Integration of multiple spatial datasets in the development of a temporal series of high-accuracy, high-resolution land use maps

Ted Huffman[1], Don Leckie[2], Mark McGovern[3], Morten Olesen[1], Melodie Green[1], Tamara Rounce[1], James Churchill[1], Jiangui Liu[1] and Nancy Hoffmann[4]

1: Agriculture Canada, Ottawa, Ontario and Regina, Saskatchewan;
2: Natural Resources Canada, Victoria, British Columbia;
3: Environment Canada, Ottawa, Ontario
4: Statistics Canada, Ottawa, Ontario

ABSTRACT

The development of accurate international reports and domestic policies related to productivity and sustainability requires high-accuracy, high-resolution and multi-temporal resource maps. A number of land cover and vegetation maps have been produced for Canada over the past 30 years, but spatial resolution varies, classification accuracy generally remains at or below 82-86% and, since modern maps are generally derived from satellite imagery, classification focuses on land cover rather than the more economically-oriented land use.

The aim of this study was to integrate a wide variety of spatial datasets in order to develop a series of national 30m land use maps covering the period from 1990 to 2010, all at the same spatial resolution and categorized according to the 6 classes of the Intergovernmental Panel on Climate Change (IPCC); Forest, Cropland, Grassland, Wetland, Settlement and Otherland. The process involved co-registering a variety of raster and vector land cover and cadastral maps and generating output at each 30m pixel through a set of rules based on logic, map accuracies and context. The methodology incorporated the best data available in each area, and thus relied on 5 or more inputs in some areas and only 1 in others. A final step termed 'discrepancy resolution' resolved apparent errors such as settlement or water becoming forest in a subsequent year.

The project resulted in 3 maps; 1990, 2000 and 2010, all at 30m resolution. Accuracy assessment based on field survey and visual interpretation of aerial photos showed pixel-level accuracies of 84.1%, 87.0% and 93.1% for 1990, 2000 and 2010 respectively. Map accuracies of the output products surpassed accuracies of all inputs, showing the synergistic effects of combining inputs, and overall accuracies show a general improvement over the study period, reflecting the greater variety and accuracy of more recent input products. The resultant maps represent the most detailed and consistent local and national estimates of land use and land use change available for Canada.

INTRODUCTION

With the advent of place-based interventions in policy and program development, and especially with respect to international reporting of GHG sources and sinks to the United Nations Framework Convention on Climate Change, the need for high-resolution, spatially explicit and time-sensitive land use data becomes more acute. In Canada, a variety of land cover/land use products have been compiled at different times, for different extents and for different purposes. These include the current series of annual crop maps produced by Agriculture and Agri-Food Canada (AAFC, 2013), as well as a variety of products of different resolutions, classifications, areal coverages and date-stamps (e.g. Ontario Ministry of Natural Resources, 2008; Wulder et al., 2008; AAFC, 2009; Natural Resources Canada, 2014b, 2014c).

Despite the number of Land Cover/Land Use (LC/LU) map products available, the use of them in temporal studies is faced with difficulty. For example, the different spatial scales and classification schemes render data from them virtually incomparable, and even if the classification schemes can be aligned, estimates of class areas often differ widely. As a result and due to the fact that these maps are generally only 85% accurate, the determination of land use change from a comparison of maps provides unreliable results. The problem of 'error accumulation' becomes even more acute when more than 2 maps are compared.

In an attempt to develop a stable land use platform for Canada, our objective in this study was to compile as many relevant, digital maps and information sources as possible and to prepare a set of rules to be applied at each pixel to generate a ‘most probable’ output class. We selected the fairly general 6 land use classes of the Intergovernmental Panel on Climate Change (IPCC) (Forest Land, Cropland, Grassland, Wetlands, Settlements and Otherland) (IPCC 2006) with the additional separation of wetlands and water as output classes. Our target accuracy was 90%, with the goal of developing 3 output maps; c.1990, c.2000 and c.2010. As the project progressed and we visually assessed numerous cases of discrepancy between maps, it became evident that the rules also enabled us to identify likely cases of real land use change by eliminating a great deal of map error.

