Validation of a large area land cover product using purpose-acquired airborne video

M.A. Wulder*, J.C. White, S. Magnussen, S. McDonald

Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, Victoria, British Columbia, Canada

* Corresponding author:
506 West Burnside Rd., Victoria, BC V8Z 1M5;
Phone: 250-363-6090; Fax: 250-363-0775; Email: mike.wulder@pfc.cfs.nrcan.gc.ca

Pre-print of published version.

Reference:

Wulder, M.A., White, J.C., Magnussen, S., McDonald, S. 2007. Validation of large area land cover product using purpose-acquired airborne video. Remote Sensing of Environment, 106: 480–491.

DOI.

DOI: http://doi:10.1016/j.rse.2006.09.012

Abstract

Large area land cover products generated from remotely sensed data are difficult to validate in a timely and cost effective manner. As a result, pre-existing data are often used for validation. Temporal, spatial, and attribute differences between the land cover product and pre-existing validation data can result in inconclusive depictions of map accuracy. This approach may therefore misrepresent the true accuracy of the land cover product, as well as the accuracy of the validation data, which is not assumed to be without error. Hence, purpose-acquired validation data is preferred; however, logistical constraints often preclude its use - especially for large area land cover products. Airborne digital video provides a cost-effective tool for collecting purpose-acquired validation data over large areas. An operational trial was conducted, involving the collection of airborne video for the validation of a 31,000 square kilometre sub-sample of the Canadian large area Earth Observation for Sustainable Development of Forests (EOSD) land cover map (Vancouver Island, British Columbia, Canada). In this trial, one form of agreement between the EOSD product and the airborne video data was defined as a match between the mode land cover class of a 3 by 3 pixel neighbourhood surrounding the sample pixel and the primary or secondary choice of land cover for the interpreted video. This scenario produced the highest level of overall accuracy at 77% for level 4 of classification hierarchy (13 classes). The coniferous treed class, which represented 71% of Vancouver Island, had an estimated user's accuracy of 86%. Purpose acquired video was found to be a useful and cost-effective data source for validation of the EOSD land cover product. The impact of using multiple interpreters was also tested and documented. Improvements to the sampling and response designs that emerged from this trial will benefit a full-scale accuracy assessment of the EOSD product.

1.0 INTRODUCTION

The classification of land cover over large geographic areas with remotely sensed data is increasingly common; regions (Homer *et al.*, 1997), nations (Loveland *et al.*, 1991; Fuller *et al.*, 1994; Cihlar and Beaubien, 1998), continents (Stone *et al.*, 1994), and the globe (Loveland and Belward, 1997; Loveland *et al.*, 2000; Hansen *et al.*, 2000) have been mapped with a variety of satellite data types. This surge of interest in large area land cover mapping projects may be explained by an increase in image availability, a need for national- and global-scale land cover products for modelling and monitoring activities, and political obligations related to international treaties such as the Convention on Climate Change (Kyoto Protocol). Standard operational protocols for the validation of these products are emerging (Loveland *et al.*, 1999; Justice *et al.*, 2000; Strahler et al., 2006; Wulder et al., 2006a), and a new initiative is addressing both the harmonization and validation of large area land cover products (Herold *et al*., 2006).

A sufficient level of accuracy is assumed in order to rationalize the applied use of these large area land cover products for a wide variety of applications (Morisette *et al.*, 2002; Stehman and Czaplewski, 2003). Accuracy assessment protocols require validation data that is independent from information used in map development; however, validation data is expensive and logistically challenging to collect for large area land cover products (Cihlar, 2000). As a result, pre-existing data are often used for validation of large area land cover products generated from remotely sensed data. Unfortunately, temporal, spatial, and attribute differences between the land cover product and the pre-existing validation data can result in poor levels of apparent accuracy (Remmel *et al.*, 2005). Furthermore, the use of pre-existing validation data may misrepresent the true accuracy of the land cover product, while at the same time, revealing problems inherent in the validation data. Aerial videography is one means of collecting purpose-acquired validation data for land cover products extending over large areas (Slaymaker, 2003).

Airborne videography became known as a flexible and cost effective remote sensing tool in the early 1980s (Meisner, 1986; Mausel *et al.*, 1992; King, 1995), and has demonstrated utility for a wide range of applications including species identification and vegetation mapping (Nixon *et al*., 1985; Bobbe *et al.*, 1993; Frazier, 1998; Suzuki *et al*., 2004), forest health and damage assessment (Jacobs and Eggen-McIntosh, 1993; Jacobs, 2000); forest inventory update (Brownlie *et al*., 1996; Davis *et al*., 2002); and validation of vegetation maps generated from medium resolution remotely sensed data (Graham, 1993; Marsh *et al.*, 1994; Slaymaker *et al*., 1996; Hepinstall, 1999; Hess *et al.,* 2002). Marsh *et al.* (1994) compared the use of systematically acquired aerial colour photography and airborne video for the validation of land cover products generated from Landsat TM imagery. The similarity between the accuracies measured by the photography and video sources was statistically significant ($\alpha = 0.01$), suggesting that video could provide validation data of similar quality and utility to that of traditional aerial photography, with the advantage of collecting a larger volume of data with the same level of effort.

Aerial videography gained additional momentum in the 1990s as a result of the GAP analysis programs implemented in the United States (Slaymaker, 2003). GAP programs operated at the state level and were initiated to assess the extent to which native animal and plant species were being protected. GAP land cover maps generated from Landsat TM data required a source of calibration data to facilitate the classification of the imagery, as well as a source of validation data to assess the accuracy of the output vegetation maps (Scott *et al.,* 1993; Slaymaker, 2003). Aerial point sampling methods developed by Norton-Griffiths (1982) for aerial photography were adapted to aerial videography by Graham (1993) and implemented in the Arizona GAP program. The automatic labelling of video frames with Global Positioning System (GPS) coordinates was a major technological advancement, "providing a way to precisely and automatically match video-recorded GPS time with the position information in the GPS data file" (Graham, 1993: 29). Slaymaker *et al.* (1996) modified this approach for the GAP program in New England. Aerial video for calibration and validation of Landsat-based land cover maps was subsequently adopted by many other state-wide GAP programs (Schlagel, 1995; Driese *et al.,* 1997; Hepinstall *et al.*, 1999; Reiners *et al.*, 2000), as well as for other land cover products (Skirvin *et al.*, 2000; Hess *et al.*, 2002; Maingi *et al.*, 2002; Skirvin *et al.*, 2004). Technological advances in the 1990s led to the development of digital video cameras and powerful multi-media capabilities in desktop computers, further enhancing the quality and affordability of video options (Hess *et al.* 2002)

The advantages and disadvantages of airborne video for validation of land cover maps are summarized in Slaymaker (2003). One of the main advantages of video is data redundancy (Mausel *et al.* 1992), which facilitates the collection of large samples of validation (and calibration) points at specified intervals (of time or distance). In the early 1990s, the inferior resolution of video (240 lines for colour, 300 lines for panchromatic) compared to that of 35 mm photo film (1500 lines) was an impediment to the widespread adoption of video for land cover applications (Slaymaker, 2003). Since this time, the quality of video has improved dramatically; with current consumer grade digital video cameras typically having more than 500 lines (colour) (Jack, 2005).

The main goal of the Earth Observation for Sustainable Development of Forests (EOSD) program is the creation of a land cover map of the forested area of Canada, produced to represent year 2000 conditions, and on track for completion in 2006 (Wulder *et al.* 2003). There are many areas in the north of Canada being mapped for the EOSD project that are inaccessible, have no pre-existing detailed forest or vegetation inventories, and minimal or out-dated aerial photography. While a framework for validating the EOSD product has been developed (Wulder *et al.*, 2006a), an alternative data source, along with a protocol for using this data source to validate the EOSD product, is required. The objective of this study was to develop a protocol and demonstrate the use of airborne video as a source of validation data for large area land cover products generated from remotely sensed data, specifically the EOSD land cover product. The rationale and approaches demonstrated here are intended to be portable to other large area mapping programs.

To fully explore the potential of airborne video for validation of the EOSD product, an operational trial was conducted on Vancouver Island, British Columbia, Canada. Video data was used for validation and the accuracy estimates were compared to the estimates obtained using pre-existing forest inventory data for validation. Pre-existing data is often considered a viable source of validation data, despite fundamental differences that often exist between the pre-existing data and the product being validated. This communication details the protocol developed through this operational trial, the results of the accuracy assessment using the airborne video, the impact of multiple interpreters, and suggested improvements for future implementation of the video system for validation.

