Quantifying error in aerial survey data

Erik W. Johnson[1,2,3] and Jennifer Ross[4]

[1]USDA Forest Service, Rocky Mountain Region, 740 Simms St, Golden, CO 80401, USA
[2]Current address: USDA Forest Service, Alaska Region, 709 W 9th St, Suite 831H, Juneau, AK 99801, USA
[3] Email: ejohnson02@fs.fed.us
[4] USDA Forest Service, Rocky Mountain Region, 740 Simms St, Golden, CO 80401, USA

Revised manuscript received 3 April 2008

Summary

Aerial survey, also referred to as aerial sketchmapping, is the technique of observing symptoms of forest damage from an aircraft and transferring the information manually onto a base map. Recent high levels of bark beetle mortality across the western United States have generated greater demands for, and more disparate uses of, aerial survey data. While aerial survey data are typically considered to be qualitative in nature, the recent application of the data has driven an interest in assessing its spatial and categorical accuracies quantitatively. This paper describes methods for assessing the accuracy of aerial survey data and discusses several implications and applications related to the error results. The error matrix and kappa (κ) statistic, commonly used to assess accuracies of image classifications in remote sensing, were used to describe errors present in the aerial survey data. Field crews collected ground data that were used to validate the aerial classifications on 233 plots across 17.3 million ha. An additional 24 plots were incorporated into the validation from a complementary project, bringing the total number of plots to 257. Errors within the aerial survey data were found to be acceptable for coarse-scale analyses but excessive for use at fine spatial scales. In addition, this paper discusses the benefits of error analysis which include the quantification of errors for reporting purposes, the inclusion of error rates in the metadata, and the ability to focus training programs and technology development by highlighting classes with the highest error. Finally, the cost associated with accuracy assessment implementation is described and weighed against the benefits.

Keywords: aerial surveys; mapping; accuracy; assessment; errors; forest damage; forest management; pests; detection; monitoring; remote sensing

Introduction

Aerial sketchmapping is the technique of observing forest damage caused by insects, disease or other damaging agents from an aircraft and documenting it manually on a map. Forested areas exhibiting damaged or dying foliage are delineated by points or polygons onto a paper map or, more commonly, a computer touch screen. When feasible, the numbers of affected trees per point or polygon are estimated. In the western United States, aerial surveys have historically been undertaken annually in order to locate areas of forest pest activity. Beginning in the late 1990s, aerial surveys have been used to monitor and report on the health of forest ecosystems as part of the national Forest Health Monitoring (FHM) program. The relative accuracy of aerial sketchmap data is therefore of interest in at least two regards.

Classification error is a term used in remote sensing when a pixel, determined by ground observation to be in one category, is assigned to another category during the classification process (Campbell 2002). It is a means of measuring the accuracy of a map by comparing it to reference data. Contrasting observed with expected classification of pixels provides a statistical basis for accuracy assessment.

In the 2005 field season, a pilot study was initiated in the Rocky Mountain Region to assess the accuracy of aerial survey data using the error or contingency matrix approach (Ciesla 2000). Levels of aerial survey accuracies, once derived, could then be incorporated into the aerial survey metadata to assist with data interpretation and use. Metadata describes a data set and often includes information on data definitions and data accuracies. While the application of error matrices to assess the accuracy of aerial survey data is fairly new to the United States, this approach has been successful in southern Brazil for calculating aerial survey accuracies (de Oliveira *et al*. 2006).

The goal of this project was to determine the spatial and classification accuracies of selected categories of aerial survey observations. No attempt was made to assess the accuracy of the mortality estimates commonly attributed during aerial surveys.

Materials and methods

Geographic area

The 2005 aerial detection survey encompassed much of the forested lands within the USDA Forest Service's Rocky Mountain Region except the state of Kansas and central Colorado's Gunnison National Forest and Sopris Ranger District (Fig. 1). In 2005 in the Rocky Mountain Region, about 17.3 million ha were surveyed resulting in 26 735 mapped observations.