MATERIALS AND METHODS

A variety of Land Cover (LC) or Land Use (LU) map products and data for the portion of Canada south of 60°N were identified, including; 1) GeoCover 1990, a commercial product of 30m resolution produced from Landsat imagery (MDA, 1995-2006), 2) Earth Observation for Sustainable Development of Forests (EOSD) maps (Wulder 2008a,b), 3) CanVec forest, roads, water, built-up and wetlands vector and point layers (Natural Resources Canada 2015), 4) Land Cover for Agricultural Regions of Canada, c.2000 30m map (AAFC 2009), 5) Census of Agriculture (Statistics Canada 2014), 6) Southern Ontario Land Resource Information System (SOLRIS) Land Use Data (Ontario Ministry of Natural Resources 2008), 7) Annual Crop Inventory maps of 2010, 2011 and 2012 (AAFC Undated) and 8) Land Cover of Canada 2005, a 250m-resolution map of Canada (Natural Resources Canada 2014) used only where no other inputs existed.

The project team consisted of domain experts from Agriculture and Agri-Food Canada, Natural Resources Canada, Environment Canada and Statistics Canada. Two pilot sites (Outaouais, Quebec and Meadow Lake, Saskatchewan) were established in order to prepare an independent assessment of the accuracy of the primary input sources and to develop and test procedures. Based on the work in the pilot sites, a number of fundamental steps were established; 1) convert input classes to the desired output classes as much as possible, 2) co-register all input products, 3) develop ‘rules’ to define an output class from the specific input data at each pixel, 4) conduct contextual assessment and rectification of conflicts, and 6) implement and assess accuracy.

Input accuracy assessment

Detailed accuracy assessment of the primary input maps at the pilot sites was performed in order to establish comparable accuracies. Accuracy was assessed through ground reference data points selected randomly and interpreted from appropriate dates of aerial photography.

Input preparation

In most cases conversion of input classes to output classes involved simply renaming. For example, the 9 different forest classes of EOSD were all renamed “Forest”. Similarly, all input classes relating to swamps, marshes and bogs were renamed Wetlands; all built-up, urban and developed were renamed Settlement, and agricultural classes such as cropland, pasture, orchards and vineyards were renamed Cropland. The EOSD class “Herbs”, which could be settlement (golf courses, parks), grassland or cropland, and the class “Shrubland”, which could be forest, wetland, grassland or otherland were retained as input classes pending other source information. Vector files from CanVec (water, wetlands, National Road Network (NRN)) were rasterized at 30m resolution.

Co-registration

All map products were resampled to 30 m pixels using NAD 1983 datum, with the AAFC2000 product serving as the base layer. This allowed the generation of classes depicted on each input product at every pixel in the country. The National Road Network with a date-stamp closest to the year of concern was “burned into” the primary input source for each of the 3 maps and the class was considered as inviolate in further manipulations.

Rule development
Rules were developed to be applied at each individual pixel, with the output class at each pixel being assigned a class based on the most accurate or most likely information for that pixel. Three 'zones' in which rules varied slightly were identified; the agricultural zone, the rangeland zone, within the agricultural zone and consisting of areas where soils data indicated that natural grassland could occur, and the non-agricultural zone. Several general principles were formulated; 1) a pixel class should not be changed unless evidence or logic indicates, 2) preponderance provides more evidence than accuracy and 3) context provides more evidence than accuracy.

We started rule development using only input sources for 1990 and 2000, with the intent of developing 2 corresponding land use maps. With all inputs co-registered, a spreadsheet detailing the number of pixels with each combination of inputs was prepared and sorted in order to identify the variety and distribution of input combinations. A draft rule for each combination based on the relative accuracies of the input classes was developed and presented to the interdisciplinary expert group. Some revisions were made based on known characteristics of the input maps, especially with respect to Grassland and Shrubland, and the ruleset was applied in the pilot areas. A review of the output map and assessment against the ground-truth identified some problems, and a revised set of rules to identify the 'most likely' output class for both 1990 and 2000 for each combination was established.

Rule development required that both differences in input classes as well as land use change be taken into consideration. Since we had only one input for 1990, we used 2000 information to supplement it, with the caveat of not eliminating legitimate land use change. LU changes such as Forest to Cropland, any other class to Settlement, Grassland to Cropland and Wetland to Forest were considered legitimate and retained, while changes such as Settlement to any other class, any other class to Grassland and Otherland to Forest were considered not legitimate and eliminated. Rules were applied as an output designation for each combination of inputs in the spreadsheet.

