Region One
Vegetation Classification, Mapping, Inventory and Analysis Report

Numbered Report NRGG16-01 **January, 2016**

Custer-Gallatin National Forest – VMap 2015
Tree Dominance Type (DOM40), Tree Canopy Cover, Tree Size Class, and Lifeform Accuracy Assessment

Stephen R. Brown Jr.
Northern Region Geospatial Group
USDA Forest Service, Region 1
Engineering Staff
200 E. Broadway, Missoula, MT 59802

1. Accuracy Assessment Defined

Accuracy assessment is an essential part of any remote sensing project. It provides the basis of comparison for different methods and/or sensors. It provides information regarding the reliability and usefulness of remote sensing techniques for a particular application. Most importantly accuracy assessment provides a validation of the data, giving an indication of reliability of the classification, so that managers are fully informed throughout the decision making process (Congalton, 1991). Too often vegetation and other maps are used without a clear understanding of their reliability. A false sense of security about the accuracy of the map may result in an inappropriate use of the map and important decisions may be made based on data with unknown and/or unreliable accuracy. Estimates of overall map accuracy and confidence of individual map classes can be inferred from an error matrix derived from the comparison of known reference sites to mapped data. Although quantitative accuracy assessment can be time-consuming and expensive, it is an integral part of any vegetation-mapping project.

Accuracy, however, is not a state variable. It is very important to evaluate the results of any accuracy assessment in the context of the intended analysis application and the management decision the data and analyses are intended to support. This evaluation needs to balance the desired level of precision (i.e., the level of thematic detail) with the desired level of accuracy (i.e., spatial location of a given attribute). For many analyses, detailed thematic classes are aggregated to produce more generalized classes, a technique that will typically increase the accuracy of a given map. It is appropriate in these instances to assess the accuracy of the aggregated classes rather than characterize the aggregations with the detailed assessment. It may even be appropriate to aggregate some classes based on the structure of the error, provided that the aggregations meet the analysis objectives. It is also important to determine the level of uncertainty that is acceptable to support a particular management decision.

Quantitative accuracy assessment depends upon the collection of reference data with which to compare the map product in question. It is therefore assumed that the reference data is “truth”, that is 100% correct. Reference data can be obtained via field site visits, photo-interpretation, existing plot data, or a combination of these methods. For the purposes of this assessment a stratified random sample design was constructed following the recommendations of Stehman and Czaplewski (1998).

After completion of the photo-interpretation process for all polygons, comparisons of these data to the mapped elements are then tabulated and presented in an error matrix, where the rows represent values of the map, and columns represent values of the reference data. Tabulated values across the diagonal of the matrix describe the number of times map and reference data sites have equal values. Conversely, the off-diagonal table elements quantify errors of either inclusion or exclusion of particular classes. Errors of inclusion are shown on the horizontal axis of classes, while errors of exclusion are shown on the vertical axis. Large numbers of inclusion or exclusion between two or more classes indicate a high degree of inter-class confusion and generally indicate a lower quality map. To illustrate these concepts, an error matrix quantifying the level of agreement in a theoretical lifeform map is given below as Table 1.

Table 1. *Error matrix of a theoretical lifeform map, with overall map accuracy of 74%*

		Reference Data Classes				
		Forest	*Shrub*	*Herbaceous*	*Water*	**Map Total**
Map Data Classes	*Forest*	**65**	4	22	24	115
	Shrub	6	**81**	5	8	100
	Herbaceous	0	11	**85**	19	115
	Water	4	7	3	**90**	104
	Ref. Total	75	103	115	141	434

Once an error matrix table has been created, several useful measures of map accuracy can be computed, including overall, producer, and user metrics. Overall accuracy is a common metric that describes how well the map compares to a reference dataset as a whole. Producer accuracy focuses on errors of exclusion and thus is a term that describes the number of samples that were incorrectly classed. User accuracy, on the other hand, is based on errors of inclusion and therefore reflects the probability that a feature of the map actually represents that category on the ground. Regardless of the measurement used, the robustness of the metric is largely dependent on the number of samples that were used for comparison. In the best case scenario a similar number of samples will be available for each map class, and each class will have a large number of samples, which generally means more than 30 instances. It is unfortunate, but an assessment of individual class accuracy cannot be conducted when there are an insufficient number of reference samples available. In such cases users of the map should be aware that while the error in some map classes is not quantified in an error matrix, it can be assessed either through additional reference data collection, or via systematic field review of the classification.