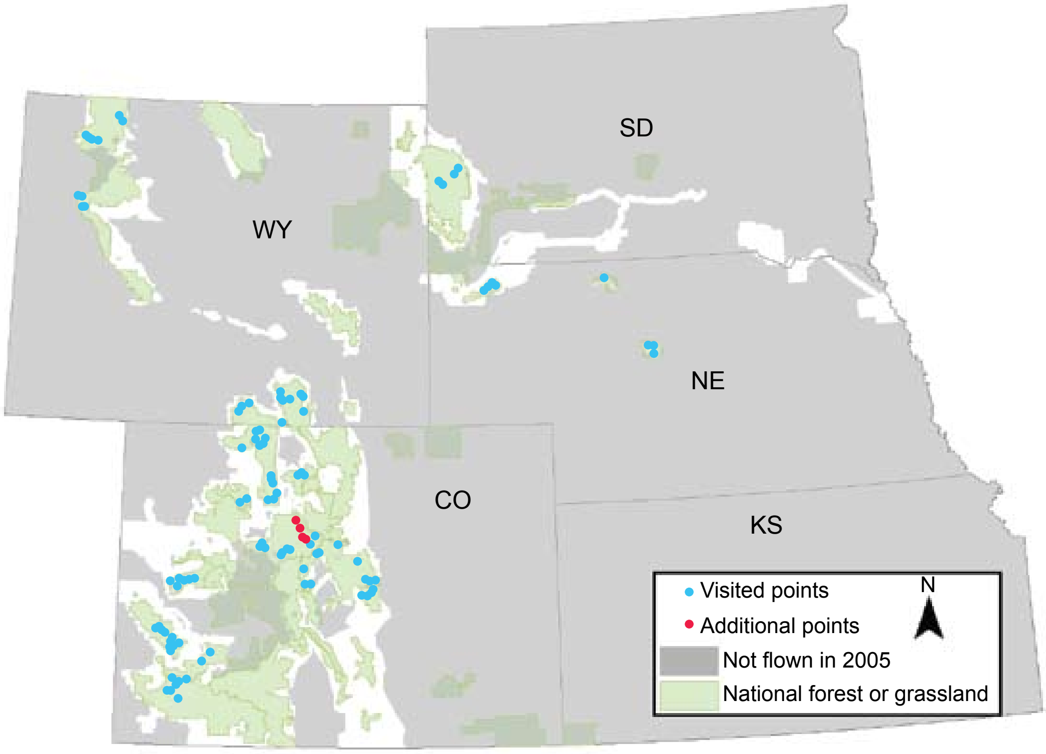

Figure 1. Locations of accuracy assessment plots within the USDA Forest Service's Rocky Mountain Region

Sample design

The authors set a goal for field crews to visit between 20 and 30 ground points per national forest during the 2005 field season. As 15 of the region's 16 national forests were scheduled to be surveyed, it was anticipated that ground information would be obtained from between 300 and 450 sites.

The following eight categories of commonly mapped forest pests or pest complexes were included in the accuracy assessment:

- Douglas-fir beetle in Douglas-fir (DFB)
- Spruce beetle in Engelmann spruce (SB)
- Mountain pine beetle in ponderosa pine (MPB-PP)
- Mountain pine beetle in lodgepole pine (MPB-LPP)
- Mountain pine beetle in limber and or whitebark pine (MPB-WP)
- Subalpine fir mortality, principally a combination of western balsam bark beetles and *Armillaria* sp. root disease (SAF)
- *Ips* sp. beetles in ponderosa and jack pines (IPS)
- No damage (NO DAM).

A two-pronged approach was used to generate sample points in order to eliminate bias by field crew personnel, some of whom also performed the aerial surveys. First, area-weighted-probability random sample points were generated within the aerial survey damage polygons. These points were constrained as follows:

- points had to fall within the defined damage polygons equalling or exceeding aerial estimates of one faded (dying) tree per 0.4 ha
- points had to fall within damage polygons having only one causal agent attribute per polygon
- points had to fall on land administered by the USDA Forest Service
- points had to fall within 2.4 km of a primary or secondary road.

Next, an equal number of secondary points were generated in areas mapped as having 'no damage'. These points were constrained as follows:

- points had to fall within vegetation polygons corresponding to the damage categories being assessed
- points must fall on land administered by the USDA Forest Service
- points had to fall within 2.4 km of a primary or secondary road.

With this design, ground points had an equal probability of falling in polygons of mapped damage and areas of no damage. Field crew personnel had no advance knowledge of the status of these generated points.