By the time of completion of draft 1990 and 2000 LU maps, crop maps for 2011 had become available and the decision was made to add a 2010 LU map. The crop maps were co-registered to the 1990 and 2000 output maps, and a similar set of rules following the same principles was developed to generate a 2010 LU map.

Contextual assessment and rectification
Context was employed in our study to eliminate small clusters of Cropland and Grassland, as we considered that the appearance of very small fields of these classes within a larger block of a different class was unlikely and probably due to misclassification. Patches of Grassland smaller than 10 ha (91 pixels) and Cropland smaller than 5 ha (46 pixels) were converted to the dominant adjacent land use type. We also applied a 3x3 majority filter to all classes except Settlement and Wetland over the entire study area in order to 'smooth' the output and eliminate cases of single pixels of one class within a larger area of a different class.

Statistical Comparison
With the completion of the 3 draft maps, the distribution of each class was compared with Census of Agriculture (Statistics Canada 2015) data at the Census Division level. In most cases, areas mapped as cropland and grassland showed slightly higher estimates than the area reported in the census, with a few exceptions. In areas along the Atlantic coast, the census indicated perennial crops in areas where the maps showed none. In those few areas we used imagery and aerial photography to visually identify and digitize open areas around residences as cropland. In other areas such as the mixed grassland/forest areas in Alberta and British Columbia, we found considerably less area mapped as grassland than reported in the Census as unimproved pasture. We expanded the Grassland category into Forest by 2 pixels on all boundaries between the two, but were unable to align the 2 estimates. We assume that the discrepancy relates to ranchers reporting pastured forest as unimproved pasture in the Census.

Accuracy Assessment
Map accuracy was assessed using randomly selected ground-truth points of the appropriate year compiled from field survey, aerial photography and high-resolution satellite imagery. In the case of points which fell within a mixed pixel, we gave them 2 correct classes, rather than discarding them as 'undefined'. For example, a point which fell within a pixel that was a mix of water and forest was considered correct for either, but incorrect for any other class. We feel that this approach avoids creating a ground-truth bias toward more easily classified 'pure' pixels, and thus provides a more realistic interpretation of map accuracy. Accuracy assessment of the LU1990 map was based on 7139 points, LU2000 on 7218 points and LU2010 on 4064 points.

RESULTS

The entire study area encompassed approximately 550 million hectares and 6 billion pixels. Input preparation required renaming a variety of classes in different input maps to IPCC categories as shown in Table 1. Results of the accuracy assessment of the 3 primary input maps are presented in Figure 1.

Table 1. Examples of different input classes and their renamed class.

Examples of input classes	IPCC class
Forest - Deciduous; Forest – Coniferous; Broadleaf Dense Forest; Dense Deciduous Forest; Coniferous Dense Forest; Mixedwood Dense Forest; Broadleaf Open Forest; Coniferous Open Forest; Mixedwood Open Forest; Broadleaf Sparse Forest; Coniferous Sparse Forest; Mixedwood Sparse Forest; Coniferous Plantation	Forest
Agriculture – General; Pasture and Abandoned Fields; Annual Cropland; Perennial Cropland; Cultivated Agricultural Land; Orchards; Vineyards	Cropland
Rangeland; Native Pasture; Native Grass; Meadow; Alvar; Tallgrass Savannah	Grassland
Wetland - Permanent Herbaceous; Wetland-Herb; Wetland-Shrub; Wetland-Treed; Inland Marsh; Conifer Swamp; Deciduous Swamp; Open Fen; Bog; Peatland	Wetland
Urban/Built-up; Developed; Industrial; Residential; Roads; Railroads; Commercial; Buildings; Transportation; Recreational; Institutional; Airport; Extraction; Pits and Quarries;	Settlement
Reservoir; River; Lake	Water
Exposed/Barren Land; Rock/Rubble; Beach; Permanent Ice; Bare Rock; Open Sand Barren and Dune; Non-Vegetated Land; Sparsely Vegetated Bedrock; Open Cliff and Talus; Bryoids;	Otherland

