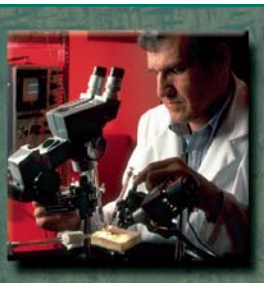
CFS
CANADIAN
FOREST SERVICE



TECHNOLOGY TRANSFER NOTE
Number 36
Forestry Research Applications

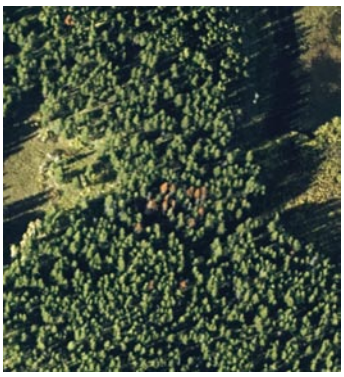
Pacific Forestry Centre

August 2007

Assessing the Accuracy of Mountain Pine Beetle Red Attack Damage Maps Generated from Satellite Remotely Sensed Data

J.C. White, M.A. Wulder, D. Grills

Background



Satellite remotely sensed data may be used to detect and map mountain pine beetle (*Dendroctonus ponderosae*) red attack damage at a variety of spatial scales. The accuracy of the output red attack damage maps should be assessed and reported in a transparent manner to allow the end user to both apply the data appropriately and to compare outputs generated using different methods and data sources.

Strategic Importance

An accuracy assessment is the comparison of a generated map product to another source of data that represents actual conditions on the ground (*i.e.*, ground truth or validation data). In the context of mountain pine beetle red attack damage, an output map produced from the analysis of satellite remotely sensed data may be compared to a sample of field survey data (that identifies known locations of red attack damage on the ground). One of the advantages of using remotely sensed data to map red attack damage is the economies of scale that are achieved from mapping very large areas; it is expensive and often logistically impossible to map such large areas using ground surveys. The objective of an accuracy assessment is therefore to compare the output map of red attack damage generated from the remotely sensed data to a small sample of very accurate ground data, to determine how reliable the map product is. An accuracy assessment not only indicates to the map user the reliability of the map for planning, geospatial analysis, and modelling, but also provides the map producer with a measure of success for their methods and/or data choices. An error matrix is a useful tool for summarizing this comparison between the ground truth data and the remotely sensed map output. In addition, an error matrix facilitates the calculation of accuracy measures, enabling a quantitative assessment of the map produced. An example of a simple error matrix with two classes (not attacked and red attack) is provided in Table 1 and we review the contents of the error matrix and the concept of accuracy assessment in the next section.

Table 1. Sample error matrix with equal sample sizes per class.

		Output Map Generated from Satellite Remotely Sensed Data				
	Classes	**Not attacked**	**Red Attack**	**Sum**	**Producer's Accuracy**	**Omission Error**
Ground Truth Data	**Not attacked**	140	10	150	93%	7%
	Red Attack	60	90	150	60% 95% CI: 52%-67%	40%
	Sum	200	100	300		
	User's Accuracy	70%	90%	**Overall Accuracy: 77%** 95% Confidence Interval: 72% - 81%		
	Commission Error	30%	10%			

Understanding the Concept of Accuracy Assessment

In the example error matrix in Table 1, 300 samples were collected during a field survey (distributed equally amongst the not attacked and red attack classes). At each of these sample locations, one or more not attacked or red attack trees were identified by the field crew

and the position of the tree(s) recorded with a Global Positioning System (GPS). The error matrix is arranged with the predicted class from the remotely sensed data along the top, and the actual class (as indicated by the ground truth data) along the left side of the table. There are three primary measures of accuracy calculated from an error matrix: producer's, user's, and overall accuracy. Producer's accuracy tells us what proportion of the ground truth samples were correctly labelled on the map. In the example in Table 1, 140 of the not attacked ground samples were labelled not attacked on the output map, resulting in a producer's accuracy of 93%. Conversely, only 90 of the 150 red attack ground samples were labelled correctly on the map, resulting in a producer's accuracy of 60% for the red attack class. The omission error is the corollary measure to producer's accuracy, and tells us what proportion of ground samples of a particular class were misclassified on the output map product (*e.g.*, 40% of the red attack ground samples in Table 1 were not classified as red attack).