Field sample

The selected points were visited by field personnel using Garmin eTrex® GPS units. As one acre (0.4 ha) represents the defined minimum mapping unit of the aerial forest health survey, each GPS-identified point was considered to represent the south-western corner of a square of 0.4 ha on the ground. Field personnel counted and recorded all overstorey trees exhibiting discoloured or 'fading' foliage contained in the 0.4 ha. For every faded tree, the species and associated damage causing agents were recorded.

It was additionally noted whether the fading was recent or old. The following supplementary items were also recorded:

- presence of green infested trees
- overstorey species composition
- understorey species composition.

The crews also noted the presence of fading trees outside of the 0.4 ha plot (to a distance of 500 m), recorded their species and estimated their distance from the plot boundary.

Additional data

To increase the sample size and demonstrate the feasibility of incorporating auxiliary data into the error matrix, a total of 24 additional data points were added to the sample. The data points were obtained from ground surveys established as part of the Dillon Ranger District's Lower Blue and Dillon Reservoir Analysis Areas salvage/sanitation program to reduce tree losses to the mountain pine beetle. Crews collected information on currently infested and year-old beetle-killed mountain pine trees from randomly located, geo-referenced, variable radius ground plots. Plots containing trees killed by mountain pine beetles during the past year were selected for site-specific comparisons with the aerial survey data. These data were included in the damage point pool that was described in the 'Sample design' section above.

Error matrix

A cross-tab query was formulated in Microsoft Access using the ground versus aerial classifications. The cross-tabulated matrix was exported as a Microsoft Excel spreadsheet that was used for the error matrix calculations (Tables 2–4).

The error matrix, containing rows, columns, row marginals, column marginals and diagonals, was used to calculate the following statistics: percentage correct or overall accuracy, commission error (EC), omission error (EO), producer's accuracy (PA), consumer's accuracy (CA) and the kappa (κ) statistic.

Commission errors occurred when an aerial survey observation was mapped and classified into a category that differed from the category found on the ground. Omission errors occurred when a category other than 'no damage' was found on the ground but no observation was recorded on the aerial survey map. Producer's and consumer's accuracies are terms expressly related to map production and are essentially complements of the omission and commission error.

The overall accuracy was calculated by dividing the total number of correct classifications, obtained by summing the diagonal cell values, by the total number of records. CA was calculated by dividing the cell in the diagonal by its corresponding column marginal. EC was calculated by subtracting the CA from 1. PA was calculated by dividing the cell in the diagonal by its corresponding row marginal. EO was calculated by subtracting the PA from 1. CA, PA, EO and EC were all expressed as percentages.

The κ statistic was used to quantitatively assess the error matrix by determining the degree of association between the two set of observations. Specifically, κ measures the difference between the agreement of the ground and aerial survey observations with what would have been attained by a chance assignment of randomly drawn polygons to randomly selected pest categories. The following equation was used to calculate κ:

$$\kappa = \frac{\text{observed} - \text{expected}}{1 - \text{expected}},$$

where *observed* is equal to the overall accuracy described above (dividing the total number of correct classifications by the total number of records) and *expected* is equal to the expected accuracy of chance agreement between aerial and ground classifications (calculated by multiplying row and column marginals for every cell; then dividing the sum of the diagonal products by the sum of all products) (Campbell 2002).

After statistics were calculated for the original data, error matrices were recalculated using tolerances of 50 m and 500 m. Relaxing the spatial tolerance for aerial survey observations acknowledges the spatial error inherent in the process of sketchmapping and allows a more complete consideration of the sketchmapper's intent. For example, if a faded tree was noted to be within 50 m of an 0.4 ha plot void of damage, that plot would change categories from 'no damage' to the appropriate damage category represented by that nearby faded tree. Similarly, if faded trees were found within a ground plot originally classified as having no damage from the aerial survey, but these faded trees were within 50 m of an aerial survey damage polygon, that ground plot's damage category would be changed from 'no damage' to the nearby aerial survey polygon's category in the relaxed tolerance error matrix.

