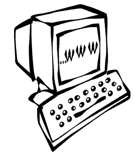
Region One
Vegetation Classification, Mapping,
Inventory and Analysis Report





Numbered Report 11-03 **April, 2011**

Beaverhead-Deerlodge National Forest VMap Accuracy Assessment; Version 11

Stephen R. Brown Jr.

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. 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 that typically increases 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. Statistical validity of the sample, however, is most easily maintained through a random selection of sites which can make the acquisition of reference data both cost and time prohibitive. To overcome this limitation a method has been devised that incorporates a random sample selection with field site visits, photo-interpretation, and existing plot data through the use of aerial resource photography and Forest Inventory and Analysis (FIA) plot data. Forest Inventory and Analysis data have been collected in a standardized, grid-like fashion across the United States for approximately 70 years. Data collected by FIA contain information about forest characteristics such as species composition, size-class, canopy coverage, health, and growth rates to name just a few. Having been collected in a consistent manner and distributed across the landscape as a network of points the information recorded by the FIA program provides a base from which an independent, systematic, assessment of VMap class accuracy can be conducted. The FIA data is not, however, collected for mapping purposes and is not directly comparable to the VMap product. It can, though, be intersected with VMap polygons to produce the random sample selection needed and then be used to inform an analyst as to the general composition of the stand in question and guide them in the photo-interpretation process.

After completion of the photo-interpretation process for all FIA intersected 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**

For a more detailed description of the accuracy assessment process used to complete the eastside R1 VMap accuracy assessment see document 'R1-VMap Accuracy Assessment Procedures for Region 1', Vanderzanden et al, 2009. CMIA # 09-11.

