification. Therefore an expert based tuning approach was employed to improve classification accuracy. This involved increasing or decreasing the DN around the first derivative inflection point until better classification accuracy was achieved.

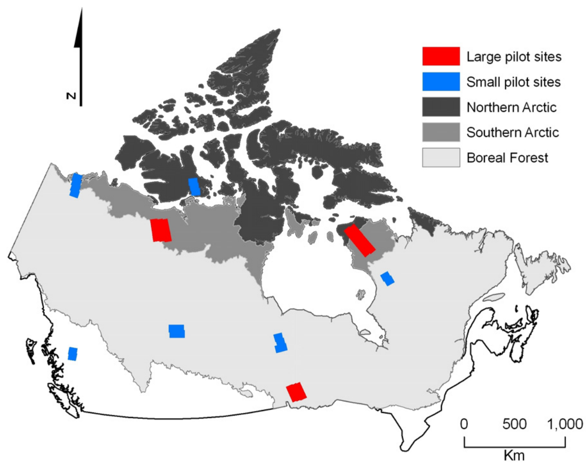

Figure 1. Pilot regions used for the methodological development of lake classification.. Each pilot region consisted of 12.5 m PALSAR FBS FBD imagery. Corresponding AVNIR-2, SPOT and Landsat imagery was also used when available to aid in accuracy assessment. All 9 regions were used in the development of the lake classification and in the extrapolation of carbon accumulation rates to regional scales. The three large pilot sites (shown in red) were used as test sites for the derived lake size distribution thematic map.

ii.) Accuracy assessment

Direct field campaigns have not been feasable up to this point becasue of the remote location and large geographic distance between study sites. Therefore classification accuracy assessment was limited to compairisons to existing hydrographic data (CanVec) and ancilliray satellite imagery from the Advanced Visible and Near-Infrared Radiometer (AVNIR-2) on board the ALOS platform, as well as SPOT 4/5 (Satellite Pour l'Observation de la Terre) and Landsat 7 imagery. When utilizing anicilliary data, the ALOS AVNIR-2 imagery provided the most accurate assessment becasue of the near simultaneous aquistion time to that of the PALSAR images. Lake area and count from both the PALSAR lake classification and the CanVec lake polygons where divided across 5 size classes ranging from less than 0.01 km^2 to over 100 km^2 allowing for direct comparison.

iii.) Carbon accumulation estimates

A preliminary assessment of carbon accumulation for Canadian lake sediments was calculated based on the PALSAR lake classification data and published lake sediment carbon accumulation rates. We used lake sediment accumulation rates from 140 Finnish boreal lakes covering a wide range of landscapes and lake sizes [5]. Finnish accumulation rates were selected because they were based on a large lake census and vary as a function of lake size. This allowed for easy application across our PALSAR lake size distributions, simplifying the process. A large part of Canada has similar physiographic conditions to boreal Finland; however there are many regions which are relatively dissimilar. In order to apply the Finnish data to these areas, a scale factor was introduced in an attempt to make the carbon accumulation rates more representative of the Canadian landscape. Here we based the scale factor on a simplified version of the terrestrial ecozones of Canada [11] to produce three regions: the boreal forest, the southern arctic and the northern arctic (Fig. 1). Combined, these regions account for almost 86% of the Canadian landmass and represent some of the major shifts in the physiographic conditions. The remainder of Canada represents some challenging and diverse landscape complexities; therefore they have been left out of the analysis for now. Future regional carbon estimates on the Canada-wide 50 meter mosaic will include other carbon accumulation rate estimates from the literature, as they come in, along with new field based measurements.