Results

Between 26 July and 10 November 2005 a total of 233 plots were measured by field crews across the Rocky Mountain Region (Fig. 1). Due to time and staffing constraints, random points were collected on only 10 of the intended 15 national forests. Of the 233 random plots visited, 145 were classified by the aerial survey as having no damage and 88 were categorised as exhibiting tree mortality (Fig. 2). The 24 additional plots were incorporated into the error matrix to bring the total number of plots used to 257.

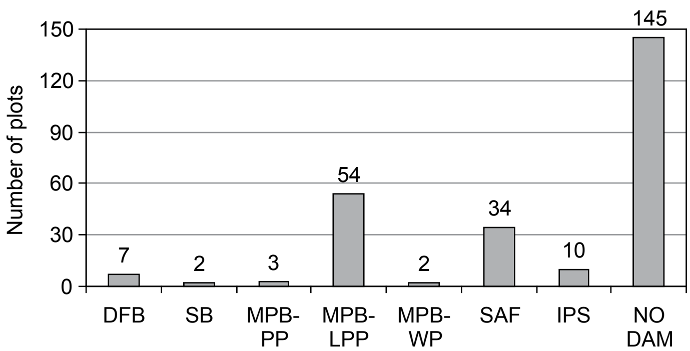

Figure 2. Histogram of visited ground survey plots by non-damage and damage causal agents as determined by field sampling of randomly selected aerial survey observations (n = 257)

Results of the accuracy assessment are shown in Tables 1–4. Accuracy statistics improved with increasing spatial tolerance (Table 1). The overall aerial survey accuracy when no spatial tolerances were allowed was 61%. When a spatial tolerance of 50 m and 500 m was allowed, the aerial survey accuracy improved to 68% and 79%, respectively. Overall κ values for the three error matrices indicate the aerial survey achieved an accuracy that was between 37% and 69% better than what would be expected from the chance assignment of randomly drawn polygons to randomly selected pest categories (Table 1).

Inspection of the error matrices (Tables 2–4) permits inferences as to how well the aerial survey represented truth on the ground. Diagonal values display the agreement between the expected classifications from aerial survey and the observed classifications from field sampling, while off-diagonal values represent disagreement. Row marginals represent the number of points (0.4 ha plots) by category from the field samples. Column marginals represent the number of points assigned to each class as defined by the aerial survey. The diagonal represents the number of points where the ground and aerial surveys agreed. For example, the error matrix in Table 4 reveals that of the 61 points determined as MPB-LPP from the ground sample (row 4 column 9), 53 were correctly classified by the aerial survey (row 4 column 4). Reading successive values along this row shows the field sampling results *that differed from the aerial survey classification* (2, 1 and 5 points belonging to the MPB-WP, SAF and NO DAM ground categories respectively). These numbers are used to calculate errors of omission. Reading successive values down the column from this diagonal reveals the values classified by the aerial survey *that differed from the field sample* (1 and 7 points belonging to the aerial survey categories SAF and NO DAM respectively). These numbers are used to calculate errors of commission.

Table 1. Summary of statistics for three error matrices used for accuracy assessment of the 2005 Rocky Mountain Region's forest health aerial survey

Tolerance	Observed accuracy (%)	Expected accuracy due to chance (%)	κ
0	61	38	0.37
50 m	68	35	0.50
500 m	79	30	0.69

When no spatial tolerances were allowed (Table 2), the MPB-PP and MPB-LPP categories had the lowest omission errors (33% and 35% respectively). The lowest commission error (30%) belonged to the MPB-LPP category. The SB and the MPB-WP categories both the highest omission errors (100%) and highest commission errors (100%).

When a 50-m spatial tolerance was allowed (Table 3) the rank of categories by errors of omission and commission remained unchanged. Omission errors of the MPB-PP and MPB-LPP categories improved to 20% and 26% respectively. The MPB-LPP commission error improved to 20%. The SB and MPB-WP omission and commission errors remained the same (100%).

When a 500-m spatial tolerance was allowed (Table 4), omission errors of the MPB-PP and MPB-LPP categories improved to 0% and 13% respectively. The MPB-LPP commission error further improved to the MPB-LPP category (13%). The SB and the MPB-WP commission errors improved to 89% and 75% respectively.