Ground Reference	GEOCOVER (Overall accuracy = 74.8%)									
	FOREST	CROPLAND	GRASSLAND	SETTLEMENT	WATER	WETLAND	OTHERLAND	SHRUBLAND	TOTAL	% Correct
FOREST	**572**	32	3	1	0	3	0	75	686	**83.4**
CROPLAND	42	**358**	14	0	0	3	2	59	478	**74.9**
SETTLEMENT	13	13	2	**87**	0	2	1	3	121	**71.9**
WATER	4	1	0	0	**84**	1	0	2	92	**91.3**
WETLAND	52	14	4	0	2	**17**	0	24	113	**15.0**
OTHER LAND	0	3	0	0	0	1	**0**	0	4	**0.0**
TOTAL	683	421	23	88	86	27	3	163	1494	**74.8**

Ground Reference	EOSD (Overall accuracy = 82.2%)									
	FOREST	CROPLAND	GRASSLAND	SETTLEMENT	WATER	WETLAND	OTHERLAND	SHRUBLAND	TOTAL	% Correct
FOREST	**649**	13	0	5	3	2	0	14	686	**94.6**
CROPLAND	16	**380**	0	54	0	2	0	26	478	**79.5**
SETTLEMENT	19	28	0	**61**	1	4	**0**	8	121	**50.4**
WATER	7	0	0	0	**85**	0	0	0	92	**92.4**
WETLAND	28	13	0	8	6	**52**	0	6	113	**46.0**
OTHER LAND	0	1	0	0	0	0	**2**	1	4	**50.0**
TOTAL	736	435	0	128	95	43		55	1494	**82.2**

Ground Reference	AAFC LC2000 (Overall accuracy = 81.7%)									
	FOREST	CROPLAND	GRASSLAND	SETTLEMENT	WATER	WETLAND	OTHERLAND	SHRUBLAND	TOTAL	% Correct
FOREST	**582**	4	5	7	23	14	22	29	686	**84.8**
CROPLAND	3	**439**	22	2	4	0	0	8	478	**91.8**
SETTLEMENT	9	0	0	**97**	3	1	4	7	121	**80.2**
WATER	2	0	0	0	**90**	0	0	0	92	**97.8**
WETLAND	15	0	0	0	16	**55**	0	27	113	**48.7**
OTHER LAND	0	0	0	1	0	0	**2**	1	4	**50.0**
TOTAL	611	443	27	107	136	70	28	72	1494	**81.7**

Figure 1. Error matrices showing the accuracy of each class at the Outaouais study site for Geocover, EOSD and AAFC LC2000

An example of a spreadsheet with pixel count, input classes and output classes as defined by the rules is presented in Figure 2.

COUNT (pixels)	INPUTS				OUTPUTS	
	GEOCOVER (1990)	AAFC (2000)	EOSD (2000)	CANVEC (c.2000)	LU90	LU00
89,320,256	Cropland	Cropland	Unclassified	Unclassified	Cropland	Cropland
73,097,451	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified
72,871,204	Shrubland	Unclassified	Forest	Unclassified	Forest	Forest
44,420,304	Cropland	Cropland	Herbs	Unclassified	Cropland	Cropland
43,413,067	Forest	Unclassified	Forest	Unclassified	Forest	Forest
34,925,671	Water	Unclassified	Water	Water	Water	Water
15,696,365	Shrubland	Unclassified	Shrubland	Unclassified	Shrubland	Shrubland
12,269,881	Water	Water	Water	Water	Water	Water
12,215,505	Shrubland	Forest	Forest	Unclassified	Forest	Forest
8,651,350	Grassland	Grassland	Unclassified	Unclassified	Grassland	Grassland
8,264,079	Forest	Unclassified	Shrubland	Unclassified	Forest	Forest
8,088,053	Shrubland	Cropland	Unclassified	Unclassified	Cropland	Cropland
6,753,562	Grassland	Cropland	Unclassified	Unclassified	Cropland	Cropland
5,981,113	Shrubland	Unclassified	Wetland	Unclassified	Wetland	Wetland
...	...	...	...	...	...	...

Figure 2. An example of a spreadsheet showing pixel counts, combinations of 4 inputs and 1990 and 2000 outputs as defined by the rules.

The study resulted in 3 maps, one for each of 1990, 2000 and 2010, covering all areas of Canada south of 60°N. An example of an output map (2010) is presented in Figure 3. The overall accuracy of all 3 output maps surpassed the accuracies of the input maps as determined at the pilot sites, with the 2010 output map showing an overall accuracy of 92.7% (Figure 4).

