AN INTEGRATED FRAMEWORK FOR ASSESSING THE ACCURACY OF GEOBIA LANDCOVER PRODUCTS

G. Castilla [a, *], A. Hernando[b], C. Zhang[a], D. Mazumdar[a], G.J. McDermid [a]

[a] Dept. of Geography, University of Calgary, 2500 University Drive, Calgary AB T2N1N4, Canada - (gcastill, chunzhan, dmazumda, mcdermid)@ucalgary.ca
[b] Research Group For Sustainable Management, Polytechnic University of Madrid, Spain – ana.hernando@upm.es

KEY WORDS: Object Based Accuracy Assessment, Polygon Sampling, Landcover Mapping

ABSTRACT:

Vector-based landcover (LC) maps derived from GEographic Object-Based Image Analysis (GEOBIA) are increasingly replacing the traditional raster maps from per-pixel classification, but our strategies for assessing their quality are not yet fully developed. We contend that a complete accuracy assessment of a vector LC map must provide answers to the following questions: (1) What is the proportion of area assigned to each LC class that is actually covered by that class? (2) How does the area wrongly assigned to a class get distributed into the other classes? If we were flying at a low altitude over any given polygon, what is the likelihood that we would agree that (3) the LC class best representing the interior of the polygon is the one appearing on the map; (4) the area enclosed by the polygon can be seen as a self-contained unit or patch; (5) there are no regions, either next to the outside of the polygon or on its inside, that would have better be included in the polygon or excluded from it; and (6) the outline of the polygon (excluding parts affected by 5) follows reasonably well the LC transitions we appreciate from air? Questions 1 and 2 can be answered using a confusion matrix, but not the rest. We discuss the conceptual foundations of our integrated object-based approach to accuracy assessment, and demonstrate its implementation for a wall to wall vector LC map of Alberta, Canada.

1. INTRODUCTION

Landcover (LC) maps created using GEOBIA typically consist of a mosaic of non-overlapping LC objects, each representing a *patch*, i.e., an area that, in terms of land cover, is both relatively homogeneous internally and different from the surroundings. As any other GIS product, object-based LC maps need to be accompanied by quantitative data on their accuracy. Unfortunately, there are currently no widely accepted methods to assess the latter. Conventional pixel-based accuracy assessment is not applicable in this context. Objects have to be evaluated as wholes in context with their surroundings; hence it is not possible to use a few pixels or plots within them to assess their goodness. For example, some of the reference pixels for a polygon could happen to lie on a small area of different LC that because of its reduced size was intentionally subsumed in the polygon under evaluation. Thus, what would be counted as an error in the confusion matrix is in reality a necessary spatial generalization. Neither can a pre-existing LC map be used as reference. The patches, represented as polygons in the LC map, are fiat objects whose existence as individual entities primarily depends on human cognition: different classification schemes will yield different patches, and even with the same legend and input data, different segmentation algorithms and classification strategies will lead to different partitions of the same landscape, which are not necessarily better or worse than the rest. Since, as result, there is no objective ‘ground truth’ to use as reference, the accuracy assessment of an object-based LC map must use the map itself as a starting point.

Moreover, given that the quality of that map is a multifaceted issue that cannot be captured by a single metric, the assessment must also include other parameters than just the overall accuracy or the per-class user or producer accuracy. In particular, there are, for object-based LC maps, three more aspects of accuracy in addition to those covered by the confusion matrix:

1. **Thematic**: is the LC class best representing the interior of the polygon the one appearing on the map?
2. **Structural**: is the area enclosed by the polygon truly a patch? That is, are all its surroundings covered by a different LC class? Are there sizeable parts in its interior that had rather been placed in a different polygon?
3. **Positional**: are those polygon outlines that correspond to true landcover transitions close enough to them on the ground?

Clearly, new accuracy parameters beyond the confusion matrix are required. Here we propose an integrated framework for assessing the accuracy of GEOBIA LC products using a single, streamlined process. We are currently applying it to a Landsat-derived LC map of Alberta (660,000 km^2), Canada (fig. 1) that contains over two million polygons belonging to 18 possible LC classes (which at a higher hierarchical level become *water, non-vegetated, wetland, shrub, herbaceous* and *forest*), and which has a MMU (i.e., minimum polygon size) of 0.5 ha for water, 1 ha for wetland, and 2 ha for the rest. We provide details on the *sampling design* (how and how many polygons are selected); the *response design* (how each selected polygon is assessed); and the *analysis* (what accuracy parameters and how they are derived).

