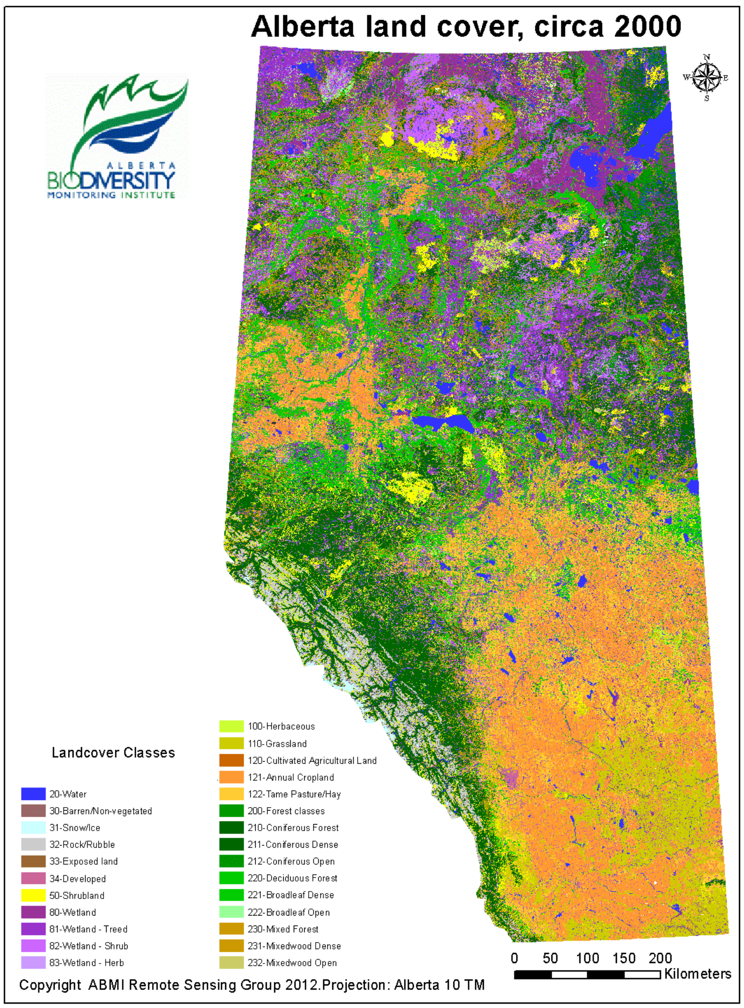
Accuracy Assessment of the Alberta wall to wall landcover polygon vector layer circa 2000, beta version (ABw2wLCV2000beta)



June 2012

UNIVERSITY OF CALGARY

ALBERTA BIODIVERSITY MONITORING INSTITUTE

About this document

Title	Accuracy Assessment of the Alberta wall to wall landcover polygon vector layer circa 2000, beta version (ABw2wLCV2000beta)
Purpose	To provide background, methods and results for the accuracy assessment of the of the ABw2wLCV2000beta, and to suggest legend and production changes for the final version of this landcover map
Filename	ABw2wLCV2000beta_AA.pdf
Contributors	Guillermo Castilla (design of the accuracy assessment procedures, implementation of sampling and analysis tools, writing of this document); Typhenn Brichieri-Colombi,,Vanessa Cuervo, Riley Iwamoto, Delliana Rasheva (photo-interpretation); Ana Hernando (assistance in developing the methods); Chunhua Zhang (streamlining of validation procedures in ArcGIS); Deep Mazumdar (data curation, QC control, lab management); Greg McDermid (feedback and facilitation).
Document history	V20120608, first draft V20120630, final draft incorporating suggestions from the ABMI science centre and other stakeholders
Current version	V20120630
Changes to previous version.	Minor edits
Date	2012/06/30
Status	Final draft
Target readership	Users of the ABMIw2wLCV2000
General readership	Public
Correct reference	ABMI Remote Sensing Group (2012) Accuracy Assessment of the Alberta wall to wall landcover polygon vector layer circa 2000, beta version (ABw2wLCV2000beta). 1st Draft.

UNIVERSITY OF CALGARY

ALBERTA BIODIVERSITY MONITORING INSTITUTE

Table of Contents

EXECUTIVE SUMMARY

This document describes the background, materials, methods, and results for the accuracy assessment of the Alberta wall to wall landcover (LC) polygon vector layer circa 2000, beta version (ABMIw2wLCV2000beta). The latter is a map describing the spatial distribution of 18 LC classes (at the finest hierarchical level, which become 14 and 6 classes at the two upper levels) across the province of Alberta at a cartographic scale of 1:125,000 and for the reference year 2000. The map itself consists of a mosaic of over two million polygons of various sizes, from half a hectare to thousands of hectares. The ABMIw2wLCV2000beta was derived by applying a spatial generalization algorithm to a combination of two raster datasets: the Canadian Forest Service (CFS) Earth Observation for Sustainable Development (EOSD) LC dataset, and the Land Cover for Agricultural Regions of Canada, circa 2000 (LCARC) dataset of Agriculture and Agri-Food Canada (AAFC). Both raster datasets were created using digital classification of Landsat 5 and Landsat 7, 30 m ortho-images acquired around year 2000. The accuracy assessment consisted in validating (i.e., determining the LC classes within), and eventually correcting the boundary of a representative subset of over 6,000 polygons randomly selected from this map. Validation was carried out by visual inspection of each selected polygon on 1 m resolution aerial ortho-photos also acquired around year 2000. The overall accuracy of the map was estimated to be 61% with 18 classes, 70% with 14 classes, and 81% with 6 classes. The classes that are more often confused are conifer vs. treed wetland, and annual crops vs. pasture. The likelihood that a given polygon has the correct finest-level label was estimated at 47%. Larger polygons (>300 ha), which cover around half of the landbase, have a greater likelihood of being correctly labeled (72%), but they are just a 2% of the total number of polygons. Only 35% of the polygons actually represent full patches, while the rest are, in reality, parts of a much larger patch that was misclassified in the area occupied by the polygon, meaning that there is a considerable number of outlines (>50%) that have the same landcover on both sides. A set of changes, to both the legend and the spatial generalization methods, are suggested to create a final version of the map that can be used as a source of reliable information on landcover.

1 INTRODUCTION

The Alberta wall to wall landcover (LC) polygon vector layer circa 2000, version beta, (ABMIw2wLCV2000beta) is a map describing the spatial distribution of LC across the province of Alberta for the reference year 2000. The map legend consists of 18 LC classes at the finest level (L3), 14 classes at L2, and 6 classes at L1 (Appendix 1). The map itself consists of a mosaic of over 2 million non-overlapping polygons of various sizes, from less than 1 hectare (ha) to thousands of ha. Each polygon represents a contiguous area relatively homogeneous in terms of LC, where the specific LC class of the polygon is different from that of adjacent polygons. The minimum mapping unit (MMU, or minimum polygon size) is 0.5 ha for aquatic features, 1 ha for wetland features, and 2 ha for the rest. The cartographic scale of reference (i.e., the scale at which the map would be printed if distributed in hardcopy) is 1:125,000. The target positional accuracy of polygon outlines is 0.5 mm at that scale, or 60 m on the ground (i.e., the true boundary of the polygon must lie within 60 m of the outline 95% of the times). The file format chosen for the ABMIw2wLCV2000beta is ESRI file geodatabase. The latter contains two feature classes, one consisting of 48 tile polygons, into which the province was divided for cartographic production purposes, and another containing the actual landcover polygons for the entire province.