User's accuracy tells us what proportion of the samples that are labelled a particular class on the map, are actually that class on the ground. In other words, if we were to go out in the field with the map and visit every site identified as a particular class on the map, what proportion of the sites we visited would actually be the class labelled on the map? In Table 1, we see that of the 200 samples that were mapped as not attacked on the output map, only 140 were actually not attacked on the ground, resulting in a user's accuracy of 70% for the not attacked class. Similarly, 90 of the 100 samples mapped as red attack were actually red attack on the ground, resulting in a user's accuracy of 90%. Commission error is the corollary of user's accuracy and tells us what proportion of the samples assigned to a particular class on the output map were actually a different class on the ground. In Table 1 we see that only 10% of the samples identified as red attack on the map were actually not red attack on the ground.

The overall accuracy is a measure of the total proportion of samples that were classified correctly by the map, regardless of the class. This measure is calculated by summing along the diagonal of the error matrix (shaded gray in Table 1). In this example, we see that 230 samples (140 for not attacked and 90 for red attack) were mapped correctly, resulting in an overall accuracy of 77%. The overall accuracy provides a general indication of the accuracy of the map if all classes are of equal importance. If all classes are not of equal importance, the overall accuracy may misrepresent the accuracy of the map. In this example, we are more interested in how well our map identified areas of red attack damage. If we just reported the overall accuracy, we may misrepresent our results, since our success at mapping the not attacked class has bolstered our measure of overall accuracy. The producer's accuracy provides a better indication of how well our approach is working. From this we can conclude that our methods are missing a substantial amount of the red attack damage (40%).

On the other hand, the user's accuracy and commission error for the red attack class tell us that where we are identifying red attack on the map, we are indeed identifying red attack on the ground and not confusing it with the not attacked class. There may be two practical ways for us to interpret this information. Firstly, we know that there is opportunity for us to improve our detection of red attack damage, and therefore we may want to look at altering our methods or trying a different remotely sensed data source to increase the amount of damage we can detect and map. For example, some of the methods used to map red attack damage from remotely sensed data rely on the establishment of a threshold, which distinguishes the red attack damage from other disturbance types. The accuracy assessment may indicate that the selected threshold is inappropriate and requires adjustment. We may also want to examine the quality of the validation data we have selected. Secondly, from an operational perspective, a high commission error can be very expensive, since deploying ground crews to areas that the map says are red attack, but that are not really red attack on the ground, is costly. We therefore not only want to reduce our omission error, but also want to try and maintain a low commission error as well.

The example above indicates the value of the information contained in an error matrix. Reporting only the overall accuracy does not provide the full context to the end user, particularly in a situation where there are only two classes being mapped. We therefore recommend that the full error matrix be reported. Table 2 demonstrates why this is the case; the producer's accuracies are identical to those reported in Table 1, but in this case, we are assuming that the ground truth samples were not evenly distributed amongst the classes (two-thirds of the ground samples were collected from the not attacked class). In this example, we see that although our ability to detect red attack damage has not changed (our producer's accuracy is still only 60%), our overall accuracy has increased to 82%. Also note that our ability to detect the not attacked class has not necessarily improved either, but because we have more samples in the not attacked class, and since our mapping method does a good job of detecting not attacked, our overall accuracy result has increased.

Table 2. Sample error matrix with unequal samples sizes per class.

	Output Map Generated from Satellite Remotely Sensed Data				
Classes	**Not attacked**	**Red Attack**	**Sum**	**Producer's Accuracy**	**Omission Error**
Not attacked	186	14	200	93%	7%
Red Attack	40	60	100	60% 95% CI: 50%-69%	40%
Sum	226	74	300		
User's Accuracy	82%	81%	**Overall Accuracy: 77%** 95% Confidence Interval: 77% - 86%		
Commission Error	18%	19%			

Methods

Selecting Ground Truth Data

The selection of ground truth data is critical to any accuracy assessment. Data collected on the ground by survey crews are often the most desirable data for accuracy assessment; however, since ground data are also the most expensive data to collect, they are not always the most practical choice. Another consideration is the timing of the survey. Ideally, any data used for validation would be acquired at the same time as the remotely sensed data used to generate the output map, which can prove challenging if archived images are used. All of these factors invariably lead to some compromise in data selection. However, there are viable alternatives to ground surveys of red attack damage, and these include aerial photography (White et al., 2005) and helicopter GPS surveys (Nelson et al., 2006).