2. METHODOLOGY

2.1 The framework at a glance

1. A set of validation polygons is selected using stratified (by LC class) random sampling with equal intensity by area.

---

[*] Corresponding author

2. Selection and assessment is carried out separately for regular-size polygons and for very large polygons.

3. Each validation polygon is visualized in a true color orthophoto of submetric resolution; split into homogeneous parts (if necessary); and assigned a LC class (eventually to each split part in it).

4. Regular-size polygons are also reshaped when they are incomplete representations of a patch and the latter can be captured without going too far away from the polygon.

5. The final edited validation layer is intersected with the LC map and the accuracy parameters are derived through a series of automated GIS scripts.

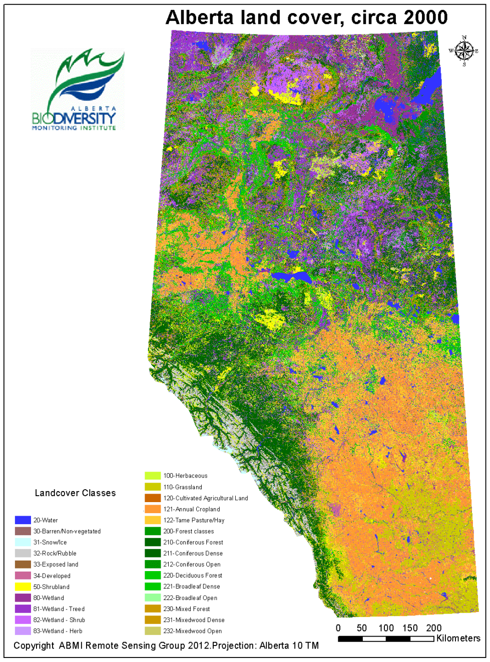


Figure 1. The LC map that motivated the framework

2.2 Sampling design

Our LC map was produced stitching together some 40 non-overlapping tiles of various sizes, each completely encompassed within a Landsat scene that was processed independently. Thus we use the tiles as primary sampling units, from which the secondary sampling units (i.e., validation polygons) are extracted. We apply the same sampling scheme and sampling intensity to each tile, so that we can later on lump together or combine the results irrespective of the tile. Initially we selected 16 tiles, amounting to approximately half of the map extent. Further tiles could be later processed if the variation in the results advises so.

In each tile, polygons are assigned to one of two size-strata according to whether they are smaller than 300 ha (*regular stratum*) or they exceed this size (*oversize stratum*). We use this separation to avoid having to assess structural accuracy parameters in very large polygons, since it would be too time consuming. We can do so because polygons > 300 ha are less than 2% of all polygons in the map, therefore they would make a negligible contribution to the value of those per-polygon parameters. The 300 ha threshold roughly divides the province in two halves (i.e., half of the map extent consist of polygons < 300 ha, and the other half is occupied by oversize polygons). Therefore both size strata have the same weight when it comes to estimate area-based accuracy parameters.

For each tile and size stratum, we randomly select, using an ad hoc tool we created (Castilla et al. in preparation), polygons sequentially, until a 1% of the area covered by each LC class present in the tile is sampled, thus ensuring a proportional representation of the different LC classes. For the regular stratum, we impose the constraint that no two adjacent polygons can be in the sample, so as to minimize the likelihood that the corrections from different validation polygons overlap.