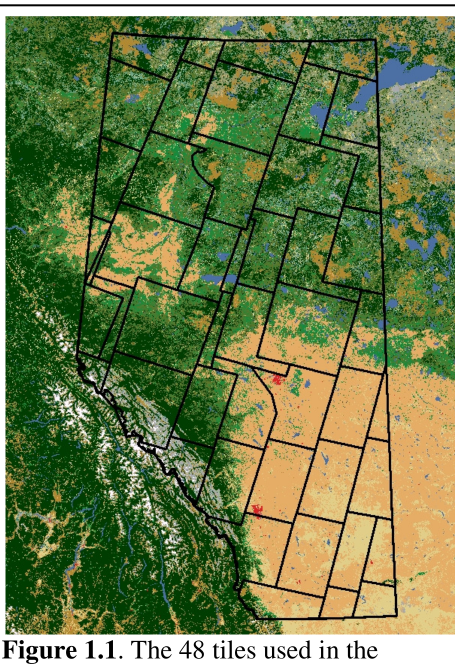


Figure 1.1. The 48 tiles used in the production of the ABw2wLCV2000beta

The seamless feature class containing the landcover polygons was obtained by merging the 48 non-overlapping tiles into which the province was divided during map production (Figure 1.1). The tiles range from 1,000 to 29,000 km^2 in size, and each is fully encompassed within a single Landsat scene. The individual landcover vector layers that were generated for each of the tiles are in the NAD83 UTM projection, zone 11 or 12, depending on location. The seamless, overall feature class is provided in the Alberta 10 Degree Transverse Mercator (10TM) projection. The ABw2wLCV2000beta was derived by combining and applying a spatial generalization algorithm to two raster datasets: the Canadian Forest Service (CFS) Earth Observation for Sustainable Development (EOSD) LC dataset, and the Land Cover for Agricultural Regions of Canada, circa 2000 (LCARC) dataset of Agriculture and Agri-Food Canada (AAFC). Both raster datasets were created using digital classification of Landsat 5 and Landsat 7 ortho-images acquired around year 2000. In both rasters, each Landsat scene was classified individually and then mosaicked into the final raster product.

The ABMIw2wLCV2000beta tiles coincide with the Landsat scenes employed in these two raster datasets, and their frames roughly follow the seam lines used to stitch together individually-classified scenes into the aforementioned raster mosaics. Within each tile, the EOSD information was given priority in forested areas, and LCARC in agricultural areas. In addition, hydrography features from the Government of Alberta (GoA) GIS layers were used for 'burning in' wetlands, water bodies, and major rivers into the LC rasters. The same process was also applied to roads, railways, powerlines, and pipelines using GoA's access layer.

As any other GIS product, the accuracy of the ABMIw2wLCV2000beta needs to be assessed so that information on its quality is available to end users. In particular, this accuracy assessment aims to answering the following questions: **(1)** What is the proportion of area assigned to each LC class that is actually covered by that class? **(2)** What is the degree of confusion among the different LC classes? **(3)** What is the likelihood that a given polygon has the correct LC label? **(4)** What is the proportion of polygons that actually represent full patches? **(5)** How accurate are the polygon boundaries? And **(6)** what can be improved in the final version? The rest of the document is organized as follows: the next two sections describe the materials (section 2) and the methods (section 3) used for the accuracy assessment. Section 4 discusses the results, and section 5 puts forward some proposed changes for the final version of the map.

2 MATERIALS

2.1 Image data

The validation polygons (i.e., those included in the sample) are inspected and corrected using as a backdrop a color aerial ortho-photo mosaic of of 1 m spatial resolution acquired around year 2000 (hereafter the Valtus2000 mosaic). Being coetaneous with the Landsat imagery used in the ABMIw2wLCV2000beta, this dataset is ideal for validation purposes. The dataset, procured from Valtus Imagery Services Limited, was created using ortho-rectified colour aerial photographs produced by North West Geomatics. Table 2.1 contains some metadata about this image mosaic. The Valtus2000 is made available to the interpreters via an ArcSDE connection to a local server in our lab.

Table 2.1 : Valtus2000 image metadata for NTS sheet 72E

Parameter	Details
Image	Ortho photo
Data format	Tiff 6.0
Data type	Ortho rectified colour aerial photography
Year of photography	2000
Scale of photography	1:50,000
Geographic region	72E
Horizontal datum	NAD 83
Projection	10TM with 500,000 meter false easting
Image resolution	1.0 m
Scanner type	LH DSW 500
Scan resolution	14.0 microns
Software used for ortho	LH Socet set
Control-aerial triangulation source	Alberta Government 1:60,000 photo
Control-aerial triangulation accuracy	± 5.0 m x, y ± 3.0 m z
Control-GPS source	Differential Kinematics GPS
Control-GPS accuracy	± 0.50 m
Relative position accuracy	± 10 m
DEM data source	AltaLis DEM spacing 100 m
DEM vertical accuracy	± 5.0 m x, y ± 3.0 m z

2.2 Vector data

Each tile dataset consists of a polygon feature class (stored in an ESRI file geodatabase of generic name T4X0XX_AA.gdb), plus two shapefiles providing a guiding sequence and context for the corrections. Each tile has a 5-digit numeric identifier, where the first two digits correspond to the path and the last three to the row of the corresponding Landsat scene (e.g., tile T42025 corresponds to path 45 row 25 of the Landsat 4-5-7 World Reference System [WRS2]). The validation dataset for each tile consists of the following GIS layers (either as a shapefile or geodatabase feature classes):

[1] **T4X0XX_unlabeled.shp**

A polygon shapefile with the ABMIw2wLCV2000beta for tile 4X0XX with the landcover labels removed, which is meant to provide the spatial context of the selected polygons.

[2] **T4X0XX_spoly.shp**

A polygon shapefile that contains the randomly selected non-adjacent polygons < 300 ha extracted from [1], where their total area exceeds a 1% of the area of the tile covered by polygons < 300 ha.

[3] **T4X0XX_spoly_edited**

A polygon feature class of T4X0XX_AA.gdb that is initially a copy of [2], which will later include (i) the break-down into incomplete homogeneous *parts* > MMU for heterogeneous validation polygons; and (ii) a geometrically corrected version of those validation polygons (or parts within them) that could be modified in a way that they become self-contained homogeneous areas that can be seen as *wholes*, i.e., as independent patches (see the response design section, 3.2).

[4] **T4X0XX_spoly_xl**

A polygon feature class containing the selected polygons for the oversize stratum of this tile. These are validated in a separate session.

[5] **T4X0XX_spoly_xl_edited**

A polygon feature class of T4X0XX_AA.gdb that is initially a copy of [4], which will later include the break-down into homogeneous parts > MMU for heterogeneous validation oversize polygons.