Discussion

The general target for the accuracy of assessments found in remote sensing literature is an overall accuracy of >85% where all of the classes have a relatively uniform accuracy, but it needs to be stressed that this target accuracy is seldom attained (Foody 2002). While digital image classification is regarded as quantitative in nature, aerial sketchmapping is not. It is commonly considered as much an art as it is a science (McConnell *et al.* 2000). If

Table 2. Error matrix comparison of aerial survey results to ground reference results without spatial tolerances. CA = consumer accuracy, PA = producer accuracy, EO = errors of omission, EC = errors of commission.

Ground data[A]	Aerial survey[A]									PA (%)	EO (%)	EC (%)
	DFB	SB	MPB-PP	MPB-LPP	MPB-WP	SAF	IPS	NO DAM	Totals			
DFB	2							5	7	29	71	33
SB								2	2	0	100	100
MPB-PP			2					1	3	67	33	67
MPB-LPP				35				19	54	65	35	30
MPB-WP								2	2	0	100	100
SAF		1		2	1	15		15	34	44	56	42
IPS							5	5	10	50	50	74
NO DAM	1	4	4	13		11	14	98	145	68	32	33
Totals	3	5	6	50	1	26	19	147	257			
CA (%)	67	0	33	70	0	58	26	67				

Overall classification accuracy = 61%
Overall kappa statistic = 0.37

[A]DFB = Douglas-fir beetle in Douglas-fir, SB = spruce beetle in Engelmann spruce, MPB-PP = mountain pine beetle in ponderosa pine, MPB-LPP = mountain pine beetle in lodgepole pine, MPB-WP = mountain pine beetle in limber pine, SAF = subalpine fir mortality (principally western balsam bark beetle and or *Armillaria* sp. root disease), IPS = *Ips* sp. bark beetle in ponderosa or jack pines, NO DAM = no damage

Table 3. Error matrix comparison of aerial survey results to ground reference results with a spatial tolerance of 50 m. CA = consumer accuracy, PA = producer accuracy, EO = errors of omission, EC = errors of commission.

Ground data[A]	Aerial survey[A]									PA (%)	EO (%)	EC (%)
	DFB	SB	MPB-PP	MPB-LPP	MPB-WP	SAF	IPS	NO DAM	Totals			
DFB	2							5	7	29	71	33
SB								2	2	0	100	100
MPB-PP			4					1	5	80	20	33
MPB-LPP				43				15	58	74	26	20
MPB-WP								2	2	0	100	100
SAF		2		2	1	17		13	35	49	51	39
IPS						1	10	4	15	67	33	47
NO DAM	1	4	2	9		10	9	98	133	74	26	32
Totals	3	6	6	54	1	28	19	140	257			
CA (%)	67	0	67	80	0	61	53	70				

Overall classification accuracy = 68%
Overall kappa statistic = 0.50

[A]DFB = Douglas-fir beetle in Douglas-fir, SB = spruce beetle in Engelmann spruce, MPB-PP = mountain pine beetle in ponderosa pine, MPB-LPP = mountain pine beetle in lodgepole pine, MPB-WP = mountain pine beetle in limber pine, SAF = subalpine fir mortality (principally western balsam bark beetle and or *Armillaria* sp. root disease), IPS = *Ips* sp. bark beetle in ponderosa or jack pines, NO DAM = no damage

Table 4. Error matrix comparison of aerial survey results to ground reference results with a spatial tolerance of 500 m. CA = consumer accuracy, PA = producer accuracy, EO = errors of omission, EC = errors of commission.

Ground data[A]	Aerial survey[A]									PA (%)	EO (%)	EC (%)
	DFB	SB	MPB-PP	MPB-LPP	MPB-WP	SAF	IPS	NO DAM	Totals			
DFB	5							2	7	71	29	17
SB		1						1	2	50	50	89
MPB-PP			5						5	100	0	29
MPB-LPP				53	2	1		5	61	87	13	13
MPB-WP					1			1	2	50	50	75
SAF		4		1	1	24		9	39	61	38	23
IPS						1	15	3	19	79	21	26
NO DAM	1	4	2	7		5	5	98	122	80	20	18
Totals	6	9	7	61	4	31	20	119	257			
CA (%)	83	11	71	87	25	77	75	82				

Overall classification accuracy = 79%
Overall kappa statistic = 0.69

[A]DFB = Douglas-fir beetle in Douglas-fir, SB = spruce beetle in Engelmann spruce, MPB-PP = mountain pine beetle in ponderosa pine, MPB-LPP = mountain pine beetle in lodgepole pine, MPB-WP = mountain pine beetle in limber pine, SAF = subalpine fir mortality (principally western balsam bark beetle and or *Armillaria* sp. root disease), IPS = *Ips* sp. bark beetle in ponderosa or jack pines, NO DAM = no damage

quantitative methods routinely fall below the 85% mark we suggest that aerial sketchmapping, a qualitative method, should have an appreciably lower target accuracy of 70%.