2.0 STUDY AREA

Vancouver Island has a total land area of 31,284 square kilometres (Figure 1). Much of the island (85%) lies in the Coastal Western Hemlock biogeoclimatic zone (Klinka *et al.* 1991). This zone is characterized as one of Canada's wettest climates and most productive forest areas, with cool summers and mild winters. The rugged physical features of Vancouver Island include long mountain-draped fjords on the west coast, coastal plains on the eastern coast, and a chain of glaciated mountains running along the north-south axis of the Island. Elevations range from 0 to 2200 metres. Forests cover 91% of Vancouver Island and forest species, in order of prevalence, include Hemlock (*Tsuga spp*.), Western red cedar (*Thuja plicata* Donn ex. Don), Western hemlock (*Tsuga* *heterophylla* (Raf.) Sarg.), Yellow cedar (*Chamaecyparis nootkatensis* (D. Don) Spach.), and Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco). The existing forest inventory indicates that approximately 45% of the forest on the island is 250 years in age or older, with the remaining forest in managed second-growth stands.

3.0 DATA

3.1 Earth Observation for Sustainable Development of Forests (EOSD)

In support of meeting national and international monitoring and reporting commitments, the Canadian Forest Service, in partnership with the Canadian Space Agency, is classifying Landsat data over the forested area of Canada. Over 400 Landsat scenes are being used to map the approximately 400 million hectares of treed land present in Canada (Wulder and Seemann 2001), with 20 different land cover classes (plus additional classes for no data, shadow, and cloud) (Wulder and Nelson, 2003) as listed in Table 1. Including image overlap into non-forest areas, over 60% of the country will be mapped by the EOSD project, with image classification undertaken by federal, provincial, and territorial agencies (Wulder *et al.* 2003). Source images cover acquisition dates ranging from 1999 to 2002 and the EOSD product has a pixel size of 25 metres.

The land cover classification system used for the EOSD product is based on the classification system originally developed for the National Forest Inventory (NFI) program. The six levels of the NFI classification hierarchy include vegetation cover, tree cover, landscape position, vegetation type, forest density, and species diversity (Figure 2; Gillis, 2001). The EOSD classification hierarchy is identical in principle, but is limited to a level of detail discernable in Landsat imagery (Wulder and Nelson, 2003; Wulder et al. 2003). Thus, Levels 3 (landscape position) and 6 (species diversity) are not included. Using the NFI class structure as a base, level 4 (vegetation type), along with level 5 forest density descriptors, the closed legend in Figure 2 emerged. Building upon the existing NFI hierarchy enables classification and class generalization. The generalization or collapsing of classes can be useful for reporting validation results for different levels of classification detail, thereby accommodating a wider range of end-users. The EOSD-NFI legend has been cross-walked to a number of regional, national, and international land cover legends (Wulder and Nelson, 2003) augmenting the utility of the land cover products generated.

3.2 Forest inventory

Existing provincial forest inventory data were available in digital format in a GIS database for approximately 26,000 square kilometres of Vancouver Island; inventory information was not available for areas not subject to inventory activities, including parks and private lands. The inventory is the primary source of information on the distribution and areal extent of forest stands, logging roads, and past natural and human disturbances in the study region. The inventory includes species composition (of up to six species, with estimates of species prevalence to the nearest 10%), stand age in years, crown closure (to the nearest 5%), stand height in metres, diameter at breast height in centimetres, and stand area in hectares (Gillis and Leckie, 1993). Some stands were labelled non-productive forest. Much of the inventory information was projected to represent 1999 forest conditions; the original data collection reference years for individual stands ranged from 1954 to 1999. A method was developed for this study to generate EOSD-equivalent labels from attributes in the forest inventory using various combinations of species and species composition, age, crown closure, height, and non-productive forest codes. As a result, each forest inventory polygon was assigned an EOSD land cover label.

3.3 Airborne video

The equipment used to collect the video data consisted of a Sony™ DCR-PC330 digital camcorder and a Red Hen A-VMS 300 device. The Red Hen (Red Hen Systems, Fort Collins, Colorado, USA) device included a Thales™ B12 GPS engine (a real-time differential GPS (DGPS) receiver), which is horizontally accurate to within 3 metres, and is used to encode/decode GPS positional information onto the audio track of the video recording. The Sony™ DCR-PC330 is a MiniDV format camcorder that is compact and lightweight, with 530 lines of resolution and a single Advanced HAD™ 1/3" CCD with 3310K total pixels. The effective video resolution is 2048K pixels. The video camera and Red Hen device were mounted in a Cessna 206 aircraft, which had a 12-inch photo port on its underside. Due to the highly variable topography on Vancouver Island, and associated difficulties in maintaining a consistent flying height, the scale of the video ranged from 1:500 to 1:30000, with pixel sizes on the ground ranging from 1 cm to 60 cm and frame extents from 2.4 by 1.8 m to 144 by 108 m. Four flight lines, extending the full length of Vancouver Island, were flown over a two day period on August 10th and 11th 2004, resulting in approximately 11 hours of continuous video (Figure 1), representing approximately 0.5% of the land area of Vancouver Island. The flight lines were designed to capture the greatest range of land cover classes possible. Ancillary data such as biogeoclimatic zones and base planimetric layers were compiled and referenced when the flight plan was being developed. The average flying height above the terrain, was 585 metres; however, flying heights varied in areas of steep topography. Post-flight, the video and GPS information were integrated to create a completely digital and fully georeferenced record of the flight lines using the Red Hen MediaMapper® software.

4.0 METHODS

4.1 Sampling design

A two-phase sampling design was used, with the first phase being the collection of the airborne video, followed by post-stratification of the video samples by the EOSD land cover strata. The sample frame for the study was the contiguous landmass of Vancouver Island (Figure 1). At the time the video acquisition was being planned, the EOSD product for Vancouver Island was not complete, and the spatial distribution of the EOSD land cover classes was not known. Given limited resources for flying and video acquisition, a surrogate source of ecological mapping was used to ensure the flight lines were ecologically representative and that sufficient numbers of samples could be collected for all possible land cover classes. To achieve this, biogeoclimatic ecosystem zones and subzones, combined with a digital elevation model, were used to guide the placement of flight lines (Figure 3). Biogeoclimatic ecosystem subzones represent unique sequences of geographically related ecosystems, which are grouped into broader biogeoclimatic zones, of which there are four unique zones on Vancouver Island (Meidinger and Pojar, 1991). The table included in Figure 3 indicates that the flight lines characterized the spatial distribution of biogeoclimatic zone/subzone combinations. Although the flight lines were not selected via a probability sampling design, the resulting sample is assumed to be representative of the land cover distribution of Vancouver Island. Individual video frames were extracted at 15 second intervals, resulting in a total of 2651 possible frames, from all of the flight lines. Each video frame was assigned a unique identifier. Points representing the centroid locations of these video frames were overlaid with the EOSD product to derive the corresponding EOSD land cover class for the video frame in order to facilitate proportional allocation of the sample units to the EOSD land cover strata. The individual video frames were then pooled (i.e., no longer associated with their source transect) and sorted randomly.

4.1.1 Sample size

The selection of sample size often involves tradeoffs between the requirements of statistical rigour and logistical realities. The objective of this study was to develop a protocol for the use of video for validation, and one of the key issues requiring exploration was the necessary number of interpreters to assign land cover attributes to the video frames. The assumption was that the greater the number of interpreters, the less subjective the interpretation was likely to be (since the mode of all seven interpreters would be compared to the EOSD product label). Seven interpreters familiar with EOSD and vegetation on Vancouver Island were selected. Sufficient resources were not available for the seven interpreters to label all 2651 video frames that were collected, and therefore a subset of video frames had to be selected. The sample had to be large enough to facilitate statistical testing and provide confidence in the overall accuracy estimate, while at the same time ensuring each land cover class had sufficient samples for reporting user's and producer's accuracies (Czaplewski and Patterson, 2003).

An overall sample size of approximately 500 was determined by specifying a 90% confidence interval for p with a margin of error of 0.03, and an assumption of 80% true accuracy (Cochran, 1977):

$$n = \left[\left(\frac{z}{m} \right)^2 \right] \times p \times (1 - p) \tag{1}$$

where z is the percentile of the standard normal (1.65 for 90% confidence interval), m is the margin of error (0.03), and p is the assumed population proportion (0.8).