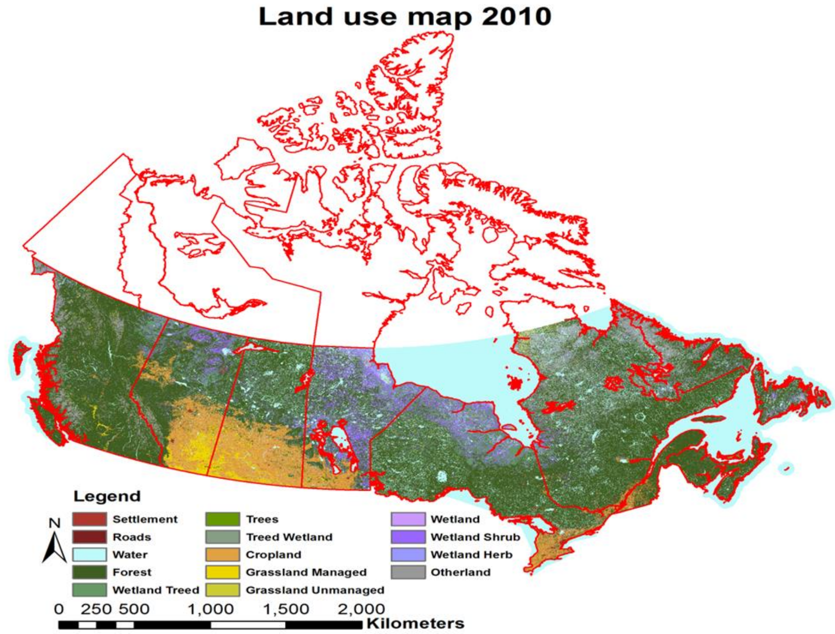

Figure 3. An example of a land use output map (2010).

1990
Overall Accuracy: 84.0%

		Ground Truth								
		Forest	Water	Cropland	Settlement	Wetland	Otherland	Grassland	Total	Users Accuracy
Map Class	Forest	2781	25	103	4	172	87	78	3250	85.6%
	Water	32	1372	3	0	74	8	7	1496	91.7%
	Cropland	56	6	524	18	40	5	26	675	77.6%
	Settlement	23	2	37	190	9	8	1	270	70.4%
	Wetland	113	9	26	5	859	30	29	1071	80.2%
	Otherland	11	1	5	0	17	91	2	127	71.7%
	Grassland	12	3	35	0	9	13	177	249	71.1%
	Total	3028	1418	733	217	1180	242	320	5994	7138
Producers Accuracy		91.8%	96.8%	71.5%	87.6%	72.8%	37.6%	55.3%	7138	

Total Observations: 7138 Total Correct: 5994 KHAT Accuracy: 78.0%

2000
Overall Accuracy: 87.1%

		Ground Truth								
		Forest	Water	Cropland	Settlement	Wetland	Otherland	Grassland	Total	Users Accuracy
Map Class	Forest	2870	18	84	3	109	60	61	3205	89.5%
	Water	28	1181	27	0	55	6	5	1302	90.7%
	Cropland	48	6	1004	15	35	3	27	1138	88.2%
	Settlement	19	3	25	217	1	4	0	269	80.7%
	Wetland	85	8	36	5	776	27	23	960	80.8%
	Otherland	8	1	0	0	17	75	2	103	72.8%
	Grassland	13	2	45	0	7	11	164	242	67.8%
	Total	3071	1219	1221	240	1000	186	282	6287	7219
Producers Accuracy		93.5%	96.9%	82.2%	90.4%	77.6%	40.3%	58.2%	7219	

Total Observations: 7219 Total Correct: 6287 KHAT Accuracy: 82.0%

2010
Overall Accuracy: 92.7%

		Ground Truth								
		Forest	Water	Cropland	Settlement	Wetland	Otherland	Grassland	Total	Users Accuracy
Map Class	Forest	1719	8	12	3	43	41	16	1842	93.3%
	Water	12	839	3	0	14	3	0	871	96.3%
	Cropland	20	1	429	0	2	0	18	470	91.3%
	Settlement	5	0	5	50	0	0	0	60	83.3%
	Wetland	24	3	3	2	464	3	1	500	92.8%
	Otherland	7	1	1	2	19	182	2	214	85.0%
	Grassland	8	0	7	0	5	4	82	106	77.4%
	Total	1795	852	460	57	547	233	119	3765	4063
Producers Accuracy		95.8%	98.5%	93.3%	87.7%	84.8%	78.1%	68.9%	4063	

Total Observations: 4063 Total Correct: 3765 KHAT Accuracy: 90.0%

Figure 4. Error matrices of the 3 land use output maps

Conclusions

The study has shown that integration of multiple raster input layers to produce high-resolution and high-accuracy land use maps can be accomplished using rulesets, and it further shows that the accuracy of the output maps can surpass the accuracy of any of the inputs. Accuracy increases as more input information is integrated.