Overall Accuracy is computed by dividing the total number of correct samples by the total number of assessment sites found in the bottom right cell of the error matrix table.
It is often the most commonly reported accuracy measure because it takes advantage of samples from all classes. Not all map classes will have large enough samples available for comparison. With Table 1 as an example, it can be seen that 434 sites were evaluated against their known condition on the ground. By adding the total number of times mapped classes were in agreement with their known condition and dividing that total by the total number of sites that were evaluated the overall accuracy of the map can be assessed as follows:

[*Forest* (**65**) + *Shrub* (**81**) + *Herbaceous* (**85**) + *Water* (**90**) = 321] / 434 = **74%**

Producer Accuracy is the probability of a reference site being correctly classified, and is calculated by dividing the total number of correctly mapped sites for a class by the total number of reference sites for that class. Using data from Table 1, Producer’s class accuracy values are assessed as follows in Table 2:

Table 2. *Computation of Producer Map Accuracy*

Map Class	# of correct sites	# of all reference sites	Relative Accuracy (%)
Forest	**65** divided by	**75**	**= 87**
Shrub	**81** divided by	**103**	**= 79**
Herbs	**85** divided by	**115**	**= 74**
Water	**90** divided by	**141**	**= 64**

User Accuracy is the probability that a feature on the map actually represents that category on the ground, and is calculated by dividing the number of agreements for a category by the total number of sites that were mapped into that category. Using data from Table 1, User class accuracy values are assessed as follows in Table 3:

Table 3. *Computation of User Map Accuracy*

Map Class	# of correct sites	# of all mapped sites	Relative Accuracy (%)
Forest	**65** divided by	**115**	**= 57**
Shrub	**81** divided by	**100**	**= 81**
Herbs	**85** divided by	**115**	**= 74**
Water	**90** divided by	**115**	**= 87**

2. Results

Following the recommendations of Stehman and Czaplewski (1998), a stratified random sample was designed for the area covered by the Custer-Gallatin National Forest VMap 2015 revision. For both the Lifeform and Tree Canopy Cover attributes strata were constructed for each of the available classes and 750 samples/class, well distributed across the area, were selected and assessed. For the Tree Dominance Type (DOM40) and Tree Size Class attributes the assessment was conducted a little differently due to the fact that neither of these attributes lend themselves very well to being photo-interpreted. In this case, then, a random selection of 10% of each class were held back from the classification routine to compare to the resulting output.

For the Lifeform attribute there were 6 strata selected: Herbaceous, Shrub, Sparsely-Vegetated, Water, Deciduous Tree, and Coniferous Tree, with 750 sample sites selected and evaluated for classification into a corresponding VMap Lifeform class for comparison to the existing map. The results are shown in Table 4 below.

CusGal Vmap v15 NFS Lands Lifeform Accuacy Assessment								
Class	Herbaceous	Shrub	Conifer	Water	Sparsely Vegetated	Deciduous Tree	Total	User Accuracy
Herbaceous	**1010**	73	16	3	21	8	**1131**	**89%**
Shrub	112	**872**	18	0	12	13	**1027**	**85%**
Conifer	16	18	**555**	0	4	3	**596**	**93%**
Water	2	1	2	**346**	4	1	**356**	**97%**
Sparsely Vegetated	22	1	5	3	**536**	0	**567**	**95%**
Deciduous Tree	10	38	16	0	0	**489**	**553**	**88%**
Column Total	**1172**	**1003**	**612**	**352**	**577**	**514**	**4230**	**Overall Accuracy**
Producer Accuracy	**86%**	**87%**	**91%**	**98%**	**93%**	**95%**		**90%**

Table 5. Custer-Gallatin NF VMap 2015 Lifeform Error Matrix

There were 10 classes evaluated for DOM40: MX-PIPO, MX-PSME, MX-PICO, MX-ABLA, MX-PIEN, MX-PIAL, MX-PIFL, MX-JUSC, TMIX and IMIX. A 10% withholding for each class was then compared to the resulting map to provide the quantification in the error matrix shown in Table 5 below.

Custer Gallatin National Forest Vmap V15 Dom_MID_40 Accuracy Assessment												
	Reference Data											
Map Class	MX-PSME	MX-ABLA	MX-PICO	MX-PIEN	MX-PIAL	MX-JUSC	MX-PIPO	MX-PIFL	TMIX	IMIX	Total	User Accuracy
MX-PSME	187	4	14	8	0	3	0	1	0	0	217	86
MX-ABLA	3	10	3	7	2	0	0	0	1	5	31	32
MX-PICO	30	6	62	3	2	0	0	0	1	0	104	60
MX-PIEN	3	8	6	13	2	0	0	0	0	0	32	41
MX-PIAL	0	3	2	0	14	0	0	0	0	1	20	70
MX-JUSC	1	0	0	0	0	13	0	3	0	0	17	76
MX-PIPO	0	0	0	0	0	0	62	0	0	0	62	100
MX-PIFL	0	0	0	0	0	1	0	6	0	0	7	86
TMIX	0	0	2	0	0	0	0	0	0	0	2	0
IMIX	2	0	1	1	0	1	1	0	0	0	6	0
Total	226	31	90	32	20	18	63	10	2	6	498	Overall Accuracy
Producer Accuracy	83	32	69	41	70	72	98	60	0	0		74

Table 5. Custer-Gallatin NF VMap 2015 DOM40 Error Matrix

There were 4 strata selected for the Tree Canopy Cover class evaluation: Low Canopy Cover Tree (10-24.9%), Moderate-Low Canopy Cover Tree (25-39.9%), Moderate-High Canopy Cover Tree (40-59.9%), and High Canopy Cover Tree (60% +), with 600 sample sites selected within each strata. This allows for ample site availability should the need arise to discard some sites as unsuitable for labeling (i.e., deep shadows where the canopy cover cannot be ascertained). These sites were then evaluated for classification into a corresponding VMap Tree Canopy Cover class and compared with the existing Map. The results are displayed in Table 6 below.