We allocated half of the 500 samples proportional to the area of the land cover stratum, and the remaining half were allocated to improve estimates for rare classes, as per Czaplewski (2003):

$$n_i = [\, p_i * (n/2) + (1/k) * (n/2)] \tag{2}$$

Where:
n_i = sample size allocated to mapped stratum i
p_i = proportion of sample population mapped as i
n = total sample size
k = # of map categories

Under a strictly proportional allocation scheme, several classes present in the EOSD product would have had no samples allocated to them. Recall that the video frames were previously tagged with their associated land cover strata, and the number of samples required for each land cover stratum, as listed in Table 1, was then selected at random from the 2651 available video frames.

4.2 Response design

4.2.1 Evaluation potocol

The evaluation protocol is the procedure used to collect the reference information (Stehman and Czaplewski, 1998). In defining the evaluation protocol, the spatial support region (SSR) is chosen. The SSR is defined as "the size, geometry, and orientation of the space on which an observation is defined" (Atkinson and Curran, 1995: 768). Given the variability in flying heights, the area of the instantaneous field of view of each video frame varied. Therefore, it was not practical to specify an SSR as a fixed areal unit. The centroid of the frame, whose location was captured by the GPS unit during flight, was selected as an easily understood reference point for all interpreters. The centre of the video frame, determined by the collected GPS UTM easting and northing, was assigned a land cover class by the interpreters.

4.2.2 Video atribution

Interpreters were given a land cover class key which provided samples of the video for each of the EOSD land cover classes. The key was intended to improve consistency between interpreters, who manually assigned a land cover label to the centroid of the selected video samples. Interpreters were selected based on their familiarity with the EOSD product and the EOSD classification scheme, their understanding of forest inventory, and their experience in interpreting aerial photography. The interpreters had differing levels of field experience on Vancouver Island; including two of the interpreters who were involved in the acquisition of the video and therefore had some additional understanding of the spatial distribution and visual representation of the land cover classes. Strand *et al.* (2002) demonstrated that direct field experience did not improve the accuracy with which photo interpreters could label land cover types; however, Drake (1996) demonstrated that even a minimum amount of training can improve the interpretation of vegetation classes from airborne video.

The key points communicated to the interpreters included (Wulder *et al.*, 2004):

- Tall and low shrubs would be grouped into a single shrub class.
- Density classes, particularly the distinction between dense and open classes, are known to be problematic. Interpreters were advised when not certain of the density class to select both a primary and secondary class label (*e.g.* primary label: dense conifer, secondary label: open conifer).
- Guidelines for defining mixed wood stands were provided (*i.e.*, when neither coniferous nor broadleaf species account for more than 75% of the total basal area in the stand).
- Several examples of video frames were provided for each land cover class to promote consistent interpretations.

4.2.3 Labelling protocol

The interpreters were provided with a listing of the sample video frames. Recall that the sample video frames had been pooled and sorted randomly by their unique identifier. The interpreters did not know the geographic location of the video frame, or the flight line that the video frame originated from. The interpreters were also not privy to the EOSD land cover stratum from which the video frame was selected. Each interpreter independently determined an appropriate label for the land cover type existing at the centre of the video frame. As per Stehman *et al.* (2003), the interpreters selected the most likely class (primary choice) with an option to specify a second choice if necessary. The secondary choice captures the confusion between classes when the frame falls in the transition between two cover-type classes (Figure 4), thereby acknowledging thematic and non-thematic errors (Foody, 2002).

4.2.4 Defining agreement

Several scenarios for defining agreement between the EOSD product and the interpreted video frames were explored:

1. If the land cover class of the EOSD product matched the primary land cover choice of the interpreted video.
2. If the land cover class of the EOSD product matched either the primary or secondary land cover choice of the interpreted video.
3. If the modal class of a 3 by 3 pixel SSR around the target EOSD pixel matched either the primary or secondary land cover choice of the interpreted video.

The first scenario is a direct comparison between the interpreted video and the EOSD product, making no allowances for possible errors in attribution and/or positional accuracy. The second scenario listed above accommodates thematic ambiguity, while the third choice accommodates both thematic ambiguity and positional uncertainty (Stehman *et al.*, 2003). Under the third scenario, if more than one modal class existed, the sample was dropped. For the scenario where the pre-existing forest inventory data is used for validation, the land cover of the EOSD pixel is compared directly to the label of the inventory polygon within which the EOSD pixel falls (see Wulder et al. 2006b for details).

4.3 Analysis

The EOSD land cover class had previously been extracted for each of the sample video frames, based on the GPS position of the video frame centroid. Similarly, the mode land cover class of a 3 by 3 pixel neighbourhood surrounding the centroid pixel was generated and extracted for the centroid of each video frame. The EOSD classifications for each of the samples were then compared to the interpreted video samples using a confusion matrix and measures of accuracy were estimated. Since post-stratification was used to allocate the total number of samples proportional to the area of each of the land cover stratum, estimation methods provided in Cochran (1977) and Czaplewski (2003) were used to calculate accuracy measures and associated confidence intervals.

The hierarchical nature of the EOSD land cover classification system supports generalization to higher levels. Many applications do not require information on vegetation density, using only information on vegetation type. In this situation, reporting accuracy results, which may be negatively impacted due to confusion associated with the interpretation of density classes, does not provide a fair representation of the accuracy of the level 4 product. For the sake of transparency and comparison in this study, accuracy values are reported for both levels 4 and 5.

5.0 RESULTS

5.1 Accuracy assessment

The first validation scenario defined agreement where the land cover class of the EOSD product matched the modal primary land cover choice of the interpreters (Table 2). Results are reported for level 5 and level 4, which is cover type only (Figure 2). Accuracy estimates were obtained by using the modal class of all seven interpreters. The level of overall accuracy for level 5 was 53%, with user accuracies for the three coniferous density classes ranging from 9% to 72%. At level 4 the overall accuracy was 71%, with the coniferous class at 84%. The second scenario defined agreement where the land cover class of the EOSD product matched either the modal primary or secondary land cover choice of the interpreted video. For this scenario, the overall accuracy was 61% (level 5) and 73% (level 4). User's accuracy for the three coniferous density classes ranged from 23% to 76%, with user's accuracy for level 4 coniferous at 85% (Table 2). The third scenario defined agreement between the modal class in a 3 by 3 pixel window and either the modal primary or secondary land cover choice of the interpreted video (Table 2). The overall accuracy for this scenario was 67% for level 5 and 76% for level 4. User's accuracies for the three coniferous density classes ranged from 15% to 82%, and the user's accuracy for the coniferous class was 86%. With pre-existing forest inventory data used as reference data the overall accuracy was 41% for level 5 and 68% for level 4. User's accuracies for the coniferous density classes ranged from 0% to 55% and the user's accuracy for the coniferous class was 74%.

5.2 Multiple interpreters

A summary of overall accuracy, as estimated by each interpreter for each scenario, is provided in Table 3. These results provide an indication of what the accuracy estimates would have been if only one interpreter had been used for the project. Overall accuracies varied by an average of 8% for level 5 and 11% for level 4. Pairwise comparisons for each scenario were made between interpreters, and between interpreters and the EOSD product, and the proportion of agreement is summarized in Table 4.

6.0 DISCUSSION

6.1 Use of purpose-acquired video for validation

There are a number of issues to consider when determining the appropriateness of video data for an accuracy assessment. Video can cover large geographic areas efficiently, and areas that are identified in the flight plan as unique or rare can be sampled intensively. When compared to field surveys, video facilitates the collection of a large number of sampling locations, providing sufficient samples for both calibration and validation, and the data redundancy and ability to view transitions between land cover classes can be advantageous. Video also provides a permanent record of the survey and interpretation may be done by highly trained professionals, mobilized at minimal cost, who have extensive experience in identifying vegetation types and land cover from airborne imagery. Technology has created advances in digital video that have resulted in higher quality, compact cameras at increasingly affordable prices, and since there is little post-processing required, the data can be available almost immediately after collection. Furthermore, sophisticated and readily available off-the-shelf software and hardware products can generate video that is fully georeferenced, facilitating sampling protocols. Digital photography has experienced similar advances in technology and reductions in cost. Video acquisition must be planned so that the video meets the level of detail required by the classification system, and flying height and speed must be set accordingly. The efficiency, simplicity, and cost effectiveness of video make it an attractive source of calibration and validation data for large area land cover products generated from remotely sensed data.