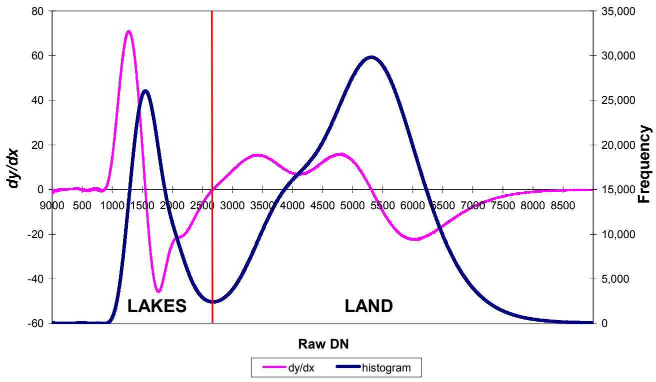

Figure 2. An example of the first derivative approach to threshold value selection. The inflection point between the first mode (water) and the second mode (land) is where the first derivative is equal to zero shown by the red vertical line. This provides a starting threshold for a classification and can be tuned according to an initial accuracy assessment.

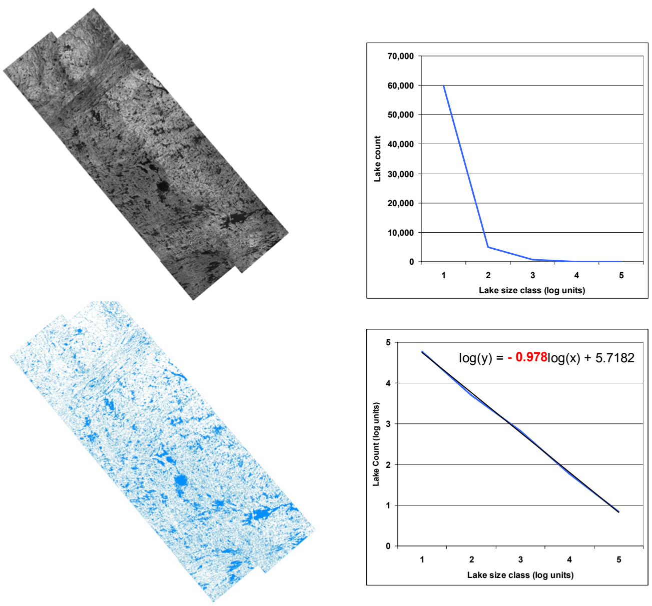

Figure 3 **Left:** The PALSAR mosaic from Northern Quebec, Ungava Peninsula and the corresponding threshold based lake classification. **Right:** Frequency of lakes within each size class (class **1:** < 0.1 km^2, class **2**: 0.1 – 1.0 km^2, class **3:** 1.0 – 10 km^2, class **4:** 10 – 100 km^2, class **5:** > 100 km^2). Lake count increases tenfold with decreasing lake size class. When frequency is log scaled the data becomes linear and can be fitted with linear best fit line where slope (shown in red) is representative of lake size distribution [12]. ALOS K&C © *JAXA/METI.*

iv.) Derived thematic product

The PALSAR lake classification describes the lake size distribution for each of the pilot areas. For the three large pilot areas (Fig. 1), size distribution was parameterized and thematically mapped to visually show how it changes across the landscape. The density of lakes generally increases from large lakes to small lakes. This increase is roughly tenfold when going from one lake size class to the next smaller class at a log scale [12, 13]. This relationship can therefore be defined by a simple log linear regression (1) (Fig. 3).

$$\mathbf{log(lake\ density) = mlog(lake\ size) + b} \qquad (1)$$

The slope (m) of the line will vary as a function of the underlying lake size distribution. A more negative slope is indicative of proportionally more small lakes whereas a more positive slope value indicates proportionally more large lakes. This simple relationship can be utilized to thematically express how lake size distribution changes across the landscape.

A grid with 100 km^2 cells was overlain on top of the lake classification for the three large pilot sites. For each grid cell the underlying lake size distribution was log transformed and fitted with the log linear regression line (1). The resulting slope values were then assigned to the corresponding grid cell (Fig.4). A cell size of 100 km^2 was selected for evaluation purposes. A final grid cell size for the 50m mosaic is under evaluation.