The authors feel that the incorporation of other inputs such as slope and soil maps could improve overall accuracy considerably, especially with respect to cropland, water and wetland. It also seems that even unclassified Earth Observation data could be incorporated to improve accuracy.

The experience of this study indicates that in order to improve the accuracy of spatial data, new Earth Observation mapping efforts should ‘build’ on existing maps rather than focus exclusively on new sensors and algorithms. By using existing information as a base, it appears that overall map accuracy could readily reach 95% or more.

References

AAFC. 2009. Land Cover for Agricultural Regions of Canada, circa 2000. http://data.gc.ca/data/en/dataset/16d2f828-96bb-468d-9b7d-1307c81e17b8 (Accessed 11 Dec 2014).

AAFC. Undated. Annual Crop Inventory. http://open.canada.ca/data/en/dataset/ba2645d5-4458-414d-b196-6303ac06c1c9. (Accessed 20 May 2015).

IPCC. 2006. Guidelines for National Greenhouse Gas Inventories, Volume 4: Agriculture, Forestry and Other Land Use; Chapter 3: Consistent Representation of Lands. http://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html. (Accessed 20 May 2015).

MDA. 1995-2006. EarthSat GeoCover LC Overview. http://www.mdafederal.com/geocover/geocoverlc/gclcoverview/. (Accessed 20 May 2015).

Natural Resources Canada. 2014a. Canada Land Inventory (1:250,000) - 1966 Land Use. http://geogratis.gc.ca/api/en/nrcan-rncan/ess-sst/1cd7d940-309b-5965-b181-a5221dd898f3.html (Accessed 11 Dec 2014).

Natural Resources Canada. 2014b. National Earth Observation Data Framework Catalogue; SPOT/HRV landcover of Canada from 2005-2010

Natural Resources Canada. 2014c. National Earth Observation Data Framework Catalogue; MODIS annual landcover time series of Canada 2000-2011

Natural Resources Canada. 2015. Geogratis. https://www.nrcan.gc.ca/earth-sciences/geography/topographic-information/free-data-geogratis/11042

Natural Resources Canada. Undated. Canadian National Fire Database. http://cwfis.cfs.nrcan.gc.ca/ha/nfdb

Ontario Ministry of Natural Resources. 2008. Southern Ontario Land Resource Information System (SOLRIS) Land Use Data. Toronto, Ontario. https://www.javacoeapp.lrc.gov.on.ca/geonetwork/srv/en/main.home# (Accessed 11 Dec 2014).

Statistics Canada. 2014. Census of Agriculture; land use. http://www5.statcan.gc.ca/cansim/a26?lang=eng&retrLang=eng&id=0040203&pattern=0040200..0040242&tabMode=dataTable&srchLan=-1&p1=1&p2=50#customizeTab. (Accessed 20 May 2015).

Statistics Canada. 2015. 2011 Census of Agriculture. http://www.statcan.gc.ca/eng/ca2011/index?fpv=920 (Accessed 2 June 2015).

Wulder, M.A., White, J.C., Cranny, M., Hall, R.J., Luther, J.E., Beaudoin, A., Goodenough, D.G. and Dechka, J.A. 2008a. Monitoring Canada's forests. Part 1: Completion of the EOSD land cover project. Canadian Journal of Remote Sensing, 34(6): 549-562, 10.5589/m08-066

Wulder, M.A.; Cranny, M.M.; Hall, R.J.; Luther, J.E.; Beaudoin, A.; White, J.C.; Goodenough, D.G.; Dechka, J.A. 2008b. Satellite land cover mapping of Canada's forests: the EOSD land cover project. Pp. 21-30 (Chapter 3) in J.C. Campbell, K.B. Jones, J.H. Smith, and M.T. Koeppe (eds). North America Land Cover Summit. American Association of Geographers, Washington, DC, USA. https://cfs.nrcan.gc.ca/publications?id=29220