The editable fields in the _edited feature class are the following:

a. **LC_orig**: The LC label from the ABw2wLCV2000beta. This field is empty in the version available to the interpreters, to avoid classification bias.
b. **LC_corr**: Corrected LC class, as interpreted visually from the VHR image.
c. **LC_corr2**: a second plausible LC class to which the modified polygon could also belong. This field can **only** be filled in cases of ambiguity about what the correct LC class should be, either because the setting is a borderline case between those two classes (e.g., conifer dense vs. conifer open), or because there are insufficient clues in the imagery (e.g, this area here is likely a grassland, but it could also be a pasture).
d. **TYPE**: a short integer field with two valid values, 0 ('Part' polygon, default value), and 1 ('Whole' polygon). See section 3.2 for an explanation of 'part' and 'whole' polygons. NB. This field only exists for the regular stratum.

3 METHODS

3.1 Sampling design

The goal of the sampling design is to select a subset of polygons in such way that the selected sample can be deemed representative of the entire population of polygons of the map, so as to ensure that the conclusions derived from the sample can be validly extrapolated to the entire map. Initially we only had high resolution imagery available from the Valtus 2000 mosaic (which covers about 2/3 of the province; hashed areas in Figure 3.1), so we had to constrain our choice of tiles to those covered by the Valtus 2000. Since areas outside the Valtus had a zero probability of inclusion, we cannot legitimately claim that the assessment is valid for the entire map, unless we assume that the areas not covered by Valtus are similar than the rest in terms of their expected accuracy. Given that all tiles were produced using the same type of input data and methods, this seems a reasonable assumption. A caveat, though, is that there are some classes in the validation dataset, such as rock/ruble and snow/ice (common in the Rockies, which lie out of the Valtus), which are underrepresented or absent in the validation dataset, and therefore, we could not assess them. But since these classes are radiometrically distinct in Landsat imagery,

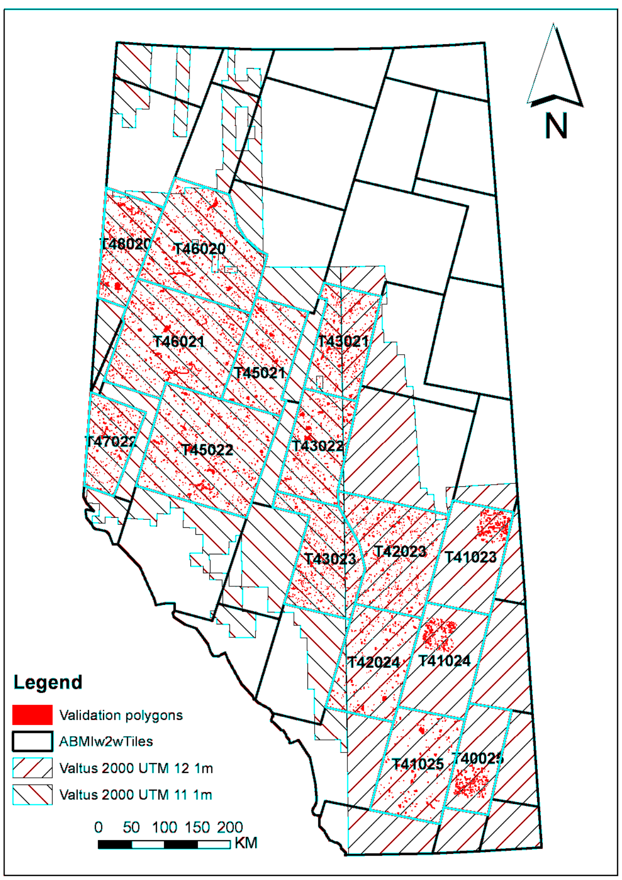


Figure 3.1. The 15 tiles (highlighted in cyan) used in the accuracy assessment of the ABw2wLCV2000beta

the accuracy in the mountains can be expected to be a little higher than what we report for the rest of the map. Another caveat is that the White Area is overrepresented in our validation dataset, since the proportion of agricultural areas in the chosen tiles is higher than the provincial average.

With the constraint imposed by the Valtus dataset, we selected the 15 largest tiles (highlighted in cyan in Figure 3.1), totaling 257,000 sq km, out of the set of 25 that have an almost complete coverage in the Valtus2000 imagery. Using the map tiles as primary sampling units makes sense, since each was independently classified in the original raster landcover maps that we used as input in the creation of the ABMIw2wLCV2000beta. Polygons within each of these 15 tiles are assigned to one of two different strata, which are sampled independently and validated using different strategies. The first stratum (hereafter the **regular stratum**) consists of all polygons smaller than 300 ha, which are subject to a modified[1] *stratified random sampling,* with one (sub) stratum per existing LC class. This sampling scheme consists in randomly selecting a number of polygons of each present LC class, adding up a minimum of 1% of the area covered by that class in the tile, with the constraint that the selected polygons cannot be adjacent. The 300 ha threshold was chosen for practical purposes (it is difficult and time consuming to implement the methods of the regular stratum for very large polygons), and also because it roughly leads to a 50/50 split of the Alberta landbase into the two strata. The second stratum (hereafter the **oversize stratum**) consists of all polygons exceeding 300 ha in the tile, which are subject to conventional (i.e., without taking into account adjacency) stratified simple random sampling, with the same 1% sampling intensity by area, this time with sub-strata defined according L2 labels. The reason for using 14 classes instead of 18 for these sub-strata is that this promotes more polygons per stratum, which at a 1% sampling intensity is necessary, since there are only a few hundred oversize polygons in a tile. By using sub-strata based on LC we wanted to obtain a balanced

[1] Modified by a non-adjacency constraint. That is, no two polygons in the sample can be adjacent. We have studied the impact of this constraint on the inclusion probabilities, and ascertained that at the sampling intensity we applied, it can be assumed that this scheme still yields an equal probability of selection (Castilla et al., submitted to IJGIS).

representation of the different LC classes present in the map. We note there are four tiles that contain exceptions to this general scheme. Tile 41025 (in the grassland natural region) was only assessed for the oversize stratum, as the interpreter in charge of it did not manage to get to the regular stratum within the available time. In tiles 41023, 41024 and 40025 (respectively, Central Parkland, Northern Fescue, and Dry Mixedgrass natural subregions), the polygons were extracted from randomly placed 80 km squares, rather than from across the full extent of the tiles.. The reason for this is that these polygons were inspected on the field, and were selected in this way so as to manage travel costs and time. In total, we validated over 6000 polygons covering over 3000 sq km, which respectively correspond to 0.3% of the polygons in the map and 0.5% of its area (Table 3.1).