The following is an interpretation for κ values given by Landis and Koch (1977): strong agreement = $\kappa > 0.80$; moderate agreement = $\kappa \geq 0.40$ and ≤ 0.80; poor agreement = $\kappa < 0.40$. While their κ interpretation was developed in the context of observer variability in medical diagnoses, their interpretations have also been used for remote sensing accuracy assessment applications (Ciesla 2000).

The κ value for the error matrix with no spatial tolerance was 0.37 which, according to Landis and Koch (1977), would indicate poor agreement, although close to the lower limit for moderate agreement. κ values for the 50 and 500 m error matrices were 0.50 and 0.70 respectively, both falling within Landis and Koch's (1977) range for moderate agreement.

Errors of omission and commission for the damage categories well represented in the field sample (MPB-LPP, IPS and SAF) were somewhat comparable to results obtained from Brazilian aerial surveys conducted between 2003 and 2004 where omission errors ranged between 0% and 50% and commission errors ranged between 0% and 10% (de Oliveira *et al.* 2006). The Brazilian aerial surveys were conducted over isolated tree plantations using detailed satellite imagery and or stand drawings as base maps. These fine-scale base maps allow polygons to be delineated more precisely, thereby reducing commission error. That is, when multiple delineations can be made at finer scales ('splitting') as opposed to more nebulous approximations ('lumping'), the likelihood that a mapped polygon contains non-damaged areas (commission error) is lower. In effect, the Brazilian surveys were analogous to what is referred to in the United States as 'special surveys' or 'event-specific surveys', more precise delineations of pest locations at the project level using enhanced aerial survey methods (McConnell *et al.* 2000). As this study assesses the accuracy of the broad annual 'overview' aerial survey where much ground is covered in a short time frame (17.3 million ha were covered by aerial surveyors between 7 July and 30 September 2005) it is not surprising that the errors we encountered were somewhat higher and more variable than those reported from Brazil.

Since only a small number of field samples were collected in the DFB, SB, MPB-PP, and MPB-WP classes, their true accuracies remain uncertain. These categories were the greatest sources of error, possibly due to the small sample size as well as other factors such as difficulty discerning pest signatures from the air and or a highly variable distribution of mortality within polygons. For example, Douglas-fir trees killed by DFB tend to occur in aggregated spatial patterns (Cielsa 2006). Often, when the sketchmapper has to delineate these areas quickly while flying at speeds exceeding 160 km h^{-1}, the aggregations of mortality are lumped into larger polygons. Other signatures, such as spruce beetle mortality, are difficult to map due to the subtlety of colour changes (Cielsa 2006) and the brevity of the biological window for detection (McConnell *et al.* 2000). Faded needles on a dead spruce tree remain only for several weeks, whereas needles remain on a dead pine for at least two years. Though spruce beetle surveys are planned for the peak signature period (mid-summer), spruce needle loss is highly variable and can occur, for example, during a spring hail storm.

Because only the MPB-LPP, SAF, IPS and NO DAM categories had samples of sufficient size to infer class accuracies, we will limit further discussion to these classes. Comparing aerial classifications to the ground reference results without spatial tolerances (Table 2), the omission errors for the MPB-LPP, SAF, IPS and NO DAM classes were 35%, 56%, 50% and 32% respectively. Commission errors for the MPB-LPP, SAF, IPS and NO DAM classes were 30%, 42%, 74% and 33% respectively. Accuracies (refer to the CA and PA results in Table 2) for the MPB-LPP and NO DAM classifications met or came near the target accuracy of 70%.

