Region One
Vegetation Classification, Mapping, Inventory and Analysis Report

Numbered Report NRGG15-01 **Feb, 2015**

Nez Perce - Clearwater National Forests – VMap 2014 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 NPC VMap 2104 revision. For both the Lifeform and Tree Canopy Cover attributes strata were constructed for each of the available classes and 1649 Lifeform samples and 1955 Tree canopy Cover samples, well distributed across the area, were selected and assessed. For the DOM40 and Tree Size Class attributes there was a 10% withholding of the field collected training data used as a comparison to the final output for a total of 307 Tree Size class samples and 442 Tree Dominance Type samples.

For the Lifeform attribute there were 6 strata selected: Herbaceous, Shrub, Sparsely-Vegetated, Water, Deciduous Tree, and Coniferous Tree, with 1649 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.

Nez Perce-Clearwater VMap 2014 Lifeform Accuracy Assessment							
	Reference Data						
VMap Class	Grass	Shrub	Tree	Water	Sparse Veg	Grand Total	User Accuracy
Grass	182	22			2	206	88%
Shrub	66	137	7	2	10	222	62%
Tree	2	3	401	2	5	413	97%
Water			1	417	4	422	99%
Sparse Veg	5		4	5	372	386	96%
Grand Total	255	162	413	426	393	1649	
Producer Accuracy	71%	85%	97%	98%	95%		92%

Table 4. NPC VMap 2014 Lifeform 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 500 sample sites selected within each strata. 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 5 below.

Nez Perce-Clearwater VMap 2014 Tree Canopy Cover Accuracy Assessment						
Vmap Class	**Low**	**Moderate-Low**	**Moderate-High**	**High**	**Grand Total**	**User Accuracy**
Low	**440**	33	6	0	479	**92%**
Moderate-Low	41	**407**	38	1	487	**84%**
Moderate-High	10	27	**430**	27	494	**87%**
High	3	8	48	**436**	495	**88%**
Grand Total	494	475	522	464	**1955**	**Overall Accuracy**
Producer Accuracy	**89%**	**86%**	**82%**	**94%**		**88%**

Table 5. NPC VMap 2014 Tree Canopy Cover Error Matrix

There were 8 classes produced for DOM40 on the NPC: MX-PIPO, MX-PSME, MX-PICO, MX-ABLA, MX-PIEN, MX-THPL, MX-TSME, and MX-ABGR. The withheld data corresponding to each type were compared with the existing map. The results are displayed in Table 6 below.

Nez Perce-Clearwater VMap 2014 DOM40 Accuracy Assessment										
	Reference Data									
VMap Class	**MX-PIPO**	**MX-PSME**	**MX-ABGR**	**MX-PICO**	**MX-ABLA**	**MX-PIEN**	**MX-THPL**	**MX-TSME**	**Grand Total**	**User Accuracy**
MX-PIPO	**61**	**7**	**1**	**1**			**1**		**71**	**86%**
MX-PSME	**6**	**45**	**4**			**2**	**2**		**59**	**76%**
MX-ABGR	**3**	**9**	**112**			**3**	**10**		**137**	**82%**
MX-PICO				**54**	**3**			**1**	**58**	**93%**
MX-ABLA		**1**			**27**	**3**		**1**	**32**	**84%**
MX-PIEN		**2**		**1**	**4**	**25**			**32**	**78%**
MX-THPL			**2**			**2**	**27**		**31**	**87%**
MX-TSME					**1**	**1**		**20**	**22**	**91%**
Grand Total	**70**	**64**	**119**	**56**	**35**	**36**	**40**	**22**	**442**	**Overall Accuracy**
Producer Accuracy	**87%**	**70%**	**94%**	**96%**	**77%**	**69%**	**68%**	**91%**		**84%**

Table 6. NPC VMap 2014 DOM40 Error Matrix

For the Tree Size Class assessment there were 5 classes produced: Seedling Tree (0-4.9" DBH), Small Tree (5-9.9" DBH), Medium Tree (10-14.9" DBH), and Large Tree (15 – 19.9" DBH), and Very Large Tree (20"+ DBH), with 307 sample sites withheld for comparison with the existing Map. The results are displayed in Table 7 below.

Nez Perce-Clearwater VMap 2014 Tree Size Class Accuracy Assessment							
	Reference Data						
VMap Class	Seedling /Sapling	Small	Medium	Large	Very Large	Grand Total	User Accuracy
Seedling /Sapling	32					32	100%
Small		96				96	100%
Medium			144			144	100%
Large			2	155	1	158	98%
Very Large				1	49	50	98%
Grand Total	32	96	146	156	50	480	
Producer Accuracy	100%	100%	99%	99%	98%		99%

Table 7. NPC VMap 2014 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.

Unfortunately all of the attributes do not lend themselves to confident interpretation through image analysis, hence the need for ground collected data to use as comparison against the Tree Size and Tree Dominance as these two classes are very difficult to reliably and consistently estimate using image interpretation alone. The advantage to the withholding approach is that there is more confidence that can be placed in the assessed site. The disadvantage is that the sample design may not entirely capture the full range of variability across the landscape. In other words, you can only really assess well those areas that were visited. In an attempt to mitigate this FIA data sites were included in both the classification and the withholding data set so that the more inaccessible lands still received representation.

In general, the accuracies exhibited in the VMap 2014 database are exceptional. Those classes with higher error rates, i.e., MX-THPL and MX-PIEN, tend to be those types that tend to occur in a mixed setting or frequently grade in and out in abundance with other species such that a mislabeled polygon could still be considered "OK" in most analysis situations. The same can be said of the Tree Canopy and Tree Size Class attributes, where most of the error is in the 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 the two classes so that the polygon would then be assessed as incorrect.

Another thing to note, the 99% overall accuracy estimated for the Tree Size class is questionable and is likely attributed to the amount of manual editing that went into the Tree Size map. When one takes into account that people are more readily able to edit areas that they have visited in the field and that the validation dataset is mostly comprised of those same areas it becomes likely that some of those same sites get edited. The original, pre-edit error matrix is shown in Table 8 below and exhibits an acceptable accuracy for a Tree Size classification. In reality, the "true" percentage likely lies somewhere in between these two estimates.

Nez Perce-Clearwater Pre-Edit Tree Size Class Accuracy Assessment							
	Reference Data						
VMap Class	Seedling /Sapling	Small	Medium	Large	Very Large	Grand Total	User Accuracy
Seedling /Sapling	28	2		1		31	90%
Small	5	34	4		1	44	77%
Medium		14	37	16	3	70	53%
Large		4	27	78	17	126	62%
Very Large			12	5	19	36	53%
Grand Total	33	54	80	100	40	307	
Producer Accuracy	85%	63%	46%	78%	48%		64%

Table 8. Pre-edit Error Matrix for the NPC Tree Size Class VMap 2014.

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 it's 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