2.3 Response design

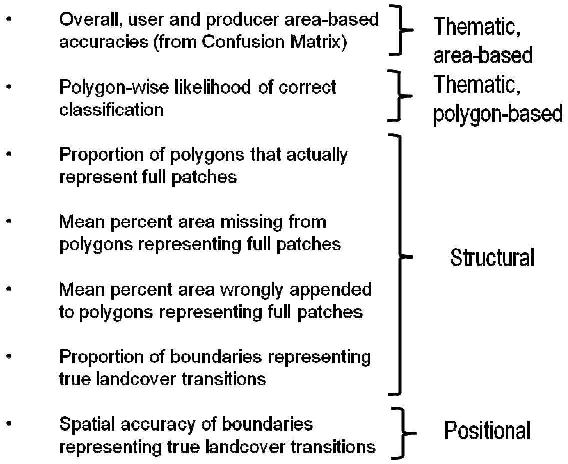
The selection for each tile and size stratum is exported to a separate shapefile stripped of LC attributes, a copy of which will be used by the interpreters during validation. The latter consists in visualizing, in a color ortho-image of submetric resolution, each polygon; assessing land cover in and around the polygon; and, following the decision flowchart in figure 2, splitting the polygon into homogeneous parts > MMU if necessary; and assigning a LC class to the polygon (or to each of its parts if it was split). In addition, polygons from the regular stratum are reshaped if they are at odds with the spatial distribution of LC appreciated in the image and could become 'whole' patches after some edits. That would be the case of a polygon that for the most part overlaps a pond in the imagery, but where there is some sizeable (i.e., > MMU) portion of pond outside the polygon. To avoid an inordinate amount of digitization work, reshaping is only allowed if the resulting patch is less than three times the size of the originating polygon or part. This threshold is a compromise between obtaining a sufficient number of validation polygons that become 'wholes' (to estimate structural accuracy parameters) and the time devoted to create them. Besides, the greater the size of the resulting 'whole' patch relative to the originating part, the more debatable is that the latter was in 'essence' the former. Since in any case patches are fiat objects, interpreters are asked to 'go' with the map in ambiguous settings (they have available the outlines of all polygons in the map –without LC information, to provide spatial context), and only change the delineation when it clearly does not make sense. Ambiguous situations can also arise from the thematic point of view, for example, when the setting is a borderline case between those two classes (e.g., conifer dense vs. conifer open), or when there are insufficient clues in the imagery to make a univocal call (e.g., the setting is likely a grassland, but there are some faint signs of grazing livestock, so it could also be a pasture). For these situations, and only for them, interpreters are allowed to enter a second LC class in an ancillary field. See figure 3 for an example of the correction of a validation polygon.

After validation, the final polygons in the edited layer of the regular stratum correspond to either 'parts' of a much larger patch, or to 'whole' patches. There is a 'type' field in the attribute table of that layer where the 'part'/'whole' membership is stored. Only 'whole' polygons are used for the computation of structural accuracy parameters. 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 Arc2Earth plugin enables us to bring Google Earth (GE) imagery directly into the ArcMap window. When the GE image is not of submetric resolution for a particular polygon, the interpreter can pull a suitable orthophoto from our intranet repository using ArcSDE. The average correction time for a validation polygon of the regular stratum is 3 minutes, and 6 minutes for an oversize polygon.

After the interpreter has completed a tile, she revisits each polygon to ascertain she still agrees with what she did. If for some polygon she feels there is a better solution, new corrections are performed. Parallel to this, she pastes screen captures of each edited polygon to a slide presentation for quick inspection by a supervisor, who is a certified photo-interpreter. Upon inspection, the supervisor may ask the interpreter to make revisions of some detected faulty corrections or biases, or may perform the corrections himself if it is pertinent (e.g., removing second LC calls when they are not warranted). All the response design procedures are precisely detailed in a manual (F3GISci 2012), which also includes LC class definitions and a photo-key.