C. *Satellite and ground data*

i.) ALOS data

Analysis thus far has focused on 12.5 meter (level 1.5 processing) Fine Beam Dual (FBD) and Fine Beam Single (FBS) polarization PALSAR imagery obtained from ALOS User Interface Gateway (AUIG). All images were acquired during the summer months- between June and August- of 2007 and 2008 to minimize ice cover interference in the far north. Corresponding AVNIR-2 images (processing level 1-B2) where also obtained from AUIG for the same time period when available.

ii.) Other data

The CanVec hydrographic dataset is produced by NRCan and was obtained through the Geogratis web portal (www.geogratis.ca). The CanVec dataset is relatively new and free cartographic reference product, although the source data spans more than 50 years. CanVec aims to accurately represent topographic entities across the Canadian landmass with thematic information grouped into 11 distribution themes- including hydrographic data. Data originates from the best available sources; mainly the National Topographic Data Base (NTDB), the Geobase initiative, and recent updates from Landsat 7 imagery.

The SPOT 4/5 and Landsat 7 imagery were obtained from the Geobase web portal (www.geobase.ca) for the pilot

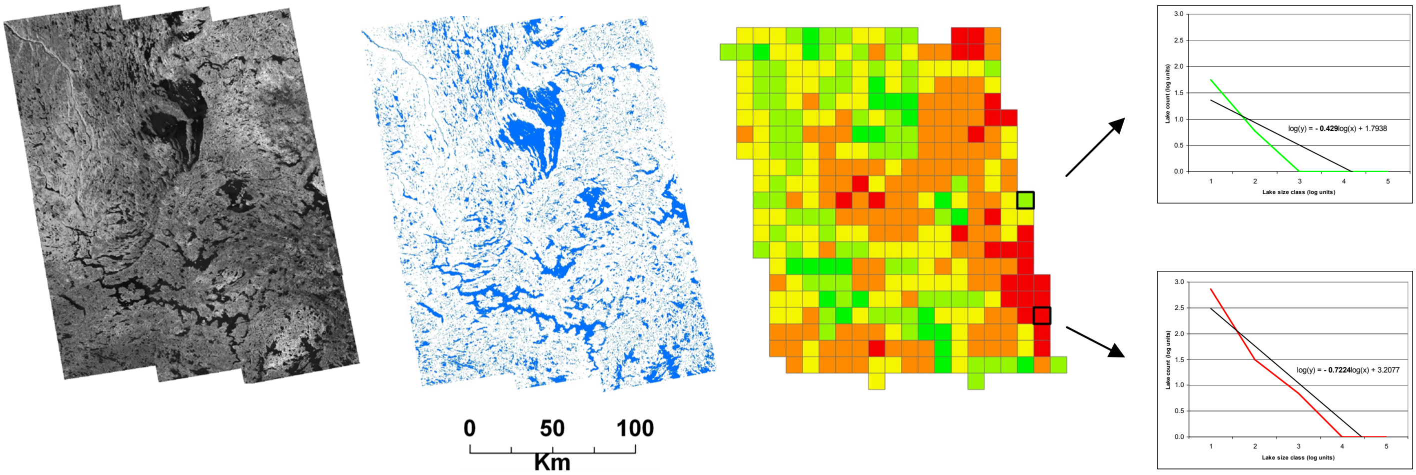

Figure 4. The development of a thematic lake size distribution map. A PALSAR image is first classified for lakes and subsequently grided to a specific cell size (in this case 100 km^2). For each grid cell the underlying lake size distribution was log transformed and fitted with the log-log linear regression line (1). The resulting slope values are then assigned to the corresponding grid cell as a raster. ALOS K&C © JAXA/METI.

Table 1. Summary lake count estimates for PALSAR lake polygons compared to CanVec lake polygons across six size classes for the 6 small pilot areas.

