Protocol for Quality Assurance and Accuracy Assessment of Ecosystem Maps

July 2000

prepared by

Del Meidinger

Research Branch,
B.C. Ministry of Forests,
Victoria, BC

for the

TEM Alternatives Task Force

Table of Contents

Acknowledgements

A draft of this protocol was prepared by Del Meidinger following a workshop session (May 1999) attended by R. Sims, D. Moon, A.Y. Omule, K. Jones, B. Love, D. Clark, T. Lea and D. Meidinger. The final manuscript was prepared after a year of application. Other individuals also provided valuable input to this protocol: Marvin Eng was available for technical discussions and provided information on creating a confusion matrix; Amanda Nemec, with the advice and assistance of Omule, prepared the sampling tables; Peter Ott provided statistical advice on the analysis of the data. This report also represents the culmination of various contracts over the past few years by R. Sims, K. Iverson, the "Sampling Team" of the BCFS Vegetation Inventory Working Group, and previous work by Moon et al. Everyone's assistance is greatly appreciated. A.Y. Omule and Dave Clark also provided useful review comments.

Contributors affiliations:

Clark, D., Resources Inventory Branch, MoELP, Victoria
Eng, M., Research Branch, MoF, Victoria
Jones, K., K. Jones and Associates, Victoria
Lea, T., Resources Inventory Branch, MoELP, Victoria
Love, B., MoF, Smithers
Meidinger, D., Research Branch, MoF, Victoria
Moon, D., CDT Core Decision Technologies, Vancouver
Nemec, A., I.S.R. Corp. Brentwood Bay
Omule, A.Y.; J.S. Thrower and Associates, Victoria
Ott, P., Research Branch, MoF, Victoria
Sims, R., EBA Environmental Consultants, Vancouver

1.0 INTRODUCTION

This report outlines a protocol designed for quality assurance and accuracy assessment for large-scale ecosystem mapping (Predictive and Terrestrial Ecosystem maps – PEM and TEM), but which can be applied to any ecosystem mapping, including, Broad Ecosystem Inventory (BEI) for example. The protocol is to be used to determine the acceptability or accuracy of a map area with respect to the ecosystem units being mapped. The results of the assessment can be used for quality assurance—acceptance of the mapping for contract administration, and/or to present unbiased information to users regarding the accuracy of the mapping.

Multiple levels of assessment are presented in the protocol. It is suggested that the lower levels (1 – 3) are acceptable for most quality assurance assessments whereas the upper levels (4 – 6) are required in accuracy assessments. In all cases, it is the thematic accuracy[1] of the map that is being evaluated. All levels of assessment require an unbiased sample of the map polygons and then an assessment of the thematic content via air photo interpretation, air calls and/or ground assessments.

Users may also want to assess the reliability of the map for application to their particular requirement. The data collected using this protocol can be used to determine map reliability for selected interpretations of the ecological information.

2.0 FEATURES OF THE PROTOCOL

The objectives of this protocol are as follows:

- To provide users with an unbiased approach to assessing the thematic accuracy of ecosystem maps; and
- To provide flexibility in the protocol to allow for either a quality assurance or accuracy assessment and to accommodate various budgets.

The protocol presents a statistically unbiased approach to evaluating the acceptability or accuracy of the mapping. The thematic content of randomly selected map polygons is assessed by various means; the methods varying in precision and objectivity. These within-polygon assessment approaches are presented as multiple 'levels' of the protocol. Users select a level based upon their intended use of the data and the project budget.

The protocol also has the following features:

- The assessment is conducted after the mapping is complete.
- The entire project map area is sampled; this may be a partial or full map sheet or an entire project area.
- The project area can be stratified (e.g., alpine and below alpine), but samples must be distributed in all strata.
- The assessment is conducted by "experts" and is backed-up with field data.
- As most PEM and TEM polygons are complexes of ecosystem mapping units, the variation within polygons is assessed by multiple plots or mapping at a larger scale.
- Where a map consists of simple units, the preparation of a 'confusion matrix' of errors of omission[2] and errors of commission[3] is recommended.
- The final scores provide data on the accuracy in describing the dominant map unit components and all (up to 3) map unit components.
- A score for the accuracy of "correct" plus "close" categories can also be reported.

[1] Thematic accuracy is the correctness of polygon labeling. A polygon is correctly labeled if the attributes of the polygon fall within the defined attribute ranges of the map unit and its components.
[2] Pixel or polygon incorrectly omitted from a class (e.g., site series).
[3] Pixel or polygon incorrectly assigned to a particular class that actually belongs in another class.

- The assessment can be used for pass/fail of a contractor's work, if the protocol level and pass mark for each assessment score is specified in the contract.
- The results of the assessment are non-spatial – that is, they identify what the level of accuracy is, but they do not show, within the map or project area, where errors or inconsistencies occur.

3.0 PROTOCOL DESIGN

The protocol involves selecting a set of polygons and assessing their thematic content using the procedures outlined for the selected 'level' of assessment. The steps are as follows:
Determine which mapping entity will be assessed (i.e., site series, site modifiers, structural stage).