Regardless of the data used for validation, several other factors must be considered when selecting validation samples, the most important of which is sample size. We have already demonstrated the importance of using equal sample sizes for each class (but acknowledge that unequal sample sizes may be preferable for other applications). The sample size will affect the confidence interval constructed for the accuracy estimates. The remote sensing literature suggests a minimum of 50 samples per class, while we would suggest collecting between 50 and 100 samples per class. The more samples acquired, the greater confidence the end user will have that the accuracy results reported are representative of the map product. Therefore, we recommend that the confidence intervals are calculated and reported for both overall accuracy and the producer's accuracy for the red attack class. Validation samples should also be selected to be spatially representative of the study area and must be independent of any samples used for calibration of the algorithm to classify the remotely sensed data.

A Note on Confidence Intervals

A confidence interval provides an estimated range, calculated from the sample data, which is likely to include the accuracy estimate. The width of the confidence interval provides information on how confident we are in our accuracy estimate; the wider the confidence interval, the less confidence we have in the accuracy estimate. Several confidence interval calculators for proportions (*i.e.*, binomial distribution) are available on the internet and provide useful approximations for reporting[1]. In the examples presented in Tables 1 and 2, we can see that the confidence intervals associated with the overall accuracy are the same width, as the overall sample size is the same. However, due to the smaller sample size for red attack in the Table 2 example, the confidence interval associated with the producer's accuracy for the red attack class is wider than in Table 1 – indicating lower confidence in this estimate. Therefore, although the producer's accuracies are the same for these two examples, we would have greater confidence in the red attack map associated with Table 1. Table 3 demonstrates how confidence intervals vary with sample size; as the sample size increases, the confidence interval narrows. Based on this relationship between sample size and confidence interval, map producers can judiciously choose how many samples are required to provide the desired level of confidence associated with their accuracy estimates, given the resources they have available for acquiring the samples.

Table 3. The impact of sample size on the width of the confidence interval, assuming the same level of accuracy.

	Producer's Accuracy = 85% 95% Confidence Interval	
Sample Size	**Lower Confidence Limit**	**Upper Confidence Limit**
500	0.82	0.87
250	0.80	0.89
100	0.77	0.91
75	0.76	0.92
50	0.74	0.93
25	0.63	0.95

Undertaking an Accuracy Assessment

The mechanics of undertaking an accuracy assessment are straightforward with modern and widely available GIS technology. The simplest approach is to have the ground truth data as a point data set, and then overlay these points with the map of not attacked/red attack generated from processing the remotely sensed data. The result will be a collection of points having attributes indicating what class the points were assigned from the ground data, and what class the points were assigned by the remotely sensed output. This information can then be used to construct the error matrix. In some cases, the possibility of positional error in either the truth data or the output map may necessitate the use of buffers or some other mechanism to account for spatial error[2]. Table 4 provides a summary of how each of the accuracy measures is calculated. Although the terminology we have used throughout this communication (*e.g.*, producer's accuracy) is fairly standard in the remote sensing literature, the reader may come across different terminology in other disciplines and we have provided a summary of these (in the context of mountain pine beetle red attack) in Table 5. Equipped with these tools, and with the examples presented earlier, the reader should be able to critically review the results of accuracy assessments associated with mountain pine beetle red attack damage detection and mapping, and/or conduct an accuracy assessment using their own ground truth and map products.

[1]For an example see: http://faculty.vassar.edu/lowry/prop1.html

[2]For more detailed discussion on the use of buffers to account for positional error see White et al. (2005).

Table 4. Calculating the estimates in an error matrix.