Pilot area	Lake Area (km^2)	< 0.1	0.1 - 1.0	1.0 - 10	10 - 100	100 <	Total
Alberta	PALSAR	20.42	58.96	118.41	151.82	205.52	555.13
	CanVec	23.23	57.51	124.98	173.20	206.54	585.46
	Difference	-2.82	1.45	-6.57	-21.38	-1.02	-30.34
	(%)	**-12.12**	**2.52**	**-5.25**	**-12.35**	**-0.49**	**-5.18**
British Columbia	PALSAR	20.97	47.83	53.17	24.20	215.03	361.20
	CanVec	11.92	36.91	63.52	27.39	228.22	367.95
	Difference	9.06	10.92	-10.36	-3.18	-13.19	-6.75
	(%)	**76.01**	**29.60**	**-16.31**	**-11.62**	**-5.78**	**-1.83**
Manitoba	PALSAR	55.44	276.06	326.00	374.69	945.66	1977.86
	CanVec	58.70	288.17	300.76	445.31	997.38	2090.32
	Difference	-3.26	-12.11	25.24	-70.62	-51.72	-112.47
	(%)	**-5.55**	**-4.20**	**8.39**	**-15.86**	**-5.19**	**-5.38**
McKenzie Delta	PALSAR	228.78	587.29	478.88	243.35	507.03	2045.32
	CanVec	281.43	654.52	543.53	542.11	248.34	2269.93
	Difference	-52.65	-67.23	-64.66	-298.76	258.69	-224.61
	(%)	**-18.71**	**-10.27**	**-11.90**	**-55.11**	**104.17**	**-9.89**
North West Territories	PALSAR	214.08	457.29	655.90	436.50	432.70	2196.47
	CanVec	286.63	571.10	715.54	358.31	574.09	2505.68
	Difference	-72.55	-113.81	-59.64	78.18	-141.39	-309.21
	(%)	**-25.31**	**-19.93**	**-8.33**	**21.82**	**-24.63**	**-12.34**
Victoria Island	PALSAR	76.32	116.20	78.78	138.74	0.00	410.04
	CanVec	113.57	139.63	82.84	164.23	0.00	500.27
	Difference	-37.25	-23.43	-4.06	-25.49	0.00	-90.23
	(%)	**-32.80**	**-16.78**	**-4.90**	**-15.52**	**0.00**	**-18.04**
ELA	PALSAR	92.16	379.71	501.66	600.46	428.94	2002.93
	CanVec	92.34	383.66	526.56	534.68	669.39	2206.62
	Difference	-0.19	-3.94	-24.90	65.78	-240.45	-203.69
	(%)	**-0.20**	**-1.03**	**-4.73**	**12.30**	**-35.92**	**-9.23**

areas. Images cover a time period from 2005 to present for SPOT 4/5 and 1999-2003 for Landsat 7.

II. Results and Summary

i.) Lake classification and Accuracy assessment

The PALSAR lake classification generally underestimated total lake area when compared directly to CanVec hydrographic data (Table 1). Differences in total lake area between CanVec dataset and the PALSAR lake classification ranged from 1.83 to 18.04%. Some of the larger differences within each size class between CanVec and PALSAR can be explained by simple class shifts. For example in the McKenzie Delta the PALSAR classification overestimates the area of lakes in the > 100 km^2 class by 104.17% compared to CanVec. If this difference is shifted into the next lower size class (10 -100 km^2) it accounts for most of this discrepancy. This is because of differences in the way PALSAR and CanVec define lakes. If a lake or water body has very narrow channels, these regions can be lost in PALSAR classification because of resolution limitations. This can result in a single lake being split into two or more lakes resulting in a different size classification. However, it may be that a lake that is split like this is behaving more like two small lakes with respect to carbon accumulation and so a PALSAR classification may be superior for this application. This is under evaluation.