The results of the accuracy assessment using the video frames for validation highlighted several issues. First, accommodating positional and thematic ambiguity can have an impact on the reported accuracy results (75% overall accuracy when these sources of ambiguity are addressed versus 53% when they are not). Second, density classes were a major source of classification confusion and result in low accuracy measures at level 5 of the EOSD classification hierarchy. This may misrepresent the accuracy of the product to end users who do not require information on vegetation density. Therefore, separate accuracy estimates should be reported for both level 4 and level 5 land cover products, allowing an end user to make informed decisions about the application-specific relevance of each estimate. End users must also understand the limitations of the agreement scenarios, and of accuracies reported for a generalized product (e.g., the mode of 3 by 3 pixel neighbourhood) (Czaplewski, 2003). Confusion between density classes also suggests that greater effort is required to calibrate interpreters. Crown closure has been identified as one of the more difficult forest inventory attributes to estimate or even measure accurately (on the ground, from the air, or from satellite imagery) (Congalton and Biging, 1992; Gill *et al.* 2000). The potential of extracting estimates of crown closure from Landsat imagery has been the subject of ongoing research (e.g., Franklin *et al*., 2003; Xu *et al.,* 2003; Pu *et al.,* 2005; Joshi *et al.,* 2006).

6.2 Pre-existing forest inventory

The use of the pre-existing forest inventory as a source of validation data produced overall accuracy results that were lower than the results produced using the purpose acquired video for validation. For level 5, using the forest inventory data produced an estimate of overall accuracy that ranged from 12% to 26% lower than that produced using the video. Individual user's accuracies for the coniferous open class were similar, ranging from 17% to 27% lower than the estimates associated with the video. At level 4, the differences between the inventory and the video estimates of overall accuracy are not as disparate, with the inventory estimating overall accuracy ranging from 3% to 9% lower than the video estimates, and user's accuracies for the coniferous class that were 10% to 12% lower than that estimated from the video. Wulder et al. (2006b) document the challenges of using pre-existing data (specifically forest inventory) to validate land cover products generated from remotely sensed data: geolocational mismatches, differences in features or classes mapped, disparity between the scale of polygon delineation and the spatial resolution of the image, and temporal discrepancies. However, the greatest challenge to using pre-existing data such as forest inventory for validation is that a forest inventory polygon is a generalization, designed to represent areas of relatively homogenous forest characteristics. In contrast, the raster based land cover product can contain much more detail and is, by its very nature, heterogeneous, with a minimum mapping unit (MMU) that often reflects the size of the image pixels (e.g., 0.06 ha) versus the 2 ha MMU of the forest inventory.

6.3 Multiple interpreters

All of the seven interpreters' primary labels agreed only 20% of the time and of these, 36% were classed as not interpretable; 26% were open coniferous; and 22% were snow and ice. This low level of agreement amongst the interpreters' primary label suggests there was either ambiguity in the application of the classification legend or confusion regarding the sample unit (i.e., the dominant land cover class for the frame as a whole versus just the cover type located at the centroid of the chip). In 80% of the cases where the interpreters' primary labels did not agree, 66% were attributed to disagreements over density classes. Reporting accuracy estimates for level 4 of the EOSD hierarchy eliminated this type of disagreement and increased interpreter agreement to 43% (compared to 20% at level 5).

One of the drawbacks to using aerial photography or video for validation is the lack of direct physical contact with the target (Congalton and Biging, 1992), therefore the classification is still based on remote interpretation (albeit with a well established and rigorous process). As such, there is no "truth", as multiple interpreters are unlikely to agree 100% of the time. The absence of an absolute truth dataset makes it difficult to assess the performance of any one interpreter, and to determine the optimal number of interpreters required to provide an unbiased result. Powell *et al.,* (2004) manufactured a truth standard by having all five interpreters assign pixels to land cover classes by consensus. The result was termed the "gold standard" and was used for the final accuracy assessment. Powell *et al.,* (2004) attributed some of the disagreement between interpreters to human error (e.g., misinterpretation of the image printout, miscoding the sample), while other disagreements related more fundamentally to identification of the vegetation in the video.

Table 3 summarizes the overall accuracies for each scenario, by interpreter. If any one of these interpreters had individually been selected to interpret the video, the overall estimates of accuracy would have varied by an average of 8% for level 5 and 11% for level 4. This result contrasts with that of Drake (1996) who concluded that individual interpreters could be used to produce similar accuracy estimates, with little effect. Table 4 summarizes the overall proportion of agreement between interpreters and between interpreters and the EOSD product. The results indicate that there was more agreement amongst individual interpreters than there was between individual interpreters and the EOSD product. For level 5, agreement between primary labels of any two interpreters ranged by 20%, with an average overall agreement of 60%. If both the primary or secondary label are considered, agreement between any two interpreters ranged by 23%, with an average overall agreement of 77% and the highest agreement at 90%. Comparisons between individual interpreters and the EOSD output were markedly lower. If only the primary label of a single interpreter is considered, then overall agreement averaged 40% (range 8%). If both primary and secondary label are compared to the modal land cover class of a 3 by 3 pixel neighbourhood, overall agreement increased to 45% (range 9%). For level 4, the patterns in accuracy estimates are similar. Agreement amongst interpreters was greater at more generalized levels of the EOSD hierarchy (70% at level 4 versus 60% at level 5), and if thematic ambiguity was accounted for (77% for both primary and secondary labels at level 5 versus 60% for primary label only at level 5).

These results suggest that the use of multiple interpreters can be important for reducing bias and improving consistency in class labeling. The number of interpreters required likely varies by application and depends on the complexity of the land cover classes. For the method whereby the modal class of interpreters is used, a minimum of three interpreters would be practical, and is recommended. An evaluation protocol that incorporates independent classification by each interpreter, followed by cross calibration, and revisit of problematic classes would be the most effective way to use fewer interpreters (and fewer resources), while still taking advantage of the benefits of multiple interpreters.

6.4 Lessons learned

The main objective of this study was to develop a protocol and demonstrate the use of airborne video as a source of validation data for large area land cover products, specifically the EOSD. Over the course of conducting this operational trial, several issues arose, providing opportunities to improve upon the methodology for future implementation. First, a fully probabilistic sample design is clearly more important than the representation of all potential classes – particularly since resources for the accuracy assessment are scarce and a key objective of the EOSD accuracy assessment is to generate robust estimates of overall accuracy and producer's and user's accuracies for the dominant forest classes. To that end, flight lines should be placed so that they traverse the study area at fixed intervals of distance in both north-south and east-west orientations. Samples should be allocated proportional to strata area, with no accounting for rare classes that have a very limited spatial extent. The video camera should be set to a fixed zoom, facilitating an increase in the consistency of the scale of video frames. In addition, the pilot should be instructed to try to maintain a consistent altitude above ground level (rather than focussing on maintaining a consistent altitude). An attitude and heading reference system device (digital gyroscope) is critical to ensure that the accurate GPS locations recorded by the Red Hen system can be used to their full potential. With a known pitch, roll, and heading, combined with the GPS position of the plane, it is relatively simple to calculate the coordinates of the center of the image on the ground. Finally, with over 60% of the disagreement between interpreters caused by differences in density estimates, greater effort must be made to calibrate interpreters and improve consistency in estimation of density classes. Drake (1996) demonstrated the utility of training in improving the interpretation accuracy of vegetation classes from airborne video.

7.0 CONCLUSION

This operational trial explored the use of purpose-acquired airborne video data for validation of an EOSD land cover product. Results indicate that airborne video is an efficient and cost-effective medium for validation of large area land cover products generated from remotely sensed data. The results also confirm that accuracy estimates generated using purpose-acquired data will be different from estimates generated using pre-existing forest inventory data as a reference source. In this study, the use of forest inventory for validation would have resulted in an underestimation of EOSD product accuracy. Furthermore, pre-existing data may not be the most appropriate data source for validation, and may be unavailable in remote or inaccessible areas. Airborne video provides a viable validation data source; the hardware, software, and expertise required to acquire the video data, generate a fully geo-referenced data set, select a representative sample, and interpret land cover classes is widely available and affordable.