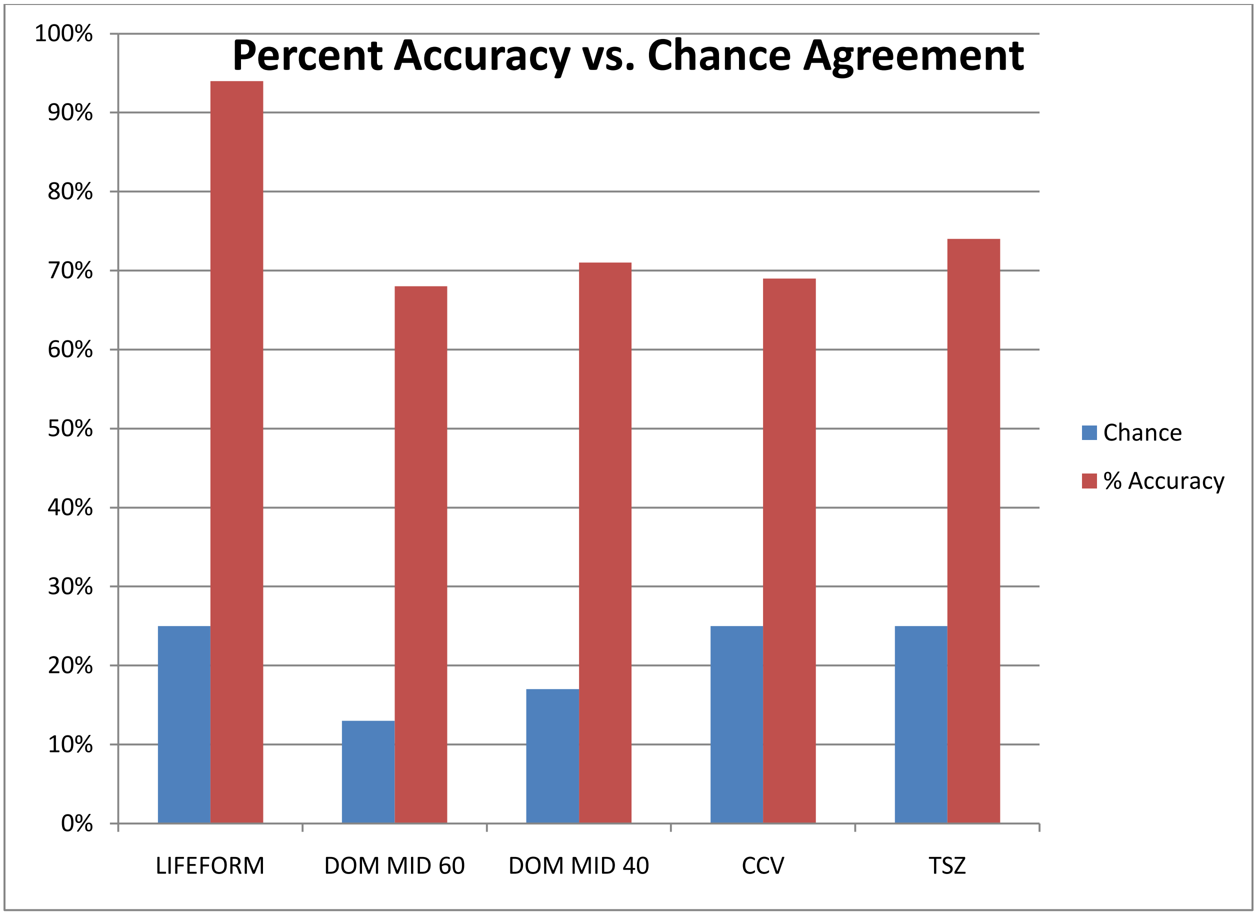
2. Results and Discussion

For the Beaverhead-Deerlodge (B-D) VMap accuracy assessment, there were a total of 544 samples available for assessment. Of these, 396 of the samples are forested and the remainder are non-forested. Of the non-forested samples, 71 could reliably be labeled as either herbaceous, shrub, or sparsely vegetated. It is possible, then, to conduct a lifeform accuracy assessment beyond simple tree/non-tree. Overall the resulting map products created for the landscapes encompassed by the B-D VMap show the highest accuracies heretofore achieved by the Northern Region Geospatial Group.

In each of the forested analysis areas, error matrices have been constructed for the mid level dominance plurality classes (DOM_MID_60, DOM_MID_40), four classes of tree canopy cover (10-24.9%, 25-39.9%, 40-59.9%, 60+%), and four classes of tree size (0-4.9", 5-9.9", 10-14.9”, = 15+".)

Overall Accuracy is a measure of the agreement between the sampled sites and mapped classes corresponding to those sites. It is simply the sum of the number of sites that agree divided by the total number of sites that were compared. As such, Overall Accuracy says nothing about individual class accuracy; rather it provides the interpreter with a measure of classification quality as a whole. It is important to consider that the value of this measure is influenced by the number of comparisons that are made in each of the classes. This can be overcome by either making the sample size the same for each class or by normalizing the elements of the error matrix. To be meaningful, each class being compared would have at least thirty samples. When such criteria are not met, assessment of classes with small sample sizes is not very meaningful, or realistic, and the Overall Accuracy statistic is the only remaining measure of map accuracy with any reliability.

Oftentimes Users do not have realistic expectations of what an acceptable level of map accuracy should be. Accuracy is generally evaluated based one’s inherent familiarity with the academic system of grading. This is a flawed comparison, however, as map accuracy is not a static variable but changes in meaning with both application and the number of classes that are being represented. A more useful interpretation of map accuracy, then, would be a comparison to the probability of chance agreement between classes. For example, a map of with 12 classes with an accuracy of 65% seems fairly limited based on an academic scale, and seems only somewhat better than flipping a coin (which has a 50% chance of getting the right answer). However when that is compared to chance agreement (which would be ~8% with 12 classes) it is seen that such a map with 65% accuracy is actually 8 times better than chance and provides the User with a high degree of confidence in the placement of classes across the map. Graph 1, below, shows a comparison of the Overall Accuracy for each VMap product to the probability of chance agreement based on the number of classes. A full discussion of the individual product accuracies follows.



Graph 1. Percent Overall Accuracy versus Chance Agreement by VMap Product.

2a. Lifeform Accuracy Assessment

Of the reference data collected for accuracy assessment approximately 73% (396 out of 544 samples) belong to the Tree lifeform class. Because the FIA plot design was the basis for our site selection, it can be safely estimated then that ¾ of the landscape is forested and ¼ of the landscape is either rangeland or sparsely vegetated. This is also indicative that the accuracy estimate for the Tree lifeform is highly reliable.

It can be seen in Table 4 below that the Overall accuracy estimate for the Lifeform Class of the B-D VMap is 94.4%. This is exceptional accuracy given the size and complexity of the B-D landscape. The matrix does show, however, that the majority of the confusion lies between the herbaceous and shrub lifeforms, with shrub perhaps being over estimated. It is difficult to ascertain this with a high degree of certainty though as the number of samples available for comparison between these types is rather small, with only 52 herbaceous samples and 21 shrub samples. Another limiting factor is the inability to distinguish low canopy Shrub sites through photo or image interpretation. It could very well be that some of the samples labeled as herbaceous could contain enough shrubs to meet the 10% canopy minimum but are simply not able to be distinguished without field verification. If a given project or analysis hinges on the distinction between shrub and herbaceous it is recommended that additional reference data be collected so that a more rigorous comparison can be made. That being said, there are enough samples to reliably estimate that tree types are mapped correctly approximately 98% of the time and non-tree types are mapped correctly approximately 92% of the time. Due to the extreme disparity in sample size between tree and non-tree lifeforms it is not possible to calculate an area-weighted overall accuracy estimate for the Lifeform class.