Table 3.1. Summary stats about sampling intensity per class and for the entire map

LC_class	LC code	N_pols	n_spols	si%(pols)	KM2	sarea (ha)	si%(area)
WATER	20	84,821	264	0.31	28,372	19,622	0.69
SNOW	31	2,081	0	0.00	1,219	-	0.00
ROCK	32	9,729	5	0.05	12,113	836	0.07
BARREN	33	18,096	53	0.29	2,677	263	0.10
DEVELP	34	238	5	2.10	24,859	367	0.01
SHRUB	50	111,135	374	0.34	20,287	10,811	0.53
WET_T	81	251,380	819	0.33	54,662	22,666	0.41
WET_S	82	197,902	536	0.27	36,797	9,534	0.26
WET_H	83	139,426	403	0.29	34,940	6,175	0.18
GRASS	110	160,648	651	0.41	56,377	26,511	0.47
CROP	121	79,935	578	0.72	82,369	70,625	0.86
P/HAY	122	106,559	588	0.55	49,982	39,891	0.80
FCD	211	257,439	421	0.16	114,137	39,565	0.35
FCO	212	47,684	37	0.08	9,208	760	0.08
FBD	221	258,882	896	0.35	78,229	43,328	0.55
FBO	222	23,771	43	0.18	2,525	759	0.30
FMD	231	288,705	641	0.22	52,921	20,550	0.39
FMO	232	11,602	3	0.03	1,349	44	0.03
	MAP TOTAL:	**2,050,033**	**6,317**	**0.31**	**663,024**	**312,307**	**0.47**

3.2 Response design

This process consists in creating validation data for each polygon in the sample. Before describing it, it is worth explaining the logical progression we followed in order to arrive at the final method. Let us start imagining a crude accuracy assessment based on polygons, wherein we would inspect each polygon of the sample and then enter in a table our choice for the majority landcover class in the polygon. This method would be fast and simple, but it is incomplete, since it cannot capture other classes that may be present in the polygon and therefore does not provide a complete assessment of the confusion between classes. A slightly more refined procedure would be entering, in a spreadsheet with as many columns as landcover classes, a rough estimate of the percent area occupied by each landcover class appearing in the polygon. While this would be relatively cost-efficient, eyeballing the areal distribution of landcover classes within a polygon is inherently unreliable. Much better estimates could be obtained by actually digitizing the different parts into which the polygon could be split, assigning a landcover class to each part, and computing the percent area using semi-automated GIS operations. This latter method would enable the construction of an accurate confusion matrix, but it does not allow for deriving information on structural or positional accuracy. In particular, we are interested in knowing (i) which of the parts that the interpreter identified within the polygon could become a self-contained *whole* (i.e., a patch of its own) after some minor edits, and which of them are just a small portion of a considerably larger landcover patch; and (ii) how close from the observed edge in the image are those portions of polygon outline that correspond to true landcover transitions. The final method we devised (Figure 3.2) enables us to collect this extra information with a small additional effort.

In addition to splitting a polygon into homogenous parts when the polygon is heterogeneous, we ask the interpreters to go beyond the confines of the actual validation polygon and complete those parts that could become self-contained wholes if merged to a neighboring area, providing that area is not much larger than the part (NB. We use a 1/3 cut-off: i.e., the resulting patch must be less than three times the current size of the incomplete part, otherwise the 'part' cannot be made a 'whole'). Finally, we also ask the interpreters to modify the outline of those polygons deemed 'wholes' wherever it strays from the actual boundary seen in the image. In order to avoid an inordinate amount of extra work, we ask for this only along sections that correspond to genuine landcover transitions, and only if the outline is more than 45 m apart from the edge in the image (if it is less than that, we deem the outline good enough, since this is well below the target positional accuracy of 60 m).

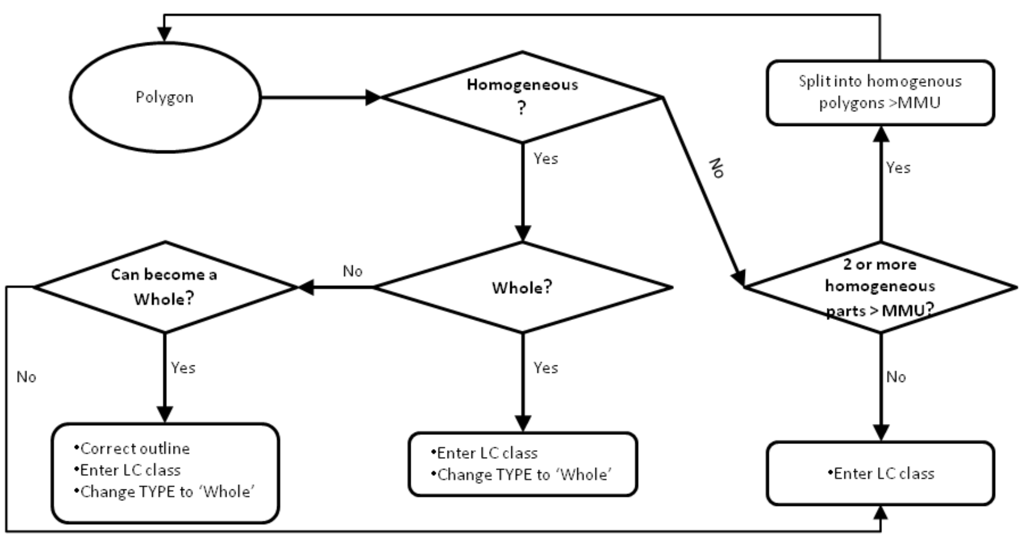


Figure 3.2. Decision flowchart for correcting a validation polygon

In order to carry out the validation, we hired four interpreters (a fresh PhD, and 3 graduate students enrolled in either MSc or MGIS programs, all with some previous experience in photo-interpretation). We trained them using some pre-defined examples and provided them with a 50-page interpreter's manual that we created, with detailed instructions and a photo-key. After the training, we evaluated the repeatability of the method (i.e., whether different interpreters will arrive to similar corrections using this method) using an *ad hoc* sample of 50 polygons that was independently corrected by each of them. The coefficient of variation of the set of accuracy parameters derived by the four interpreters was less than 15%, which we deemed reasonable given that there is an unavoidable subjective component to any assessment. To ensure enhanced consistency, a QC process was also put in place wherein 10% of the validated polygons in each tile were inspected by our RS/GIS technician. If, after inspection, some significant bias was detected, the full set of validation polygons for the tile was returned to the interpreter for revision, after which the QC process was repeated. All digitization and attribution are performed using the inbuilt editing tools of ESRI's ArcGIS. Of particular usefulness is the *Data Driven Pages* tool, which enables the interpreter to navigate from one polygon to the next by just clicking a button. A set of coded names is used for the LC attributes so that the LC class is entered from a drop down menu to avoid typos. The average correction time for a validation polygon of the regular stratum is 3 minutes, and 6 minutes for an oversize polygon. We note that in the case of the three tiles whose polygons were inspected on the ground, the correct LC class comes from the field data sheets, and the required splits are based on the imagery, but also on sketches from the field.

3.3 Analysis

We created some automated scripts to analyze the validation dataset and derive the accuracy parameters. The area-based confusion matrices were derived from the attribute table of the intersection between the ABMIw2wLCV2000beta and the _edited layers. The polygon-wise accuracy is derived likewise, but from the intersection between the initial _spoly layers and the _edited layers, and is equal to the percent of polygons in the sample that have more than 50% of their area covered by the LC class indicated in the map. The proportion of polygons that represent full patches is equal to the number of validation polygons from the regular stratum that contain in the final _edited layer a 'whole' polygon where the area of overlap between the two is at least half of the original polygon. This is derived from the intersection between the initial _spoly layers and the subset of polygons in the _edited layers deemed 'whole' patches. Finally, we computed the amount of map linework that is within 45 m of the outlines of the edited 'whole' polygons, and also visually assessed the overall accuracy of boundaries.