A full implementation of the protocol developed here requires a carefully selected sampling strategy that successfully balances the requirements of statistical rigour while at the same time accommodating logistical realities and fulfilling the primary objectives of the accuracy assessment. Multiple interpreters were found to be an effective means to improve consistency and reduce bias in the assignment of land cover classes to the video samples, although a minimum of three interpreters would likely be sufficient. Substantial effort should be directed at calibrating interpreters and improving consistency in labelling. Finally, the importance of reporting accuracy estimates and the details of the sample design in a transparent manner, for both levels 4 and 5 of the EOSD classification hierarchy, will provide the end user with the information required to assess whether the product is sufficiently accurate for a specific application. The lessons learned from this trial will contribute to a robust assessment of EOSD product accuracy.

ACKNOWLEDGEMENTS

This research is enabled through funding of the Government Related Initiatives Program of the Canadian Space Plan of the Canadian Space Agency and Natural Resources Canada – Canadian Forest Service. David Seemann, of the Canadian Forest Service, and Jamie Heath, of Terrasaurus Aerial Photography, are thanked for their assistance with collection of the video data. Paul Boudewyn, Morgan Cranny, Mike Heneley, David Seemann, and David Tammadge are thanked for their assistance with the interpretation of the video samples.

REFERENCES

Atkinson, P. & Curran, P.J. (1995). Defining an optimal size of support for remote sensing investigations. *IEEE Transactions on Geoscience and Remote Sensing, 33,* 768-776.

Bobbe, T., Reed, D., & Schramek, J. (1993). Georeferenced Airborne Video Imagery: Natural resource applications on the Tongass. *Journal of Forestry, 91*, 34-37.

Brownlie, R.K., Firth, J.G., & Hosking, G. (1996). Potential of airborne video for updating forest maps. *New Zealand Forestry*, *41*, 37-40.

Cihlar, J. (2000). Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing*, *21*, 1093-1114.

Cihlar, J. & Beaubien, J. (1998). *Land cover of Canada Version 1.1*. Special Publication, NBIOME Project. Produced by the Canada Centre for Remote Sensing and the Canadian Forest Service, Natural Resources Canada. Ottawa, Ontario, Canada.

Cochran, W.G. (1977). *Sampling Techniques*, *3rd Edition*. John Wiley and Sons: New York. 428 p.

Congalton, R.G. & Biging, G.S. (1992). A pilot study evaluating ground reference data collection efforts for use in forest inventory. *Photogrammetric Engineering and Remote Sensing, 58,* 1669-1671.

Congalton, R.G., & Mead, R.A. (1983). A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing, 49,* 69-74.

Czaplewski, R.L. (2003). Statistical design and methodological considerations for the accuracy assessment of maps of forest condition. In M.A. Wulder & S.E. Franklin (Eds.), *Remote sensing of forest environments: Concepts and case studies* (p. 115-141). Boston MA: Kluwer Academic Publishers. 519 p.

Czaplewski, R.L. & Patterson, P.L. (2003). Classification accuracy for stratification with remotely sensed data. *Forest Science, 49,* 402-408.

Davis, T.J., Klinkenberg, B., & Keller, C.P. (2002). Updating inventory using oblique videogrammetry and data fusion. *Journal of Forestry*, *2*, 45-50.

Drake, S. (1996). Visual interpretation of vegetation classes from airborne videography: an evaluation of observer proficiency with minimal training. *Photogrammetry and Remote Sensing, 62,* 969-978.

Driese, K.L., Reiners, W.A., Merrill, E.H., & Gerow, K.G. (1997). A digital land cover map of Wyoming, USA: a tool for vegetation analysis. *Journal of Vegetation Science, 8,* 133-146.

Foody, G.M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment, 80*, 185-201.

Franklin, S.E., Hall, R.J., Smith, L., & Gerylo, G.R. 2003. Discrimination of conifer height, age, and crown closure classes using Landsat-5 TM imagery in the Canadian Northwest Territories. *International Journal of Remote Sensing*, *24*, 1823-1834.

Frazier, P. (1998). Mapping blackberry thickets in the Kosciuszko National Park using airborne video data. *Plant Protection Quarterly, 13*, 145-148.

Fuller, R., Groom, G., & Jones, A. (1994). The land cover map of Great Britain: An automated classification of Landsat Thematic Mapper data. *Photogrammetric Engineering and Remote Sensing*, *60*, 553-562.

Gill, S.J., Milliken, J., Beardsley, D., & Warbington, R. (2000). Using a mensuration approach with FIA vegetation plot data to assess the accuracy of tree size and crown closure classes in a vegetation map of Northeastern California. *Remote Sensing of Environment*, *73*, 298-306.

Gillis, M.D. (2001). Canada's National Forest Inventory (responding to current information needs). *Environmental Monitoring and Assessment, 67*, 121-129.

Gillis, M.D. & Leckie, D.G. (1993). Forest inventory mapping procedures across Canada. Information Report PI–X–114, Petawawa National Forestry Institute, Forestry Canada.

Goodchild, M.F., Biging, G.S., Congalton, R.G., Langley, P.G., Chrisman, N.R. & Davis, F.W. (1994). *Final Report of the Accuracy Assessment Task Force*, California Assembly Bill AB1580. National Center for Geographic Information and Analysis, UCSB.

Graham, Lee A. (1993). Airborne video for near-real-time vegetation mapping, *Journal of Forestry, 8*, 28-32.

Hansen, M.C., Defries, R.S., Townshend, J.R.G. & Sohlberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach, *International Journal of Remote Sensing*, *21*, 1331-1364.

Hepinstall, J. A., Sader, S. A., Krohn, W. B., Boone, R. B. & R.I. Bartlett. (1999). *Development and testing of a vegetation and land cover map of Maine*. Maine Agriculture and Forest Experiment Station Technical Bulletin 173, University of Maine, Orono. 104 p.

Herold, M, Woodcock, C., di Gregorio, A., Mayaux, P., Belward, A., Latham, J., & Schmullius, C.C. 2006. A joint initiative for harmonization and validation of land cover datasets. IEEE Transaction on GeoScience and Remote Sensing, In press.

Hess, L.L., Novo, E.M.L.M., Slaymaker, D.M., Holt, J., Steffen, C., Valeriano, D.M., Mertes, L.A.K., Krug, T., Melack, J.M., Gastil, M., Holmes, C. & Hayward, C. (2002). Geocoded digital videography for validation of land cover mapping in the Amazon basin. *International Journal of Remote Sensing*, *23*, 1527-1556.

Homer, C., Ramsey, R., Edwards Jr., T., & Falconer, A. (1997). Landscape cover-type modeling using a multi-scene thematic mapper mosaic. *Photogrammetric Engineering and Remote Sensing*, *63*, 59-67.

Jack, K. (2005). *Video Demystified, 4^{th} Edition*. Eagle Rock, VA: LLH Technology Publishing. 733 p.

Jacobs, D.M. (2000). *February 1994 Ice Storm: Forest Resource Damage Assessment in Northern Mississippi*. United States Department of Agriculture, Forest Service, Southern Research Station, Resource Bulletin SRS-54. Asheville, NC. 11 p.

Jacobs, D.M. & Eggen-McIntosh, S. (1993). Airborne videography and GPS for assessment of forest damage in souther Louisiana from Hurricane Andrew. In: *Proceedings of the IUFRO conference on inventory and management techniques in the context of catastrophic events*; 1993 June 21-24; University Park, PA. 12 p.

Joshi, C., DeLeeuw, J., Skidmore, A.K., van Duren, I.C., & van Oosten, H. (2006). Remotely sensed estimation of forest canopy density: A comparison of the performance of four methods. *International Journal of Applied Earth Observation and Geoinformation, 8,* 84-95.

Justice, C., Belward, A., Morisette, J., Lewis, P., Privette, J., & Baret, F. (2000). Developments in the 'validation' of satellite sensor products for the study of the land surface. *International Journal of Remote Sensing*, *21,* 3383-3390.

King, D.J. (1995). Airborne multispectral digital camera and video sensors: a critical review of system designs and applications. *Canadian Journal of Remote Sensing, 21*, 245-273.

Klinka, K., Pojar, J., & Meidinger, D.V. (1991). Revision of biogeoclimatic units of coastal British Columbia. *Northwest Science, 65*, 32–47.