2.4 Analysis

Table 1 contains the list of accuracy parameters that are derived from the framework. All of them are computed automatically using a combination of GIS overlay operations plus scripts that operate in the relevant attribute tables. The area-based confusion matrix is computed independently for the regular and oversize strata from the intersection between the corresponding edited layers and the LC map. For each intersect polygon there is a LC class original and another (two, in ambiguous cases) from the validation. If the former coincides with (one of) the latter, the area of the intersect polygon goes to the diagonal of the confusion matrix, to the corresponding non-diagonal cell otherwise. The two confusion matrices are later added together to create the overall matrix for the tile. The polygon-wise likelihood of correct classification is proportion of validation polygons where their original class is at least half of the area of the polygon in the edited layer. The proportion of polygons that represent full patches is 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. The mean percent area missing from polygons representing full patches is the arithmetic mean, for polygons included in the previous parameter, of the ratio of area of the 'whole' polygon outside the original polygon versus area inside. The mean percent area wrongly appended to polygons representing full patches is like the previous parameter, but the ratio this time is the area of the original polygon outside the 'whole' polygon versus area of the 'whole' polygon. The proportion of boundaries representing true landcover transitions is the length of outlines from 'whole' polygons that are less than 120 m away from outlines of the LC map (120 m is double the target positional accuracy; further away is considered no longer a positional error but a thematic one), divided by the total length of outlines of validation polygons from the regular stratum. (NB. This is a rough estimate that will be refined in future versions; it likely has negative bias, since very large polygons are not included, and regular polygons catalogued as 'parts', if they are not islands or gaps, usually have a portion of their perimeter that is a true landcover boundary). Finally, the spatial accuracy of boundaries representing true landcover transitions is computed as the weighted average between the mean distance between the outlines from 'whole' polygons that are less than 120 m and more than 45 m away from outlines of the LC map, and 45 m, weighted by the relative proportion of length of both groups (the second group being outlines from 'whole' polygons that are less than 45 m apart from outlines of the LC map; the reason for this is that we don't ask to correct outlines that are less than 45 m away from the true boundary, as they are already good enough –see caption of figure 3). This is again a rough estimate that can be easily refined with a little additional work (using e.g. linear transects and evaluating the intersections with the LC map). A more detailed manuscript with further refinements will be submitted to a peer-reviewed journal this year.



Overall, user and producer area-based accuracies (from Confusion Matrix)	Thematic, area-based
Polygon-wise likelihood of correct classification	Thematic, polygon-based
Proportion of polygons that actually represent full patches	Structural
Mean percent area missing from polygons representing full patches	
Mean percent area wrongly appended to polygons representing full patches	
Proportion of boundaries representing true landcover transitions	
Spatial accuracy of boundaries representing true landcover transitions	Positional

Table 1. List of accuracy parameters derived from the framework

3. CONCLUSIONS

We have designed an integrated framework to assess the accuracy of large-area landcover polygon layers and implemented it into a single streamlined process. The manual and some of the tools will be made freely available from our website (http://www.ucalgary.ca/f3gisci).

4. ACKNOWLEDGMENTS

This research was funded by the Alberta Biodiversity Monitoring Institute (ABMI), the Natural Sciences and Engineering Research Council of Canada (NSERC), and Alberta-Pacific Industries inc. (Al-Pac).

5. REFERENCES

Foothills Facility for Remote Sensing and GIS Science (F3GISci), 2012. Interpreter's Manual for the Accuracy Assessment of the Alberta wall to wall landcover polygon vector layer circa 2000, Version 2.0. 50 pp.

Castilla, G., Hernando, A., Zhang, C., McDermid, G.J., in preparation. Random selection of non-adjacent polygons.