1. Determine level of protocol that will be applied.
2. Determine number of polygons to be assessed and select them using a simple random sample, with replacement (or a stratified random sample, with replacement).
3. Prepare a sampling plan to sample selected polygons that is consistent with the protocol level selected.
4. Conduct sampling.
5. Summarize data in a table comparing samples with map predictions.
6. Test whether site series proportions from map predictions fall within confidence intervals of that determined by samples.
7. Score each polygon, and average for the map or project (or each strata, as well), using protocol scoring to determine: 1) proportion of map where the dominant mapping entity is correct; and 2) the "overlap" score between the samples and the mapping entities.
8. If desired, also score the "close overlap" between the sample and the map unit entities allowing for "adjacent" units to score one-half that of the correct unit.

The steps are described in more detail in the following sections.

3.1 Selecting Mapping Entities for Assessment

Mapping entities can vary somewhat, depending upon the map or project objectives, available data and methods used. In TEM, the mapping entities are site series, site modifiers and structural stage. In PEM, site series are mapped, generally with slope/aspect site modifiers, with structural stage as a separate layer. Therefore, all assessments can evaluate site series mapping accuracy, however, the selection of other mapping entities for assessment will depend upon the project objectives and funding. The QA or accuracy scores should be calculated and reported separately for each mapping entity (e.g., a set of site series scores, a set of structural stage scores). In most cases, all of the mapping entities which are mapped should be assessed.

3.2 Determining Protocol Sampling Level

In order to accommodate the use of the protocol for either quality assurance or accuracy assessment, and to allow for various levels of sampling effort, six protocol sampling levels are presented in Table 1. Levels 1 – 3 are to be applied as quality assurance checks of the final mapping; levels 4 – 6 are intended for application as accuracy assessments, but may also be used for quality assurance. A level is selected by balancing the objectives with the budget available for field sampling and analysis.

TABLE 1: Quality Assurance and Accuracy Assessment Protocol Sampling Levels

Level	Primary application	Characteristics
1	Quality assurance	• 100% of sample polygons evaluated by air photo interpretation, preferably at larger scale using 'Soft-copy' technology • 10 – 25% of sample polygons field assessed by air or ground calls
2	Quality assurance	• 100% of sample polygons evaluated by air photo interpretation, preferably at larger scale using 'Soft-copy' technology • 26 – 50% of sample polygons field assessed by air and ground calls at a ratio of about 75:25
3	Quality assurance	• 100% of sample polygons evaluated by air photo interpretation, preferably at larger scale using 'Soft-copy' technology • 51 – 75% of sample polygons field assessed by air and ground calls at a ratio of about 75:25
4	Accuracy assessment	• 100% of sample polygons assessed by ground checks • either traverse polygon and map simple map entities at large scale (e.g., 1:5000) or randomly/systematically locate 3 – 5 sample plots in polygon
5	Accuracy assessment	• 100% of sample polygons assessed by ground checks • randomly/systematically locate 6 – 20 sample plots in polygon
6	Accuracy assessment	• 100% of sample polygons assessed by ground checks • randomly/systematically locate 21 – 50 sample plots in polygon

In moving from level 1 to 6, there is generally an increase in statistical rigour and precision in the assessment of sample polygons. Sample polygon selection in all levels is statistically unbiased but there is more objectivity to the within-polygon assessment with increasing level. The number of polygons sampled is the most important variable in determining the confidence and precision of the estimates, however, the method of within-polygon evaluation is also important.

3.3 Determining Sample Size and Selecting Polygons

Irrespective of protocol level, a random sample of polygons is selected. The larger the sample size, the greater the precision of the estimates. Table 2 suggests approximate minimum sample sizes for estimating a proportion of correctly classified polygons based on a selected confidence level, maximum error of the estimated proportion, and the probability of a polygon being correctly classified. The sample size is for the whole map or project area and will result in a map or project area estimate. If estimates are required for each map entity, the sample size must be multiplied by the number of map entities. This is likely to be a huge number so most assessments will be for the whole project area.

For example, if you wanted to have a maximum error of +/- 5% on the estimate of accuracy, 9 times out of 10, and you didn't know the probability of any point being correctly mapped, you would look up in the table a confidence value of 0.90, a maximum error of 0.05, and the lowest probability of a correct classification of 0.5. The value for (0.90, 0.05, 0.5) is 273. You would therefore select 273 polygons.

For TEM or PEM accuracy assessment, we recommend a confidence level of 0.80 and a maximum error value of 0.07 (7%). Without information to choose a probability of a correct point value larger than 0.5 (50%), p = 0.5 should be used. However, if an adjacent area has been assessed to be 0.75 or higher and the new area has been mapped using the same procedures, the same individuals, and is ecologically similar, the higher 'p' value can be used. This means that in most cases, a sample of 86 polygons is required for a map assessment. This may only be a small proportion of the total polygons, but the cost of ground sampling every polygon is quite high.