4 RESULTS AND DISCUSSION

4.1 Area-based accuracy parameters

The map's overall accuracy (i.e., the percent area of the map that is actually covered by the LC class indicated in the map), as estimated using the entire set of validation polygons, is 61% (Table 4.1.1). This figure seems low compared to the 85% standard recommended in remote sensing textbooks. However, this refers to the set of 18 classes existing at the finest level of the map (L3). If we aggregate the results into the 6 classes of the top hierarchical level (L1), this figure reaches 81% (Table 4.1.3). The best user's accuracies (UA, i.e., proportion of area assigned to a given LC class that is actually covered by that class) for individual classes are 94% for water, 76% for dense broadleaf forest (FBD), 69%

								REFERENCE											
MAP	WATER	SNOW	ROCK	BARREN	DEVELP	SHRUB	WET_T	WET_S	WET_H	GRASS	CROP	P/HAY	FCD	FCO	FBD	FBO	FMD	FMO	TOTAL	UA%
WATER	18,419	-	-	5	0	5	400	170	369	31	31	47	35	5	65	17	21	1	19,622	93.87
SNOW	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
ROCK	-	-	10	780	2	-	-	-	-	41	-	-	-	-	-	2	-	-	836	1.17
BARREN	-	-	-	13	119	19	-	10	9	44	6	12	-	-	7	-	-	24	263	4.78
DEVELP	7	-	-	-	54	16	17	7	12	30	55	70	5	1	55	28	10	1	367	14.66
SHRUB	7	-	17	62	50	2,626	1,104	451	191	620	26	422	1,327	1,120	985	750	819	233	10,811	24.29
WET_T	83	-	-	12	3	209	10,792	2,396	608	25	9	38	1,125	331	5,000	370	1,362	304	22,666	47.61
WET_S	37	-	-	39	21	183	3,429	1,809	548	131	507	462	201	397	797	564	236	172	9,534	18.98
WET_H	266	-	-	16	15	97	1,216	1,863	1,415	117	129	226	118	79	331	110	107	71	6,175	22.91
GRASS	18	-	10	1,520	1,479	635	524	1,539	1,143	14,752	616	2,022	231	209	741	633	331	106	26,511	55.64
CROP	58	-	-	622	260	59	358	277	965	852	48,438	18,262	3	1	208	216	36	10	70,625	68.58
P/HAY	51	-	-	52	459	311	995	434	1,294	3,278	5,850	25,698	24	21	547	791	40	44	39,891	64.42
FCD	18	-	2	18	-	71	7,996	334	152	78	526	396	23,348	377	1,103	77	4,893	175	39,565	59.01
FCO	4	-	7	-	-	18	258	8	-	4	-	-	138	12	267	4	34	7	760	1.54
FBD	16	-	-	59	33	697	887	924	291	115	415	986	515	180	33,008	2,649	2,239	314	43,328	76.18
FBO	-	-	-	7	-	29	23	5	-	2	2	19	73	45	94	254	135	71	759	33.44
FMD	21	-	-	9	1	108	1,963	602	98	19	128	52	2,231	233	4,672	416	9,623	376	20,550	46.83
FMO	-	-	-	-	-	-	15	16	-	-	-	-	9	4	-	1	-	-	44	0
TOTAL	19,005	-	47	3,213	2,497	5,083	29,977	10,844	7,095	20,138	56,736	48,713	29,383	3,016	47,881	6,882	19,886	1,910	**312,307**	
PA%	96.91	0	20.92	0.39	2.15	51.66	36	16.68	19.94	73.25	85.37	52.76	79.46	0.39	68.94	3.69	48.39	0	**OA%=**	**60.92**

Table 4.1.1. Confusion matrix for the finest level (18 classes)

for annual crops, and 64% for pasture/hay. If these latter two classes, which are often confused with one another, were collapsed into a single 'agriculture' class, its user's accuracy would raise to 89% (Table 4.1.2). Forest conifer dense (FCD) is the fourth most abundant class in the sample after agriculture (crops plus pasture) and broadleaf forest, and has a user's accuracy of 59%. The largest confusion in FCD occurs with the wetland treed (WET_T) class. Around 20% of the areas classified as FCD in the map were assigned to WET_T by the interpreters. This is not surprising, as a rich black spruce fen is a treed wetland, but it could also be seen as a dense conifer forest. This confusion is likely to be even more widespread in the far north (which was not sampled), so these two classes should be collapsed in some way if the accuracy is to be raised. Mixed forest (UA 49%) is mostly confused with broadleaf forest, and to a lesser degree with conifer forest. This confusion is understandable given the semantic and radiometric similarity between these classes, and can hardly be improved. The open forest classes (FCO, FBO and FMO, respectively open conifer, broadleaf and mixed forest) have very low UA (FCO=2%; FBO=33%; and FMO=0%), although these estimates may not be reliable given their low sampling intensity (e.g., there were only 44 ha of FMO in the validation dataset). In any case, the open forest classes are rare (<5% of the total forest area), so they have no impact on the overall accuracy. Notwithstanding, it may be more parsimonious to suppress them in the next version.

The shrub class also has a low user's accuracy (24%). Almost half of the area marked as 'shrub' in the map was found to be actually forest. This is not surprising given their radiometric similarities, but nevertheless, this is a liability that cannot be tackled unless height information is gathered from other sources, something that we hope to be able to do in the future (but not for the next version replacing the beta version). The grassland class has a better user's accuracy (56%) than shrubland, but it is confused with many different classes, from barren to broadleaf forest, meaning that its accuracy cannot be improved much for the next version. The wetland classes have in general low user's accuracies (48% for wetland treed, 19% for wetland shrub, and 23% for wetland herb), so it would be preferable to suppress them by adding an upland/lowland attribute to the map. The 'developed' class also has low user's accuracy (24%). For example, the city of Edmonton appears mostly as grassland in the map. This will be fixed in the next version. In any case, most of the developed area in the map corresponds to roads, which for the most part were not sampled. Since roads come from an official GIS layer, the actual accuracy of this class has to be necessarily higher than reported. Non-vegetated land classes also show a low accuracy, but they were heavily under-sampled as explained earlier (e.g., there were no polygons belonging to ice/snow in the validation dataset), so their accuracy is likely underestimated.