High SAF commission errors may be attributed to the tactic known to sketchmappers and aerial photo-interpreters as lumping (McConnell *et al.* 2000). Subalpine fir mortality is pervasive throughout the range of subalpine fir and the many, often scattered, spots of SAF mortality are commonly 'lumped' by sketchmappers into very large polygons (2005 mean polygon size = 31 ha). In addition, subalpine fir decline is often accorded a 'lower mapping priority'; particularly where tree species considered more important by land managers are dying across a uniform area. This probably explains why omission errors are greater than commission errors within the SAF class.

The IPS classification had an omission error of 50% and a commission error of 74%. Because most mortality caused by *Ips* sp. within the region is found within small, isolated pockets (2005 mean polygon size = 1.4 ha), the high commission error is probably caused by difficulties in precisely delineating these small areas on the 1 : 100 000 scale base maps frequently used during a sketchmap survey, resulting in the spatial displacement of these polygons. Pin-pointing these smaller infestations would be made easier by using a finer-scale base map, such as 1 : 24 000 scale, or one providing more detail, such as digital ortho quads (DOQs), satellite imagery or scanned aerial photography.

When a spatial tolerance of 50 m was allowed (i.e. the 0.4 ha minimum map unit of 63.6 × 63.6 m was buffered by an additional 50 m), omission errors improved to 26%, 51%, 33% and 26% for the MPB-LPP, SAF, IPS and NO DAM classes respectively (Table 3). Commission errors improved to 20%, 39%, 47% and 32% for the MPB-LPP, SAF, IPS and NO DAM classes respectively. By allowing a spatial tolerance of 50 m, accuracies for the MPB-LPP and NO DAM categories exceeded the 70% target accuracy mark and the SAF and IPS classes were significantly improved, moving closer to our target mark. The high commission error of the IPS class caused by polygonal displacement was greatly improved by allowing a spatial tolerance of 50 m.

Accuracies further improved when the spatial tolerance was increased to 500 m. Accuracies for the MPB-LPP and NO DAM classes exceeded 80%. Accuracies for the IPS class exceeded 70% and the SAF class nearly reached the target of 70%. The SAF class's omission error remained high (38%) which, again, was probably due to its 'lower mapping priority'.

Historically, the primary purpose for aerial surveys has been to detect and delineate new pest outbreaks and to generate maps sufficient for facile ground orientation. Thus, a moderate agreement between aerial classification and ground observation ± 500 m has been adequate for most detection purposes. Recently, however, aerial survey data have been used for monitoring and modelling functions that require a higher degree of spatial explicitness. Incorporating accuracy assessment into aerial survey programs would not only help describe the confidence levels of ensuing models, but it would improve aerial survey results by highlighting classification strengths and weaknesses (e.g. aerial survey training programs could focus on improving those classifications found to be the most errant).

One goal of this pilot study was to incorporate a quantitative measure of aerial survey accuracy into the metadata to give users a better understanding of the errors associated with the data set. That is, the 2005 aerial survey data set meets the 70% accuracy target when used and viewed at coarse scales (79% accurate ±500 m) and is somewhat suitable when used and viewed at intermediate scales (68% accurate ±50 m). If objectives called for site-specific information, however, the survey tactics would need refinement to improve spatial accuracies (e.g. planning a helicopter survey using high-resolution satellite imagery as a base map) since the non-tolerant accuracy rate was only 61%.

Another goal of our pilot project was to determine the feasibility of implementing an annual program to measure accuracies for each year's data set. The methodology we applied, along with some sampling refinements suggested below, would be a functional solution for assessing accuracy. Field personnel spent about 650 person-hours collecting the reference data over the course of three months. Additionally, our GIS specialist spent about 300 person-hours generating the sample points and creating maps for the field crews. The total cost of the effort, considering salary, per diem and fuel cost, was about US$24 000; less than 10% of the operational cost of our annual aerial 'overview' survey. The feasibility of incorporating auxiliary data into the error matrix from other sources was also demonstrated which would help offset sampling costs. A 10% increase in operating costs for establishing an accuracy assessment system within an aerial survey program appears warranted when considering the benefits; namely, quantifying spatial errors for the accompanying metadata, improving aerial survey results through signature-specific training, and improving aerial survey results through the implementation of better survey methodologies (i.e. finding the 'right tool for the right job').