		Output Map Generated from Satellite Remotely Sensed Data				
	Classes	**Not attacked**	**Red Attack**	**Sum**	**Producer's Accuracy**	**Omission Error**
Ground Truth Data	**Not attacked**	A	B	A+B	A/(A+B)	B/(A+B)
	Red Attack	C	D	C+D	D/(C+D)	C/(C+D)
	Sum	A+C	74	A+D		
	User's Accuracy	A/(A+C)	D/(B+D)	**Overall Accuracy:** (A+D)/(A+B+C+D)		
	Commission Error	C/(A+C)	B/(B+D)			

Table 5. A crosswalk for terminology commonly used in accuracy assessment reporting

Alternative Term	Which Class?	Common Term
True Positive Rate	Red Attack	Producer's Accuracy
False Positive Rate	Not Attacked	Ommission Error
True Negative Rate	Not Attacked	Producer's Accuracy
False Negative Rate	Red Attack	Ommission Error
Precision	Red Attack	User's Accuracy
Accuracy	Both	Overall Accuracy

Summary

An accuracy assessment is considered the best way to demonstrate the effectiveness with which different data sources and methods may be used to map mountain pine beetle red attack damage from remotely sensed data. Simply reporting overall accuracy, however, does not provide sufficient context to evaluate the map product and may misconstrue the accuracy with which red attack damage is detected and mapped. Accuracy assessments are not difficult to undertake, nor should they be difficult to interpret. We have made several recommendations regarding accuracy assessment in the context of mountain pine beetle red attack detection and mapping and they can be summarized as follows:

- Ground surveys provide the best source of validation data for not attacked and red attack; however, air photos and helicopter GPS surveys are viable alternatives.
- Select validation samples that are spatially representative and that are independent from any calibration data used for the classification.
- Select at least 50 to 100 samples per class for validation. The more samples used, the greater the confidence in the accuracy estimates reported.
- Use equal sample sizes for both not attacked and red attack classes.
- Report the contents of the full error matrix, not just overall accuracy.
- Report 95% confidence intervals for both overall accuracy and the producer's accuracy for red attack.

References and Additional Reading

Coops, N.C.; Johnson, M.; Wulder, M.A.; White, J.C. 2006. Assessment of Quickbird High Spatial Resolution Imagery to Detect Red-Attack Damage due to Mountain Pine Beetle Infestation. Remote Sensing of Environment 103:67-80.

Nelson, T.; Boots, B.; Wulder, M.A. 2006. Representing large area mountain pine beetle infestations. Forestry Chronicle 82:243–252.

Skakun, R.S.; Wulder, M.A.; Franklin, S.E. 2003. Sensitivity of the Thematic Mapper Enhanced Wetness Difference Index (EWDI) to detect mountain pine needle red attack damage. Remote Sensing of Environment 86: 433–443.

White, J.C.; Wulder, M.A.; Grills, D. 2006. Detecting and Mapping Mountain Pine Beetle Red Attack with SPOT Imagery. British Columbia Journal of Ecosystems and Management 7:105-118.

White, J.C.; Wulder, M.A.; Brooks, D.; Reich, R.; Wheate, R.D. 2005. Detection of Red Attack Stage Mountain Pine Beetle Infestation with Spatial Resolution Satellite Imagery. Remote Sensing of Environment 96:340-351.

Wulder, M.A.; Dymond, C.C.; White, J.C.; Leckie, D.G.; Carroll, A.L. 2006. Surveying mountain pine beetle damage of forests: A review of remote sensing opportunities. Forest Ecology and Management 221: 27–41.

Contacts:

Dr. Mike Wulder
Pacific Forestry Centre, Canadian Forest Service
Natural Resources Canada
506 West Burnside Road
Victoria BC V8Z 1M5
Telephone: (250) 363-6090
email: mwulder@nrcan.gc.ca

For additional information on the Canadian Forest Service visit our website at: cfs.nrcan.gc.ca/regions/pfc

Acknowledgements:

This study was funded by the Government of Canada through the Mountain Pine Beetle Initiative, a six-year, $40-million program administered by Natural Resources Canada, Canadian Forest Service.

Cat. No. Fo143-4/36-2007E • ISBN 978-0-662-46532-4
Cette publication est aussi disponible en français.

Canada