	REFERENCE															
MAP	WATER	SNOW	ROCK	BARREN	DEVELP	SHRUB	WET_T	WET_S	WET_H	GRASS	AGRI	CONIFER	BROADLF	MIXED_F	TOTAL	UA%
WATER	18,419	-	-	5	0	5	400	170	369	31	78	41	82	23	19,622	93.87
SNOW	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
ROCK	-	-	10	780	2	-	-	-	-	41	-	-	2	-	836	1.17
BARREN	-	-	-	13	119	19	-	10	9	44	19	-	7	24	263	4.78
DEVELP	7	-	-	-	54	16	17	7	12	30	125	6	83	10	367	14.66
SHRUB	7	-	17	62	50	2,626	1,104	451	191	620	449	2,447	1,735	1,052	10,811	24.29
WET_T	83	-	-	12	3	209	10,792	2,396	608	25	47	1,456	5,370	1,666	22,666	47.61
WET_S	37	-	-	39	21	183	3,429	1,809	548	131	969	598	1,360	408	9,534	18.98
WET_H	266	-	-	16	15	97	1,216	1,863	1,415	117	354	197	442	177	6,175	22.91
GRASS	18	-	10	1,520	1,479	635	524	1,539	1,143	14,752	2,638	440	1,375	438	26,511	55.64
AGRI	109	-	-	673	720	369	1,354	712	2,259	4,130	98,248	49	1,763	130	110,516	88.9
CONIFER	22	-	10	18	-	89	8,254	342	152	81	922	23,875	1,451	5,110	40,326	59.21
BROADLF	16	-	-	66	33	727	910	929	291	117	1,422	813	36,005	2,759	44,088	81.67
MIXED_F	21	-	-	9	1	108	1,978	617	98	19	179	2,477	5,088	9,999	20,593	48.56
TOTAL	19,005	-	47	3,213	2,497	5,083	29,977	10,844	7,095	20,138	105,449	32,400	54,763	21,797	312,307	
PA%	96.91	0	20.92	0.39	2.15	51.66	36	16.68	19.94	73.25	93.17	73.69	65.75	45.88	OA%=	69.81

Table 4.1.2. Confusion matrix for level 2 (14 classes)

Finally, the fact that the water class shows the highest accuracy (94%) is not surprising, given its radiometric distinctness and the use of hydrography GIS information for rivers and lakes.

Table 4.1.3. Confusion matrix for level 1 (6 classes)

			REFERENCE					
MAP	WATER	NON_VEG	SHRUB	WETLAND	HERB	FOREST	TOTAL	UA%
WATER	18,419	5	5	939	109	145	19,622	93.87
NON_VEG	7	977	35	54	258	133	1,465	66.69
SHRUB	7	130	2,626	1,746	1,069	5,234	10,811	24.29
WETLAND	386	106	489	24,076	1,643	11,674	38,375	62.74
HERB	127	4,403	1,004	7,531	119,767	4,195	137,027	87.4
FOREST	59	135	924	13,571	2,740	87,577	105,007	83.4
TOTAL	19,005	5,756	5,083	47,916	125,586	108,959	**312,307**	
PA%	96.91	16.98	51.66	50.25	95.37	80.38	**OA%=**	**81.15**

Regarding the accuracy of the 11 individual tiles for which both the regular and oversize strata were sampled, their overall accuracy (OA) ranges between 30% and 69% at level 3, and between 73% and 89% at level 1 (Table 4.1.4). The low OA (30%) in tile 42024 is due to the cropland class - the most abundant in that tile - being heavily confused with pasture/hay (UA 16%), to a point where 75% of the area deemed cropland in the map was pasture/hay in the reference. In fact, when these two classes are lumped together at level 2, OA raises to a good level (77%). The similarly low OA in tiles 43022 and 46020 (respectively, 46% and 48%) is explained in turn by the forest conifer class being very confused with wetland treed (more than a third of the area deemed FCD in those tiles was WET_T in the reference), which leads to a low user's accuracy for this class in both tiles (respectively, 31% and 35%). Given these differences, the coefficient of variation of the OA for the 11 tiles is 21% with 18 classes, albeit it decreases to 9% for 6 classes. This variation is due to a combination of different proportions of LC classes in different tiles, and different degrees of confusion among the classes, although the trends are very consistent from tile to tile.

Table 4.1.4. Overall accuracy and sampling intensity by tiles

TILE	NAT REG	OA% L3	OA% L2	OA% L1	KM2	sarea (ha)	si%
42023	Parkland	62.94	77.53	82.94	20,579	24,113	1.17
42024	Grassland	30.04	77.84	84.47	17,115	18,024	1.05
43021	Boreal	56.03	56.79	74.84	13,520	17,196	1.27
43022	Boreal	45.71	51.68	63.24	14,180	21,164	1.49
43023	Boreal	68.98	77.26	83.7	16,017	18,828	1.18
45021	Boreal	63.31	64.66	82.5	13,949	18,827	1.35
45022	Foothills	59.22	61.17	73.31	29,034	35,722	1.23
46020	Boreal	48.09	53.04	73.24	24,486	28,535	1.17
46021	Boreal	68.66	79.12	87.36	27,146	32,477	1.20
47022	Foothills	58.58	60.76	83.89	9,336	11,618	1.24
48020	Boreal	68.79	70.24	86.28	11,200	18,549	1.66
MEAN		**57.30**	**66.37**	**79.61**	**17,869**	**22,278**	**1.27**
STDDEV		11.97	10.48	7.44			
CV%		20.90	15.79	9.34			