Comparisons between datasets cannot be perfect because they are not time synchronous. The hydrographic data spans up to 50 years whereas the ALOS data comes from one short period of 24 months. This relatively short collection time and single data source give the ALOS product a strong advantage when comparing the current conditions as a baseline to changes in future conditions.

ii.) Carbon accumulation estimates

A summary of preliminary carbon accumulation estimates for Canadian lake sediments is shown below in Table 2. Based on these estimates for every 100,000 km^2 of boreal, south arctic and north arctic a total of 3.65, 2.45 and 0.48 Mt of carbon could be accumulating every year in Canadian lakes. This is equivalent to the CO_2 emissions of over ***4.5 million cars*** per year. Carbon estimates here were restricted to a relatively small area of 100,000 km^2 because of the limited footprint covered by the pilot areas. With improvements to carbon accumulation estimates and a complete lake size distribution based on the 50m PALSAR mosaic these estimates will be refined and extended to represent all of Canada.

Table. 2 PALSAR derived carbon accumulation in Canadian lakes per 100,000 km2 per year

Lake size km^2	Boreal	South Arctic	North Arctic
<0.1	0.15	0.24	0.08
0.1-1	1.07	0.94	0.20
1-10	1.10	0.79	0.10
10-100	0.76	0.28	0.11
>100	0.57	0.20	0.00
Total Carbon in megatonnes (Mt)	3.65	2.45	0.48
Mt CO2 Equivalent	13.39	9.00	1.77
Number of Cars[1]	2,579,000	1,733,000	342,000

1Based on EPA average annual emissions for passenger cars, April 2000. (http://www.epa.gov/oms/consumer/f00013.pdf)

Carbon accumulation rates for each size class are based on Pajunen 2004 [5]

iii.) Derived thematic product

The derived thematic slope maps for the three large pilot areas are shown in Figure 5. Each coloured box represents the lake size distribution of the underlying grid cell. Red grid cells indicate a higher proportion of small lakes to large lakes whereas green grid cells represent areas where there are more large lakes than small lakes. Carbon accumulation in lakes varies as a function of lake size [5]. With a higher density of small lakes, red grid cells will accumulate more carbon than green grid cells. Mapping lake distribution as shown allows for easy visual identification of high lake density regions and should allow for easy extrapolation of carbon accumulation rates to regional scales.

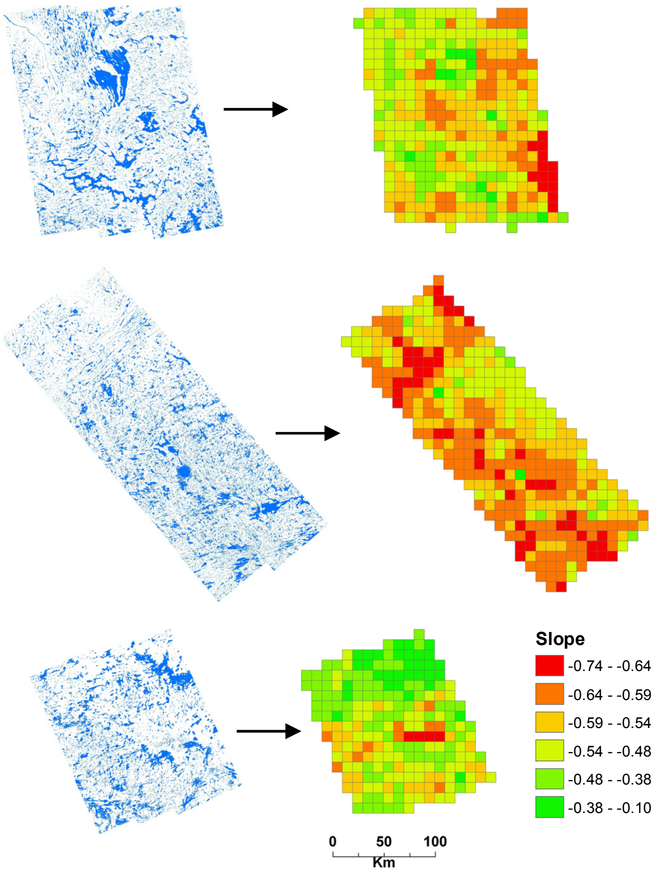

Figure 5. Lake classifications (left) and the derived lake size distribution maps (right) for each of the three large pilot areas (top to bottom): the North West Territories, Northern Quebec on the Ungava Peninsula, and the Experimental lakes in south western Ontario. Each coloured grid cell represents the lake size distribution with red indicating areas where there are more small lakes than large, and green vice versa. The grid cells were classified according to the slope of the regression fit to a classification of lake size as per Meybeck (1995) [13]. ALOS K&C © JAXA/METI.