Table 4. Lifeform Class Error Matrix

Life Form	Reference Data						
VMap Class	Grass	Shrub	Sparsely Vegetated	Tree	Grand Total	Mapped Abundance	User's Accuracy
Grass	37	1	1	4	43	8.9%	86.0%
Shrub	12	20		4	36	7.4%	55.6%
Sparsely Vegetated			14	1	15	3.1%	N/A
Tree	3		1	387	391	80.6%	99.0%
Grand Total	52	21	16	396	485	Overall Accuracy	
Sampled Abundance	10.7%	4.3%	3.3%	81.6%		94.4%	
Producer's Accuracy	71.2%	N/A	N/A	97.7%			

2b. Dominance 60% Plurality (DOM_MID_60) Accuracy Assessment

Based on the numbers seen in Table 5, below, it can be seen that 66% of the forested landscape is comprised of either Douglas-fir (PSME) or lodgepole pine (PICO), and that number increases to approximately 77% when mixes of the two are included in the calculation. The only other species with significant (defined as at least 30 samples) coverage is whitebark pine (PIAL). The individual User Accuracies estimated for these classes range from 61% (PIAL) to 74% (PSME), with most of the confusion being between the three main classes and the IMIX, that would incorporate the same types. Unfortunately not a lot can be said for the remaining classes in terms of hard numbers as there simply are not enough samples in those classes. It can be inferred, however, from the high numbers in the major classes, that the classes for sub-alpine fir (ABLA) and Englemann spruce (PIEN) are relatively pure and there will be few cases where PSME, PICO, PIAL, or IMIX is misidentified as either ABLA or PIEN.

Table 5. DOM_MID_60 Class Error Matrix

Tree DOM_MID_60	Reference Data										
Vmap Class	PSME	PICO	ABLA	PIEN	PIAL	POTR	IMIX	TMIX	Grand Total	Mapped Abundance	User's Accuracy
PSME	72	7	3	3	4		8		97	25.1%	74.2%
PICO	12	146	3	5	6	1	16	10	199	51.4%	73.4%
ABLA	1	2	6		1		2	1	13	3.4%	N/A
PIEN	1	2		2			1	3	9	2.3%	N/A
PIAL	1	3	1		14		2	2	23	5.9%	N/A
POTR						1			1	0.3%	N/A
IMIX	2	7	2	1	2		11	1	26	6.7%	N/A
TMIX			2	1	3		3	10	19	4.9%	N/A
Grand Total	89	167	17	12	30	2	43	27	387	Overall Accuracy	
Sampled Abundance	23.0%	43.2%	4.4%	3.1%	7.8%	0.5%	11.1%	7.0%		68%	
Producer's Accuracy	80.9%	87.4%	N/A	N/A	46.7%	N/A	25.6%	N/A			

2c. Dominance 40% Plurality (DOM_MID_40) Accuracy Assessment

There is a slight improvement in the accuracy assessment for DOM_MID_40 over DOM_MID_60 (Table 6). This is due to the definition of the class allowing for the inclusion of other species within each mixed class label. With the "Big 3" – PSME, PICO, and PIAL – from the discussion above there is seen a marked improvement for the MX-PIAL class, a gain of 10%, while the other two exhibit little change, 4% for MX-PICO and no gain for MX-PSME. One difference is that there are now enough samples to be able to start to say something about MX-ABLA, unfortunately it is not all good. Basically there is a lot of confusion between MX-ABLA and MX-PICO and MX-PIEN. More samples are being called MX-PICO than are correctly labeled MX-ABLA. While this is not surprising given the nature of the two species and how often they intermingle it is something to keep in mind should a given analysis concern be the precise location of MX-ABLA. It is noteworthy to mention that the estimates of abundance based on the FIA sample selection process are almost 4% higher than the VMap estimated abundance, indicating that MX-ABLA is underrepresented in the B-D VMap database. Conversely, there is an overestimate, by about 6%, in the modeling of MX-PICO. Also, it can be seen in the error matrix that there is zero confusion between the coniferous types and MX-POTR, that there are no instances where something else is mislabeled as MX-POTR. This indicates that where an object is labeled as MX-POTR one can be fairly certain it will be MX-POTR and nothing else.