Improving the sampling design

Our sample was stratified into two classes: damage and no damage. In essence, the damage categories were proportionally stratified by the areal extent of each class included in the accuracy assessment. Because the areal extent of damage from subalpine fir decline and mountain pine beetle in lodgepole pine is greatest within the region, these classes were well represented by the field sample whereas other classes, such as Douglas-fir beetle and spruce beetle damage, were under-represented. For future assessments, it may be beneficial to stratify the damage classes equally, especially in cases where determining the class accuracies for the less-represented pests is sufficiently important to justify increased sampling costs. One such example would be the spruce beetle damage class because of its high social and economic value. Thus, to equally allocate the damage classes into strata, the total number of points to be sampled (n) can be calculated using probability theory (Scheaffer *et al*. 1996):

$$n = z^2(p)(q)/E^2 ,$$

where: p = expected accuracy, $q = 100 - p$, E = allowable error (expressed in the same form as p, either percent or decimal), $z = 2$ (i.e. 95% two-sided confidence level). With an expected accuracy of 70% and a 5% allowable error, 336 points would need to be sampled. One-half of these points (168) should be generated from the no-damage category to ensure minimal sampling bias. The other half should be equally allocated to each damage class — in this case 24 points per class for each of the seven damage classes. This improvement to the sampling design would ensure that each class would have sufficient samples to infer individual class accuracies.

Conclusion

The error matrix traditionally used for assessing remote sensing classifications can also be applied to the assessment of aerial survey class accuracies. Furthermore, supplemental data can be used to augment the field sample and reduce sampling costs. Accuracies of the 2005 annual 'overview' survey, which ranged from an estimated 61% based on site-specific comparisons to an estimated 79% when a 500-m spatial tolerance was allowed, were incorporated in the aerial survey's metadata in order to give users a better understanding of its strengths and limitations. While accuracy levels met established operational objectives for broad-scale detection and monitoring, they fell short of more contemporary needs for fine-scale modelling.

Results of the pilot project presented here can be used to enhance future assessment routines. The suggestion to equally allocate the sample points by stratum will ensure that accuracies for every category are better known. Knowing which classes have the highest error rates would allow program managers to better guide data quality assurance and quality control (QA/QC) efforts.

It is critical to evaluate the accuracy of any geospatial data set, and aerial survey data should be treated no differently. Implementing accuracy assessment measures in aerial survey programs should become a standard practice since the use of the data in spatial models has increased considerably. It is increasingly important that both originators and consumers of the data understand its strengths and limitations. Originators of the data would benefit from accuracy assessment by targeting training programs and enhancing methods to resolve specific accuracy issues, while consumers would benefit by basing their assumptions on more quantitative measures.

References

Campbell, J.B. (2002) *Introduction to Remote Sensing*. 3rd edn. The Guilford Press, New York, 621 pp.

Ciesla, W.M. (2000) *Remote Sensing in Forest Health Protection*. USDA Forest Service, Forest Health Technology Enterprise Team, Fort Collins, Colorado, FHTET 00-13, 266 pp.

Ciesla, W.M. (2006) *Aerial Signatures of Forest Insect and Disease Damage in the Western United States*. USDA Forest Service, Forest Health Technology Enterprise Team, Fort Collins, Colorado, FHTET 01-06, 94 pp.

de Oliveira, Y.M.M., Rosot, M.A.D., Ciesla, W.M., Johnson, E., Rhea, R., Penteado Jr, J. and da Luz, N.B. (2006) Aerial sketchmapping for monitoring forest conditions in southern Brazil. In: *Monitoring Science and Technology Symposium: Unifying Knowledge for Sustainability in the Western Hemisphere*. 20–24 September 2004, Denver, CO. USDA Forest Service, Rocky Mountain Research Station, Fort Collins, Colorado, Proceedings RMRS-P-42CD.

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

Landis, J. and Koch, G. (1977) The measurement of observer agreement for categorical data. *Biometrics* **33,** 159–174.

McConnell, T., Johnson, E. and Burns, B. (2000) *A Guide to Conducting Aerial Sketchmap Surveys*. USDA Forest Service, Forest Health Technology Enterprise Team, Fort Collins, Colorado, FHTET 00-01, 88 pp.

Scheaffer, R.L., Mendenhall, W. and Ott, L. (1996) *Elementary Survey Sampling*. 5th edn. Duxbury Press, 501 pp.