Loveland, T.R., Merchant, J.W., Ohlen, D.O., & Brown, J.F. (1991). Development of a land-cover characteristics database for the conterminous U.S. *Photogrammetric Engineering and Remote Sensing*, *57*, 1453-1463.

Loveland, T.R. & Belward, A.S. (1997). The International Geosphere Biosphere Programme Data and Information System global land cover data set (DISCover), *Acta Astronautica*, *41*, 681-689.

Loveland, T. R., Zhu, Z., Ohlen, D. O., Brown, J. F., Reed, B. C., & Yang, L. (1999). An analysis of the IGBP global land-cover characterisation process. *Photogrammetric Engineering and Remote Sensing, 65*, 1021– 1032.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., & Merchant, J. W. (2000). Development of a global land cover characteristic database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, *21*, 1303-1330.

Maingi, J.K., Marsh, S.E., Kepner, W.G. & Edmonds, C.M. (2002). *An accuracy assessment of 1992 Landsat-MSS derived land cover for the Upper San Pedro Watershed (US/Mexico)*. United States Environmental Protection Agency, EPA/600/R-02/040. 21 p.

Marsh, S.E., Walsh, J.L. & Sobrevila, C. (1994). Evaluation of airborne video data for land-cover classification accuracy assessment in an isolated Brazilian forest. *Remote Sensing of Environment, 48*, 61-69.

Mausel, P.W., Everitt, J.H., Escobar, E.E. & King, D.J. (1992). "Airborne Videography: Current Status and Future Perspectives." *Photogrammetric Engineering and Remote Sensing*, *58*, 1189-1195.

Meisner, D.E. (1986). Fundamentals of airborne video remote sensing, *Remote Sensing of Environment, 19*, 63-79.

Meidinger, D. & Pojar, J. (Eds.) 1991. *Ecosystems of British Columbia*. Research Branch, Ministry of Forests, Victoria, British Columbia, Canada, 330 p.

Morisette, J.T., Privette, J.L. and Justice, C.O. (2002). A framework for the validation of MODIS land products. *Remote Sensing of Environment*, *83*, 77-96.

Nixon, P., Escobar, D., and Menges, R. (1985). Use of multiband video system for quick assessment of vegetation condition and discrimination of plant species. *Remote Sensing of Environment, 17*, 203-208.

Norton-Griffiths, M. (1982). Aerial point sampling for land use surveys. *Journal of Biogeography, 15,* 149-156.

Powell, R.L., Matzke, N., de Souza Jr., C., Clark, M., Numata, I., Hess, L.L. & Roberts, D.A. (2004). Sources of error in accuracy assessment of thematic land-cover maps in the Brazillian Amazon, *Remote Sensing of Environment, 90*, 221-234.

Pu, R., Yu, Q., Gong, P., & Biging, G.S. (2005). EO-1 Hyperion, ALI, and Landsat ETM+ data comparison for estimating forest crown closure and leaf area index. *International Journal of Remote Sensing, 26,* 457-474.

Remmel, T.K., Csillag, F. Mitchell, S. and Wulder, M. (2005). Integration of forest inventory and satellite imagery: A Canadian status assessment and research issues, *Forest Ecology and Management, 207,* 405-428.

Reiners, W.A., Driese, K.L., & Schrupp, D. (2000). Statistical evaluation of the Wyoming and Colorado landcover map thematic accuracy using aerial videography techniques: Final Report. Department of Botany, University of Wyoming, Laramie, Wyoming. Available online: http://ndis1.nrel.colostate.edu/cogap/report/colandcov_acc.pdf

Schlagel, J. (1995). Integrating aerial-videography with ArcInfo for satellite image interpretation and accuracy assessment in Vermont. *Proceedings of the 15th Annual ESRI User Conference*, May 22-26, 1995, Palm Springs, California.

Scott, J.M., Davis, F., Csuti, B., Noss, Reed, Butterfield, B., Groves, C., Anderson, H., Caicco, S., D'Erchia, F., Edwards, T.C., Ulliman, J., & Wright, R.G. (1993). Gap Analysis: A geographic approach to protection of biological diversity. *Journal of Wildlife Management, 57*, Supplement: Wildlife Monographs No. 123.

Skirvin, S.M., Drake, S.E., Maingi, J.K., Marsh, S.E., & Kepner, W.G. (2000). *An accuracy assessment of 1997 Landsat Thematic Mapper derived land cover for the Upper San Pedro Watershed (U.S./Mexico)*. EPA/600/R-00/097.Office of Research and Development, Washington, DC.

Skirvin, S.M., Kepner, W.G., Marsh, S.E., Drake, S.E., Maingi, J.K., Edmonds, C.M., Watts, C.J., & Williams, D.R. (2004). Assessing the accuracy of satellite-derived land-cover classification using historical aerial photography, digital orthophoto quadrangles, and airborne video data. In R.S. Lunetta & J.G. Lyon (Eds.) *Remote Sensing and GIS Accuracy Assessment* (p. 115-131). Boca Raton FL: CRC Press.

Slaymaker, D.M. (2003). Using georeferenced large-scale aerial videography as a surrogate for ground validation data. In M.A. Wulder & S.E. Franklin (Eds.), *Remote sensing of forest environments: Concepts and case studies* (p. 469-489). Boston MA: Kluwer Academic Publishers. 519 p.

Slaymaker, D.M., Jones, K.M.L., Griffin, C.R. & Finn, J.T. (1996). Mapping deciduous forests in southern New England using aerial videography and hyperclustered multi-temporal Landsat TM imagery. In J.M. Planning, T.H. Scott, & F.W. Davids (Eds.) Gap Analysis: A Landscape Approach to Biodiversity Planning (p. 87-101). Bethesda MD: American Society of Photogrammetry and Remote Sensing.

Stehman, S.V. & Czaplewski, R.L. (1998). Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment, 64*, 331-344.

Stehman, S.V. & Czaplewski, R.L. (2003). Introduction to special issue on map accuracy. *Environmental and Ecological Statistics*, *10*, 301-308.

Stehman, S. V., Wickham, J. D., Smith, J. H., & Yang, L. (2003). Thematic accuracy of the 1992 national land-cover data for the eastern United States: Statistical methodology and regional results. *Remote Sensing of Environment, 86,* 500-516.

Stone, T., Schlesinger, P., Houghton, T. & Woodwell, G. (1994). A map of the vegetation of South America based on satellite imagery. *Photogrammetric Engineering and Remote Sensing*, *60*, 541-551.

Strand, G.H., Dramstad, W. & Engan, G. (2002). The effect of field experience on the accuracy of identifying land cover types in aerial photographs. *International Journal of Applied Earth Observation and Geoinformation, 4*, 137-146.

Strahler, A.H., Boschetti, L., Foody, G.M., Friedl, M.A., Hansen, M.C., Herold, M., Mayaux, P., Morisette, J.T., Stehman, S., Woodcock, C.E. (2006). Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps. Office for the Official Publications of the European Communities, Luxembourg.

Suzuki, R., Hiyama, T., Asanuma, J. & Ohata, T. (2004). Land surface identification near Yakutsk in eastern Siberia using video images taken from a hedgehopping aircraft. *International Journal of Remote Sensing*, *25*, 4015-4028.

Wulder, M.A., Dechka, J.A., Gillis, M.A., Luther, J.E., Hall, R.J., Beaudoin, A., & Franklin, S.E. (2003). Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program. *Forestry Chronicle*, *79*, 1075-1083.

Wulder, M.A., McDonald, S., & White, J.C. (2004). *Protocols for attribution of EOSD classes to video tiles. Example application using Vancouver Island as the population.* Version 1, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, Canada, January 13, 2005, 57 p.

Wulder, M. & Nelson, T. (2003). EOSD *Land Cover Classification Legend Report*, Version 1, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, Canada, March 2003. Available from http://www.pfc.forestry.ca/eosd/cover/EOSD_Legend_Report.pdf . 81 p.

Wulder, M. & Seemann, D. (2001). Spatially partitioning Canada with the Landsat Worldwide Reference System. C*anadian Journal of Remote Sensing, 27*, 225-231.

Wulder, M.A., Franklin, S.E., White, J.C., Linke, J., & Magnussen, S. (2006a). An accuracy assessment framework for large area land cover classification products derived from medium resolution satellite data. *International Journal of Remote Sensing*, *27*: 663-683.