Table 6. DOM_MID_40 Class Error Matrix

Tree DOM_MID_40	Reference Data								
Vmap Class	MX-PSME	MX-PICO	MX-ABLA	MX-PIEN	MX-PIAL	MX-POTR	Grand Total	Mapped Abundance	User's Accuracy
MX-PSME	74	14	3	5	5		101	26.6%	73.3%
MX-PICO	17	155	13	7	11	1	204	53.8%	76.0%
MX-ABLA	3	2	8		7		20	5.3%	N/A
MX-PIEN	1	3	7	5	2		18	4.7%	N/A
MX-PIAL	1	4	3	2	25		35	9.2%	71.4%
MX-POTR						1	1	0.3%	N/A
Grand Total	96	178	34	19	50	2	379	Overall Accuracy	
Sampled Abundance	25.3%	47.0%	9.0%	5.0%	13.2%	0.5%		71%	
Producer's Accuracy	77.1%	87.1%	23.5%	N/A	50.0%	N/A			

2d. Tree Canopy Cover Accuracy Assessment

The tree canopy cover error matrix (Table 7) shows an error distribution that is fairly typical of categorized variables, with most of the confusion existing between the adjacent classes. This is not surprising given that field data collection protocols only require that accuracies be within plus or minus one class. Also, a review of the FIA plot data reveals that much of the forested area is right on the edge of a class, rarely at the midpoint. For example, many stands show a canopy cover estimate of 42%, which is just inside the 40-59.9% class but may be easily confused with the upper end of the 25-39.9% class. All in all, though, the tree canopy cover map product performs very well, especially when the Mountain Pine Beetle (MPB) effects on the B-D are taken into account. This required separate modeling of the tree canopy cover, 1 model for trees that are beetle affected and 1 model for trees that are not, in order to achieve acceptable levels of accuracy. It is seen from the table that overall the estimates for each class are close, with the largest disparity coming in the 25-39.9% class. This is easily attributable to the effects of the MPB on existing canopy as trees may be affected, and hence have a dead/dying canopy, that a field estimate would not necessarily capture.

Table 7. Tree Canopy Cover Class Error Matrix

Tree Canopy Cover	Reference Data						
VMap Class	**10-25%**	**25-40%**	**40-60%**	**60%+**	**Grand Total**	**Mapped Abundance**	**User's Accuracy**
10-25%	**29**	6	6		**41**	**10.6%**	**70.7%**
25-40%	11	**65**	37	2	**115**	**29.8%**	**56.5%**
40-60%	9	12	**148**	21	**190**	**49.2%**	**77.9%**
60%+		3	11	**26**	**40**	**10.4%**	**65.0%**
Grand Total	**49**	**86**	**202**	**49**	**386**	**Overall Accuracy**	
Sampled Abundance	**12.7%**	**22.3%**	**52.3%**	**12.7%**		**69%**	
Producer's Accuracy	**59.2%**	**75.6%**	**73.3%**	**53.1%**			

2e. Tree Size Class Accuracy Assessment

Once again, against the common expectation, tree size is the top performer of the VMap classes (Table 8). Presumably this can be attributed to two things. The first being the inclusion of the NAIP imagery in the classification process which adds an element of stand texture, a measure which corresponds to crown size and density, which enables the algorithms to more accurately model tree size. The other is that, based on the FIA estimates, the majority of the samples (approximately 91%) fall within 2 tree size classes (5-9.9" and 10-14.9"), where it then becomes statistically more likely to be correctly labeled. Even so, there is very good delineation between these two classes with individual User accuracies of 81% and 74% respectively. There are not enough samples of the other two classes to be able to say anything of merit.

Table 8. Tree Size Class Error Matrix

Tree Size Class	Reference Data						
Vmap Class	0-5"	5-10"	10-15"	15"+	Grand Total	Mapped Abundance	User's Accuracy
0-5"	**15**	3	7		**25**	**6.5%**	**N/A**
5-10"	11	**149**	22	4	**186**	**48.1%**	**80.1%**
10-15"	2	37	**119**	2	**160**	**41.3%**	**74.4%**
15"+		4	11	**1**	**16**	**4.1%**	**N/A**
Grand Total	**28**	**193**	**159**	**7**	**387**	**Overall Accuracy**	
Sampled Abundance	**7.2%**	**49.9%**	**41.1%**	**1.8%**		**73.4%**	
Producer's Accuracy	**N/A**	**77.2%**	**74.8%**	**N/A**			