Custer Gallatin Vmap V15 Tree Canopy Cover Accuracy Assessment						
	Reference Data					
Vmap Class	10-25% Cover	25-40% Cover	40-60% Cover	60% + Cover	Total	User Accuracy
10-25% Cover	488	54	7	4	553	88%
25-40% Cover	28	471	58	15	572	82%
40-60% Cover	7	52	477	46	582	82%
60% + Cover	6	15	58	483	562	86%
Total	529	592	600	548	2269	Overall Accuracy
Producer Accuracy	92%	80%	80%	88%		85%

Table 6. Custer-Gallatin NF VMap 2015 Tree Canopy Cover Error Matrix

For the Tree Size Class assessment there were 4 classes evaluated: Seedling Tree (0-4.9" DBH), Small Tree (5-9.9" DBH), Medium Tree (10-14.9" DBH), and Large/Very Large Tree (15"+ DBH), by a 10% withholding of the field sampled data within each class. These sites were then evaluated for classification into a corresponding VMap Tree Size class and compared with the existing Map. The results are displayed in Table 7 below.

Custer/Gallatin Tree Size Class Accuracy Assessment Vmap V15						
	Sample Class					
Vmap Class	0-5" DBH	5-10" DBH	10-15" DBH	15" + DBH	Grand Total	User Accuracy
0-5" DBH	75	9			84	89%
5-10" DBH	25	79	9		113	70%
10-15" DBH		12	76	5	93	82%
15" + DBH			15	95	110	86%
Grand Total	100	100	100	100	400	Overall Accuracy
Producer Accuracy	75%	79%	76%	95%		81%

Table 7. Custer-Gallatin NF VMap 2015 Tree Size Class Error Matrix

3. Discussion

There are some advantages and some disadvantages to constructing a post-classification, stratified random sample based accuracy assessment. The biggest advantage is that there is a guarantee of sufficient N in the sample base that a full assessment of each represented class is possible (i.e., the Lifeform and Tree Canopy Cover assessments). The biggest disadvantage is that the ability to estimate a true “Producer’s Accuracy” (quantification of the omission error) is lost due to the biased nature of the sample selection. All things considered, however, the advantage of having the “User’s Accuracy” available for each of the classes in being evaluated outweighs this disadvantage.

Since not all of the map attributes lend themselves to confident interpretation through image analysis, specifically Tree Size and Tree Dominance type, it becomes necessary to withhold a certain amount of the field collected samples in order to provide an independent estimate of the map class accuracy. The advantage to the withholding approach is that there is more confidence that can be placed in the assessed site over an image interpretation. The disadvantage is that not all classes receive sufficient N within the sample base to provide a valid quantification of the error for that class, i.e., TMIX and IMIX each having less than 10 occurrences in the field based dataset which is far short of the minimum 30 N for a statistically valid sample.

In general, the accuracies exhibited in the VMap 2015 database are exceptional. Those classes with higher error rates, i.e., MX-ABLA and MX-PIEN are confused, tend to be those types that occur in proximity to each other or frequently grade in and out in abundance together such that a mislabeled polygon could still be considered “OK” in most analysis situations. In this example even the Society of American Foresters do not discriminate between Engelmann spruce and subalpine fir, labeling simply as spruce-fir. The same can be said of the Tree Canopy and Tree Size Class attributes, where most of the error occurs between adjacent classes and can easily be attributed to either interpretation error or just the inherent fact that when a continuous world is parceled into discrete classes not everything will always fit neatly. For example, if a given polygon is estimated to have 61% tree canopy cover, but the analyst estimates that it has 59%, the true difference is only 2%, but 59.9% is the cutoff between two classes so that the polygon would then be assessed as incorrect.

The take home message is that even the accuracy assessment, which is judged as "truth" because there has to be some standard by which to compare the map, needs to be taken with a grain of salt. While the accuracy assessment is an attempt at a numerical quantification of the error structure in the map products, this is no substitute for a qualitative map evaluation prior to its use in any analysis. Both the "good" and the "bad" performances noted within the error matrices should be mitigated by a solid qualitative evaluation of the map products based on the User's understanding of the vegetation classification system and in-depth knowledge of the on the ground conditions.

4. Literature Cited

Congalton, R.G. (1991), A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment* 37:35-46

Stehman, S.V. and Czaplewski, R.L. (1998), Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles. *Remote Sensing of Environment* 64:331-344