Wulder, M.A., White, J.C., Luther, J.E., Strickland, G., Remmel, T.K., & Mitchell, S. (2006b). Use of vector polygons for the accuracy assessment of pixel-based land cover maps. *Canadian Journal of Remote Sensing, 32*, 268-279.

Xu, B., Gong, P. & Pu, R.L. (2003). Crown closure estimation of oak savannah in a dry season with Landsat TM imagery: comparison of various indices through correlation analysis. *International Journal of Remote Sensing, 24*, 1811-1822.

Table 1. Proportional allocation of first phase samples by EOSD stratum. Half of the samples were allocated proportional to the area of each stratum, while the other half of the samples were allocated to improve estimates for rare classes (see Czaplewski, 2003).

EOSD class	Proportion of sampled population mapped as x	Sample size of x
water	0.0461	25
snow/ice	0.0217	19
rock/rubble	0.0000	14
exposed land	0.0324	22
shrub	0.0420	24
wetland treed	0.0074	16
wetland shrub	0.0016	14
wetland herb	0.0013	14
herb	0.0792	34
coniferous dense	0.0705	32
coniferous open	0.5391	149
coniferous sparse	0.1015	39
broadleaf dense	0.0185	19
broadleaf open	0.0377	23
broadleaf sparse	0.0007	14
mixed wood dense	0.0000	14
mixed wood open	0.0000	14
mixed wood sparse	0.0002	14
TOTAL	**1.0000**	**500**

Table 2. Summary of results for all agreement scenarios with accuracy estimates followed by 90% confidence intervals.

	LEVEL 5		LEVEL 4	
Agreement Scenario	**Overall Accuracy**	**User's Accuracy** *(for coniferous; variable density)*	**Overall Accuracy**	**User's Accuracy** *(for coniferous)*
EOSD and video primary	53% (49% - 57%)	coniferous dense 9% (0.68% - 17%) coniferous open 72% (66% - 78%) coniferous sparse 24% (13% - 35%)	71% (68% - 74%)	coniferous 84% (80% - 88%)
EOSD and video primary or secondary	61% (57% - 65%)	coniferous dense 23% (11% - 35%) coniferous open 76% (70% - 82%) coniferous sparse 33% (21% - 45%)	73% (70% - 76%)	coniferous 85% (81% – 89%)
EOSD 3x3 mode and video primary or secondary	67% (64% - 70%)	coniferous dense 15% (5% - 25%) coniferous open 82% (77% - 87%) coniferous sparse 38% (25% - 51%)	77% (74% - 80%)	coniferous 86% (82% - 90%)
EOSD and forest inventory	41% (37% - 45%)	coniferous dense 18% (7% - 29%) coniferous open 55% (48% - 62%) coniferous sparse 0% (0% - 2.27%)	68% (66% - 71%)	coniferous 74% (69% - 79%)

Table 3. Overall accuracies estimated by each interpreter for three different scenarios of agreement (for levels 4 and 5 of the EOSD classification hierarchy).

	LEVEL 5			LEVEL 4		
interpreter #	**primary only to EOSD label**	**primary or secondary to EOSD label**	**primary or secondary to mode of a 3x3 neighbourhood surrounding EOSD label**	**primary only to EOSD label**	**primary or secondary to EOSD label**	**primary or secondary to mode of a 3x3 neighbourhood surrounding EOSD label**
3	30.17%	36.85%	40.19%	43.97%	46.36%	47.63%
7	30.27%	35.43%	39.90%	46.19%	48.84%	52.24%
5	33.92%	42.57%	46.77%	47.45%	53.68%	54.77%
6	35.55%	41.00%	44.92%	51.66%	52.74%	55.69%
4	31.32%	39.49%	44.13%	50.39%	55.57%	57.39%
1	33.88%	42.56%	48.56%	50.00%	57.36%	59.07%
2	37.80%	42.82%	48.23%	53.83%	58.64%	59.81%

Table 4. Summary of overall interpreter agreement.

		MINIMUM %	MAXIMUM %	MEAN %
Interpreter vs. interpreter (primary label only)	Level 5	51	71	60
	Level 4	61	82	70
Interpreter vs. interpreter (primary or secondary label)	Level 5	67	90	77
	Level 4	70	92	81
Interpreter vs. EOSD (primary label only)	Level 5	30	37	33
	Level 4	44	54	49
Interpreter vs. EOSD (primary or secondary label compared to the mode of a 3x3 neighbourhood around EOSD target pixel)	Level 5	40	49	45
	Level 4	48	60	55

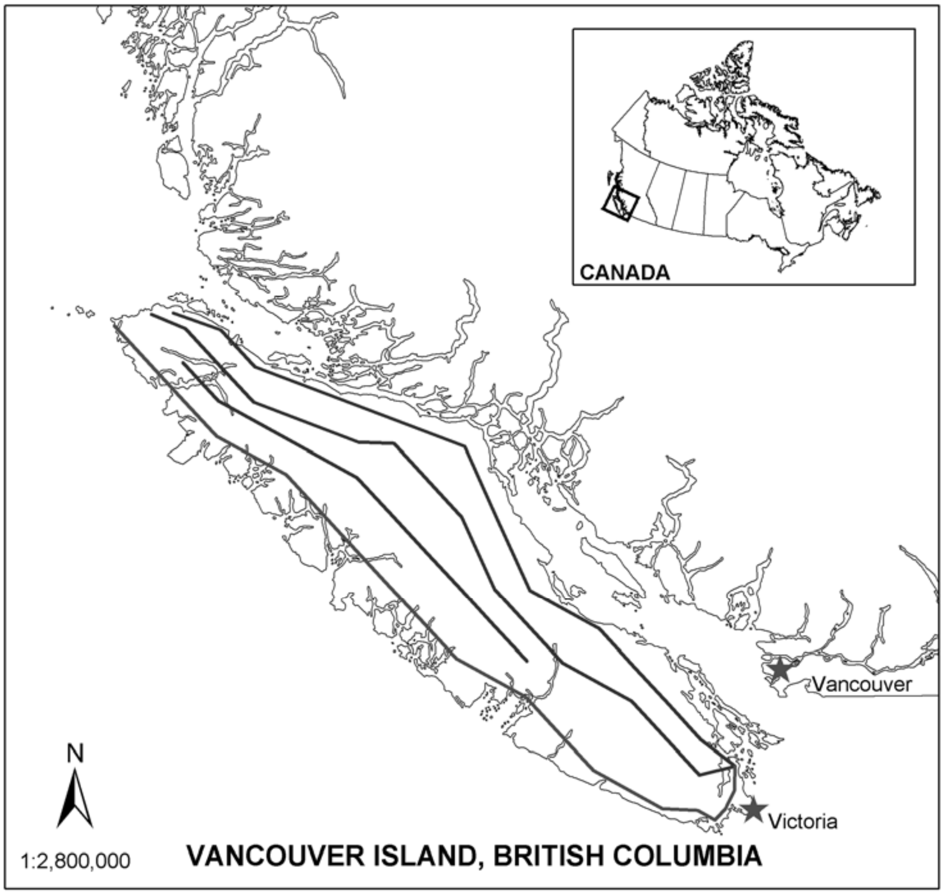

Figure 1. Vancouver Island, British Columbia is located on Canada's western coast. Actual flight lines used for video collection are superimposed.