4.2. Polygon-wise likelihood of correct classification

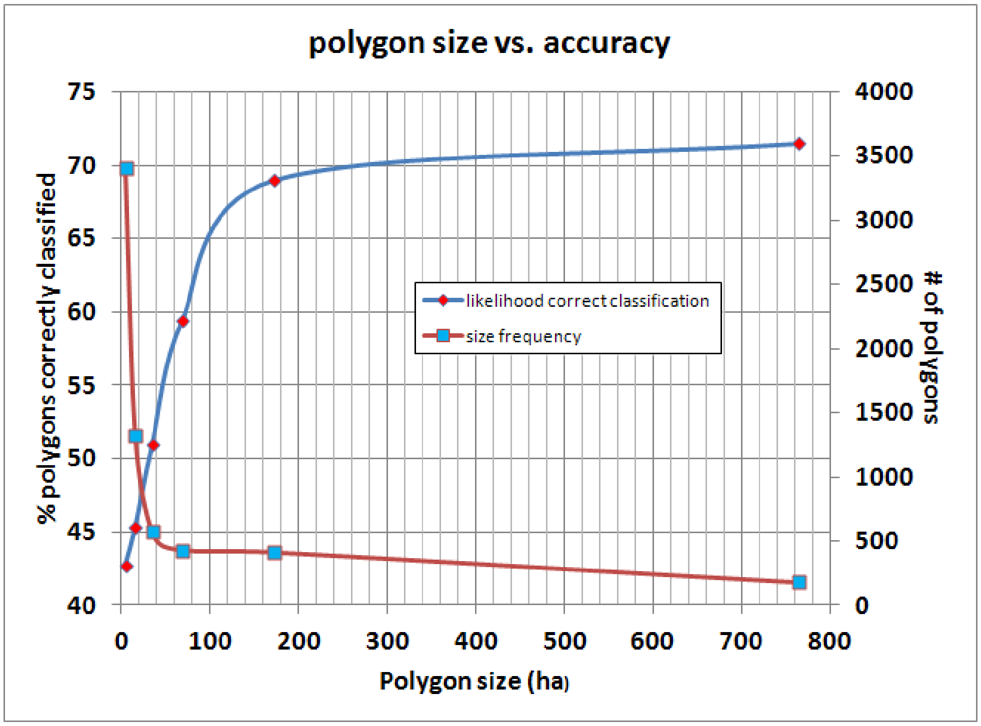


Figure 4.2. Polygon accuracy increases with size

This parameter, which could also be called per-polygon overall accuracy, is the percent of polygons in the sample that have more than 50% of their area actually covered by the LC class indicated in the map. This condition is fulfilled by 3012 out 6317 validated polygons, which yields a **per-polygon overall accuracy of 48%**, that is, the average polygon in the map has roughly a 50/50 chance of having the right label. At first sight, this looks surprising, since this measure, being more forgiving than the area-based OA (it is based on the majority class rather than on the mixture of classes within the polygon), shows a lower value (48% vs. 61%). To understand why this is so, we divided the set of validation polygons in six size intervals (<10 ha; 10-25 ha; 25- 50 ha; 50-100 ha; 100-300 ha; > 300 ha), and computed the per-polygon OA for each interval. We found that accuracy increases with size very rapidly for sizes under 100 ha, and much more slowly for polygons larger than this (Figure 5.2). At the same time, the number of polygons decreases rapidly with size. Combining both observations, one soon realizes that more than half of the polygons in the sample are under 10 ha and have an accuracy of roughly 40%, while the size interval having an accuracy equal or greater than 60% is just a tenth of the total, which explains why this parameter has a lower value than the area-based OA. An important implication of this behavior is that if we randomly removed a portion of the polygons < 10 ha, we would automatically increase the accuracy of the map. If a more selective method of removal were used, where polygons with a greater likelihood of incorrect classification had a greater chance of removal, the gains in accuracy would be considerable.

4.3. Proportion of polygons that represent full patches

This parameter is the percent of validation polygons that originated a 'whole' polygon in the _edited layer, and where at the same time that 'whole' polygon covers at least half of the original

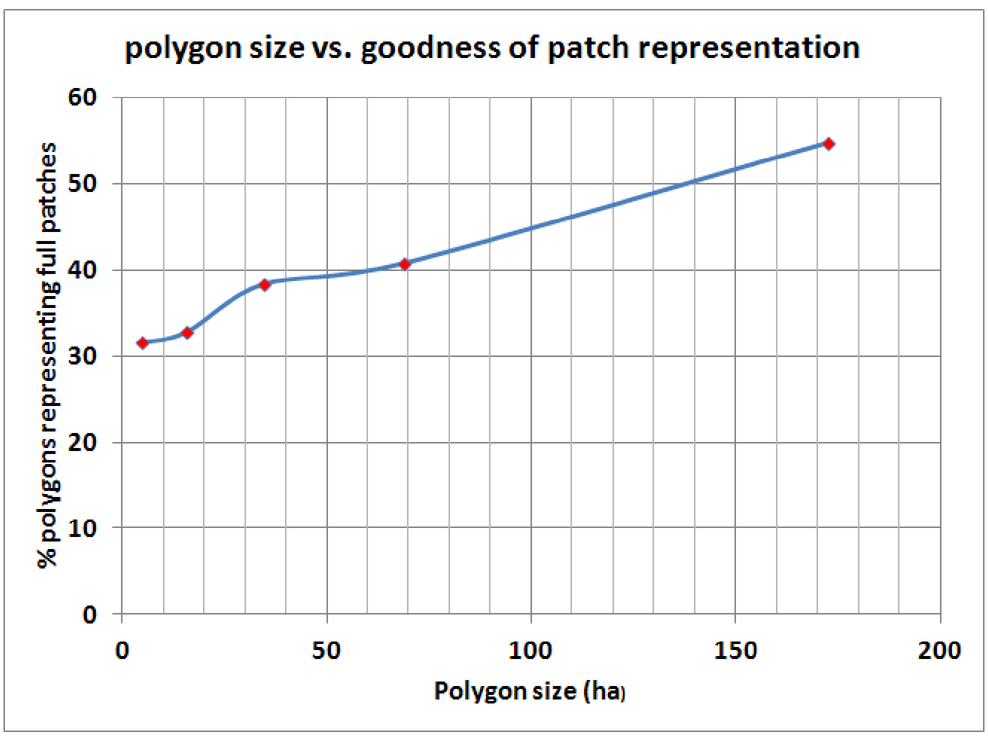


Figure 4.3. Larger polygons are more likely of representing a full patch

polygon, so that the latter can be seen as an (imperfect) representation of a self-contained unit or patch. This gives an indication of how good the mosaic of polygons represents landscape structure. In an ideal map, each polygon should coincide with a distinct patch, meaning that it should encompass all connected local areas of the same landcover type, and exclude areas greater than the MMU from a different type. In a real-world map, this will be hardly the case, but at least a good proportion of polygons should be able to be viewed after inspection as an imperfect representation of a patch, i.e., as having some local errors of omission or commission, but in essence a patch. In the ABMIw2wLCV2000beta, this occurs only for 35% of the polygons (2,129 out of 6,138 polygons in the regular stratum –the oversize stratum could not be checked for this). The remaining 65% are in reality parts of a much larger patch that what misclassified in the area occupied by each of these polygons. The main implication of this is that this map cannot be recommended as input for landscape pattern analyses. Again, the proportion of 'good' polygons increases with size (Figure 4.3). The increase is more irregular and less steep than for polygon-wise accuracy, but it nevertheless confirms that removing a portion of the smaller polygons will result in considerable accuracy gains.

4.4. Accuracy of polygon outlines

We used the set of outlines from the edited 'whole' polygons as a surrogate for a representative subset of accurate boundaries. While a more complex analysis should be possible, for the present we only calculated the length of outlines in the ABMIw2wLCV2000beta that were less than 45 m away from the edges of the edited 'whole' polygons, which yielded 7,244 km. The total length of outlines of 'whole' polygons is 8,256 km, which can be interpreted as follows: from those outlines in the ABMIw2wLCV2000beta that correspond to true landcover transitions, 88% of their length is within 45 m of the actual boundary. However, we could not compute the proportion of map outlines that correspond to true landcover transitions, because we did not mark what line segments of the polygons deemed 'parts' were such (NB. unless a 'part' polygon is a gap or island, there will be at least one of its 'sides' that is a true boundary). Notwithstanding, our qualitative visual assessment suggests that whenever there are crisp edges in the satellite images, they are well captured in the map, but there is high proportion of outlines, maybe more than 50%, that in reality, have the same landcover on both sides.

5 CONCLUSION AND PROPOSED CHANGES

All accuracy parameters derived from our analysis suggest that the ABMIw2wLCV2000beta map, as it stands in this beta version, cannot be considered a reliable source of landcover information. This is true for most applications, except perhaps for non-local analyses where a rough estimate is sought about the composition of LC classes in an area larger than a township. The map's poor accuracy has likely been inherited from the source raster layers used as input to create it, and thus can only be tackled by applying a more aggressive thematic and spatial generalization, at the cost of losing some detail in both dimensions. With the information gathered through this accuracy assessment, we have a fair idea of the trade-offs involved. We suggest, as an overarching measurable target, that the area-based overall accuracy be raised from 61% to at least 85%. For this, we propose the following changes for the final version (which would be version 2.1 rather than 1.0, as it would be the third version –2nd to go public):

1) Collapse the classes 'annual crops' (121) **and 'pasture/hay'** (122) into 'agriculture' (120), given their high degree of confusion. There is no way out around this.

2) Suppress the dense/open distinction in forest classes. 'Open' classes (crown closure under 50%) have poor user accuracies and occupy less than 5% of the forest. Therefore we would be gaining some additional accuracy without losing much information. If desired, a new field, 'crown closure,' could be appended to the attribute table that could be filled only for polygons within AVI coverage, for the present.