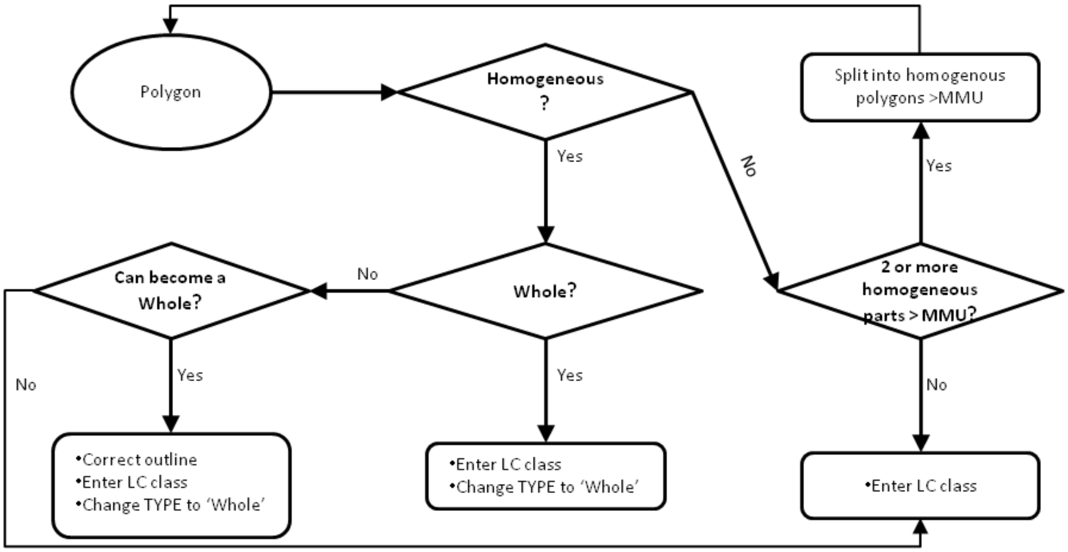


Figure 2. Response design decision flowchart

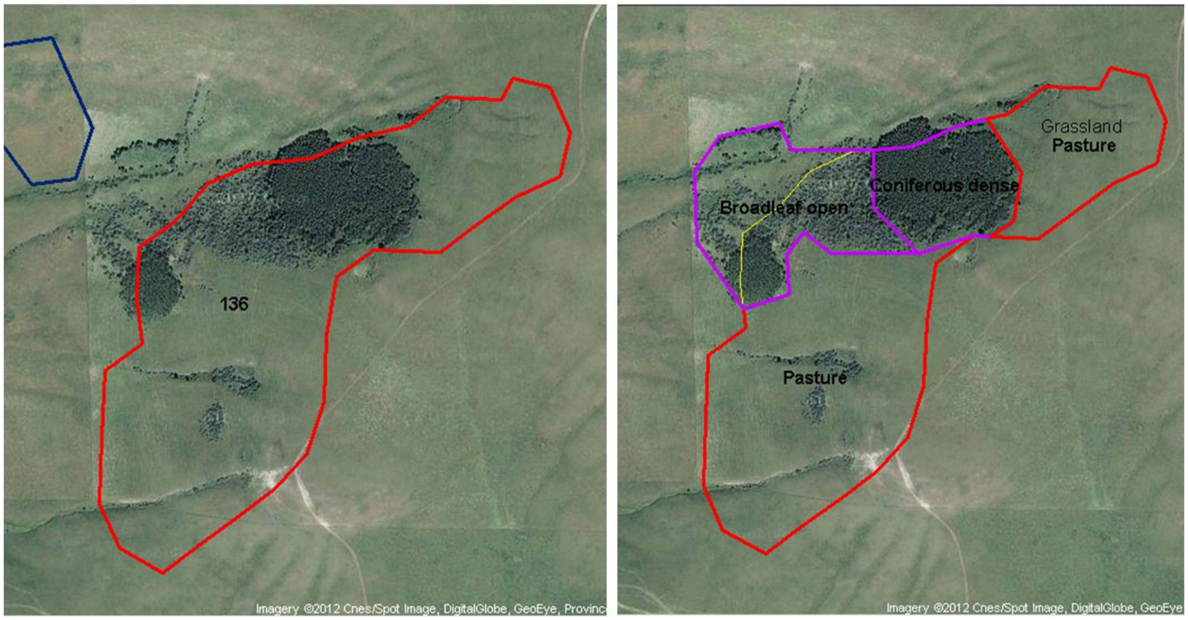


Figure 3. Two 1 km^2 screen captures of a GE image with the same validation polygon (ID 136, regular stratum) before (left) and after (right) correction. The validation polygon was splitted into four polygons, two 'parts' and two 'wholes' (purple outlines). From the two 'wholes', one required reshaping to add an area outside of the original validation polygon. Note that the northern outline of the 'conifer dense' patch was not reshaped because it is within 1.5 Landsat pixels of the true boundary (the target positional accuracy of this map is 60 m). Note also that the NE part has a second label (grassland) because it wasn't clear for the interpreter if it was a 'pasture' (NB. The supervisor later removed this second label after the QC). Finally, note that the dense pocket of coniferous trees to the midwest of the polygon was included in the 'broadleaf open' 'whole' patch because it is <MMU and thus was appended to the most similar adjacent patch. The two little treed pockets further south area also < MMU and thus were ignored. The entire validation process for this polygon should take less than 3-4 minutes using our framework.