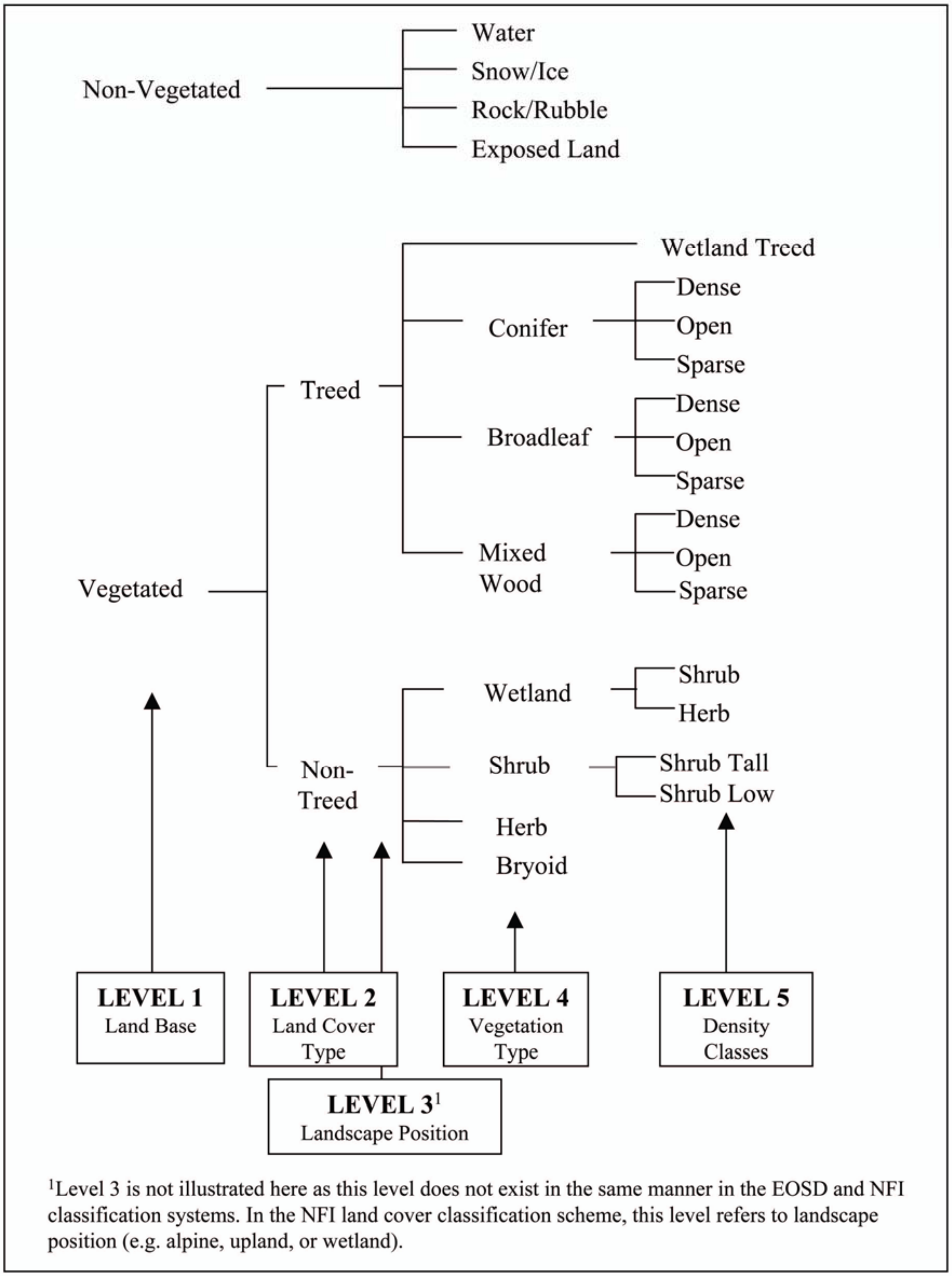

Figure 2. The five levels of the NFI classification hierarchy that correspond to the EOSD classes. The hierarchical nature of the classification scheme enables generalization and reporting at higher levels of the hierarchy.

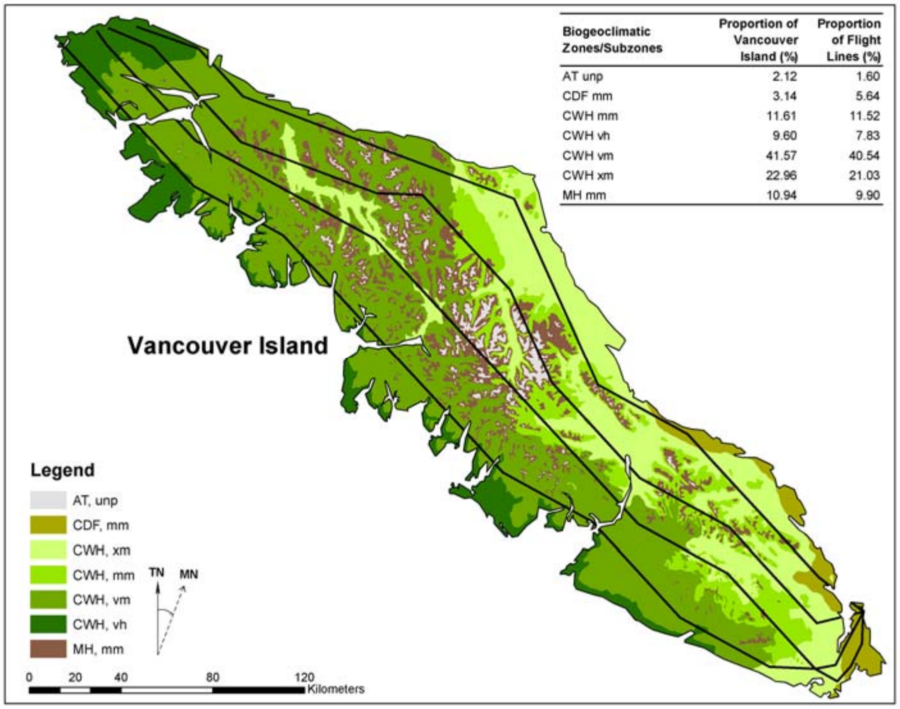

Figure 3. The planned flight lines for video acquisition draped over biogeoclimatic zones/subzones and a digital elevation model. Actual flight lines are shown in Figure 1.

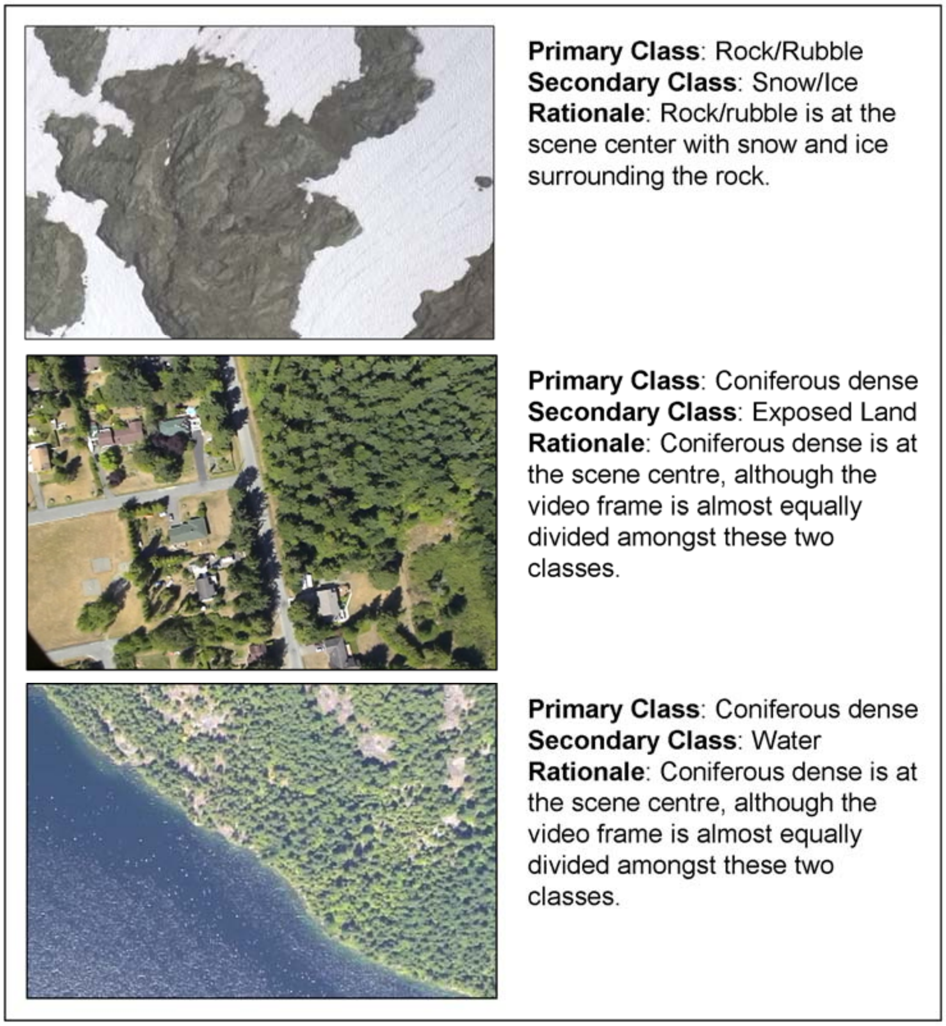

Figure 4. Examples of the need for using primary and secondary labels.