iv.) Future work

An initial lake classification for all 9 pilot sites has been completed resulting in a digital map describing the number, size, and location of lakes. Accuracy has been validated by comparisons to other existing digital lake databases and corresponding high-resolution satellite imagery. Additionally the derived slope maps have been completed for the three large pilot sites.

Preliminary estimates of carbon accumulation in lakes have also been completed for the pilot study areas. This has provided an initial assessment of lake carbon accumulation which has been scaled up to 100,000 km^2 regions of Canada. However, these preliminary estimates are limited by the small footprint covered by the pilot sites and will improve greatly with the utilization of the Canada wide coverage 50m PALSAR mosaic to be completed during the extension phase.

Acknowledgements

This work has been undertaken within the framework of the JAXA Kyoto & Carbon Initiative. ALOS PALSAR data have been provided by JAXA EORC.

References

[1] W. E. Dean and E. Gorham, "Magnitude and significance of carbon burial in lakes, reservoirs, and peatlands," *Geology,* vol. 26, pp. 535-538, Jun 1998.

[2] G. Einsele, J. P. Yan, and M. Hinderer, "Atmospheric carbon burial in modern lake basins and its significance for the global carbon budget," *Global and Planetary Change,* vol. 30, pp. 167-195, Oct 2001.

[3] P. Kortelainen, H. Pajunen, M. Rantakari, and M. Saarnisto, "A large carbon pool and small sink in boreal Holocene lake sediments," *Global Change Biology,* vol. 10, pp. 1648-1653, Oct 2004.

[4] P. Mulholland and J. W. Elwood, "The role of lake and reservoir sediments as sinks in the perturbed global carbon cycle," *Tellus,* vol. 34, pp. 490-499, 1982.

[5] H. Pajunen, "Lake sediments as a store of dry matter and carbon," in *Report of Investigation 160 (In Finnish with English Summary)* Espoo: Geological Survey of Finland, 2004, p. 308.

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

[7] K. H. Telmer and M. P. F. Costa, "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,* vol. 17, pp. 289-304, May 2007.

[8] B. a. P. D. Lehner, "Development and validation of a global database of lakes, reservoirs and wetlands," *Journal of Hydrology,* vol. 296, pp. 1-22, 2004.

[9] "CanVec- http://ftp2.cits.rncan.gc.ca/pub/canvec/doc/CanVec_product_specifications_en.pdf," Natural Resources Canada, 2007.

[10] N. Kozlenko and M. O. Jeffries, "Bathymetric mapping of shallow water in thaw lakes on the North Slope of Alaska with spaceborne imaging radar," *Arctic,* vol. 53, pp. 306-316, Sep 2000.

[11] E. Wiken, "Terrestrial ecozones of Canada," 1986.

[12] M. Maybeck, *Global distribution of lakes*, 2nd ed. New York: Spinger-Verlag, 1995.

[13] J. A. Downing, 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, "The global abundance and size distribution of lakes, ponds, and impoundments," *Limnology and Oceanography,* vol. 51, pp. 2388-2397, Sep 2006.

Jamie MacGregor is a graduate student at the School of Earth and Ocean Sciences, University of Victoria. Jamie is an outdoors enthusiast, an experienced and professional forest fire fighter, and a remote sensing expert; Dr. Kevin Telmer is a professor in the School of Earth and Ocean Sciences with expertise in the biogeochemistry of carbon and GIS and RS tools; Dr. Maycira Costa is a professor in the Department of Geography, University of Victoria specializing in applying remote sensing techniques to understand the ecology of wetlands and the ocean.