3) Eliminate the wetland classes. ‘Treed wetland’ and ‘conifer forest’ are two classes that are much confused in the boreal, given the large expanses of black spruce fens that occur there. Furthermore, they overlap semantically (the above fens can be classified as both treed wetland and conifer forest). So instead of having wetland LC classes, it makes more sense to append a new attribute called ‘topographic position’, with two values; ‘upland’ and ‘lowland’. All polygons classified as wetlands would have the value ‘lowland’ in this attribute. Polygons larger than 300 ha formerly belonging to treed wetland would be automatically be relabeled conifer (as they will be covered in all likelihood by black spruce or tamarack), and for those smaller, the broadleaf/conifer decision would be based on a combination of geographic position, proximity to open water, and appearance in the image (this has yet to be implemented into a specific set of rules).

4) Remove clutter. Most areas of the map display a myriad of small isolated polygons that for the most part, are actually covered by the same LC class as their surroundings. The removal of these polygons would not only increase accuracy, but it would also lead to a more aesthetically pleasing map. For each of these polygons, the decision on whether to remove or keep them could be based on a contrast threshold in the Landsat images, which we could derive from the validation dataset. Polygons with less contrast than this threshold would be merged to the surrounding polygons, or, if they have more than one neighbor, to the radiometrically nearest neighbor (i.e., most similar in the image).

In conclusion, although the results from this accuracy assessment are disappointing, they will allow us to make better decisions on how to improve the final version so that it can be rendered a source of reliable information on landcover.

[**Postscriptum**. These changes were implemented in version 2.1 of the ABMIw2wLCV2000, and as a result the overall accuracy rose from 61% to 75 % (88% at level 1). The accuracy assessment of version 2.1 used the same validation areas used for the beta version and this time only included the area-based confusion matrix, since the other parameters would require polygons extracted from the new map]

APPENDIX 1 : Description of Level 3 LC classes (used for beta version)

20 – **Water**: lakes, lagoons, rivers, canals, and artificial water bodies. Shallow open water is included in this category, unless there is more than 20% vegetation cover, in which case it belongs to the relevant wetland class.

31 – **Snow/Ice**: areas permanently covered by snow or ice, including glaciers.

32 – **Rock/Rubble**: bedrock, rubble, talus, blockfield, lava beds, or other natural impervious surfaces.

33 – **Exposed Land**: bare soil (barren, non-agricultural), river sediments and cutbanks, pond or lake sediments, reservoir margins, beaches, landings, recently burned areas, mudflat sediments, surface mining, or other non-vegetated (less than 6% trees, or less than 20% shrub/herb) surfaces.

34 – **Developed**: urban and built-up areas (including industrial sites), impervious artificial surfaces (e.g. airport runaways), railways and roads. Acreages and farmsteads are included in this class. Oil and gas well pads are included in this class if connected to a road and not abandoned or under reclamation. Urban terrain under development is included in this class, even if the land is exposed. Urban green areas are excluded of this class if larger than 2 ha and if they have less than 2 buildings per hectare.

50 – **Shrubland**: At least 20% ground cover which is at least one-third shrub, with no or little presence of trees (<6% crown closure). Examples of plants belonging to this class are prickly rose, Rocky Mountain juniper, and sagebrush. Alder and most willows are also shrubs, but depending on the soil moisture regime, they can be wetland shrubs. Note that a dense patch of regenerating young trees is still considered forest and not shrub, no matter that the trees are still small.

81 – **Wetland – Treed**: flat land and depressions with a water table near/at/above soil surface for enough time to promote the accumulation of peat and foster the presence of at least 10% cover of wet-tolerant trees. This class includes treed bogs (black spruce, Sb) and fens (Sb and/or larch), which occupy big expanses of the boreal forest. Treed swamps along river floodplains or treed wet flats around lakes are also included in this class.

82 – **Wetland – Shrub**: Land with a water table near/at/above soil surface for enough time to foster the presence of hydrophytic shrubs such as willows, alder and Labrador tea. At least one-third of vegetation is shrub, and < 6% cover are trees. Usually associated with floodplains and the shores of lakes and streams. Often form a transitional zone between marshes or shallow open water and uplands.

83 – **Wetland - Herb** Land with a water table near/at/above soil surface for enough time to promote wetland or aquatic processes. At least 20% vegetation cover, of which trees are < 10%, shrubs are < 30%, and the rest is either graminoids (sedges), forbs, or bryoids (the latter only in the far north). Marshes (the most common wetland in the Southern Alberta) belong to this class, but also non-woody peatland and swamps.

110 – **Grassland**: Predominantly native grasses and other herbaceous vegetation with a minimum of 20% ground cover, may include some shrub cover (but less than a third of the vegetated area) or a few trees (but the tree cover cannot exceed 6%). Land used for range or native unimproved pasture (e.g., rough fescue) is included in this class. Alpine meadows fall into this class.

121 – **Cropland**: annually cultivated cropland and woody perennial crops. Includes annual field crops, vegetables, summer fallow, orchards and vineyards. Excludes forage crops (they belong to class 122). Bare agricultural (i.e., tilled) soil belongs to this class.

122 – **Pasture/Hay**: cultivated fields planted with tame (non-native) grass (such as clover) and legume species (such as alfalfa), either with livestock directly grazing on the fields (tame pasture), or which are periodically cut to provide forage for livestock (i.e., hay fields).

211 – **Coniferous – Dense**: greater than 50% crown closure; coniferous trees are 75% or more of crown closure.

212 – **Coniferous – Open**: 6-50% crown closure; coniferous trees are 75% or more of crown closure.

221 – **Broadleaf – Dense**: greater than 50% crown closure; broadleaf trees (trembling aspen, balsam poplar and white birch, the latter only in northern Alberta) are 75% or more of crown closure.

222 – **Broadleaf – Open**: 6-50% crown closure; broadleaf trees (trembling aspen, balsam poplar and white birch) are 75% or more of crown closure.

231 – **Mixed Wood – Dense**: greater than 50% crown closure; neither coniferous nor broadleaf tree account for 75% or more of crown closure.

232 – **Mixed Wood – Open**: 6-50% crown closure; neither coniferous nor broadleaf tree account for 75% or more of crown closure.

UNIVERSITY OF CALGARY

ALBERTA BIODIVERSITY MONITORING INSTITUTE

ABMIw2wLCV2000beta hierarchical legend

- 0-No data
- 10-Unclassified
- 11-Cloud
- 12-Shadow
- 20-Water
- 30-Barren/Non-vegetated
  - 31-Snow/Ice
  - 32-Rock/Rubble
  - 33-Exposed land
  - 34-Developed
- 50-Shrubland
- 80-Wetland
  - 81-Wetland - Treed
  - 82-Wetland - Shrub
  - 83-Wetland - Herb
- 100-Herb
  - 110-Grassland
  - 120-Cultivated Agricultural Land
    - 121- Cropland
    - 122- Pasture
- 200-Forest/Tree classes
  - 210-Coniferous Forest
    - 211-Coniferous Dense
    - 212-Coniferous Open
  - 220-Deciduous Forest
    - 221-Broadleaf Dense
    - 222-Broadleaf Open
  - 230-Mixed Forest
    - 231-Mixedwood Dense
    - 232-Mixedwood Open