TABLE 2: Approximate Sample Sizes for Assessment of Map Accuracy

Confidence level	Maximum error (+/-)	Probability random point correctly classified on map					
		0.5	0.75	0.8	0.9	0.95	0.99
0.80	0.01	4108	3081	2630	1480	782	164
0.80	0.02	1028	772	659	371	197	42
0.80	0.03	458	344	294	166	89	20
0.80	0.04	258	194	166	94	51	12
0.80	0.05	166	125	107	61	33	8
0.80	0.06	116	87	75	43	24	6
0.80	0.07	86	65	55	32	18	5
0.80	0.08	66	50	43	25	14	5
0.80	0.09	53	40	34	20	12	4
0.80	0.10	43	33	28	17	10	2
0.90	0.01	6766	5075	4331	2437	1287	270
0.90	0.02	1693	1271	1085	611	324	69
0.90	0.03	754	566	483	273	145	32
0.90	0.04	425	319	273	155	83	19
0.90	0.05	273	205	176	100	54	13
0.90	0.06	190	143	123	70	38	10
0.90	0.07	140	106	91	52	29	8
0.90	0.08	108	82	70	40	22	7
0.90	0.09	86	65	56	32	18	6
0.90	0.10	70	53	46	27	15	5
0.95	0.01	9607	7206	6149	3460	1828	383
0.95	0.02	2404	1804	1540	867	459	98
0.95	0.03	1070	803	686	387	206	45
0.95	0.04	603	453	387	219	117	27
0.95	0.05	387	291	249	141	76	18
0.95	0.06	270	203	174	99	54	14
0.95	0.07	199	150	128	73	40	11
0.95	0.08	153	115	99	57	31	9
0.95	0.09	121	92	79	46	25	8
0.95	0.10	99	75	64	38	21	7

Another test, a chi-squared test, is conducted to determine whether the site series proportions observed (from the samples) are different from those obtained by the mapping (the expected). Appendix 1 presents the assumptions and explanation of the procedure for determining sample size. What is apparent is that a sample size of 80 – 90 polygons will be reasonable for the chi-squared test.
Polygons are selected as a simple random sample, with replacement. Although polygons can be selected using other methods, like probability proportional to area (with replacement), or by a simple random sample of points (where the polygon that the point falls in is sampled). However, the simple random sample of polygons is easy to do and is appropriate for the variable being estimated (the proportion of 'correct' polygons).

It may be desirable to stratify the map into a few strata and sample some strata more than others, however, all strata require some sampling. Examples of useful strata include:

- alpine and below alpine, where the alpine is a large proportion of the map area and you want to have more data, proportionately, from the forested areas;
- high-elevation (inaccessible areas) vs. low-elevation, where you suspect that inaccessible areas may have a lower accuracy; in this case you may sample each strata with a reasonable number of samples.

3.4 Preparing a Sampling Plan

A sampling plan is important for two reasons:

1. It clarifies the assessment objectives, protocol sampling level, sample size, etc., and,
2. It provides for the efficient conduct of field work.

The protocol has many options including: accuracy assessment vs. quality assurance, sampling level, stratification, sample size, method of polygon assessment, number of plots per polygon, to name a few. Therefore, the objectives of the assessment need to be clearly stated and the sampling decisions outlined. This will allow for straightforward planning of the field work. The sampling plan should contain the following information:

- A description of the area to be assessed (e.g., administrative unit, map sheets, area)
- Objectives
- Protocol sampling level (from Table 1)
- Sample accuracy decisions (confidence level, maximum error, 'p' value, strata)
- Sampling decisions (number of polygons to sample, method of selection, number to be field checked, method of field assessment, description of strata)
- Any deviations from this protocol
- For each polygon, a sampling plan for the within-polygon assessment and, if polygons to be field sampled, the access plan (e.g., air photos, UTM coordinates, closest access, distance and bearing from access point, difficulty of access).

3.5 Sampling

Polygon assessment can include air photo interpretation, air calls, ground calls, large-scale mapping, or plot sampling (see Table 1). Each method is discussed below.

3.5.1 Assessment using air photo interpretation

Assessing polygons using air photo interpretation is a viable procedure for quality assurance, if the following conditions can be met:

- for TEM, larger scale photos or 'soft-copy' assessment, where photos can be enlarged, and
- for PEM, preferably larger scale photos or 'soft-copy' assessment, but even same scale photos are OK.

The entire polygon is assessed for map entity composition and proportions.

3.5.2 Assessment using air calls

Assessing polygon map entity composition and proportions using air calls is possible by hovering over the polygon in a helicopter. It is sometimes possible to do an 'air call' from a viewpoint.

3.5.3 Assessment using ground calls

In the case of polygon assessment, a ground call is conducted by traversing enough area of a polygon to assess map entity composition and proportions.

3.5.4 Assessment using large-scale mapping

Mapping a polygon at a larger scale, using procedures similar to that for Silviculture Prescriptions, is another way of estimating map entity composition and proportions. Procedures are presented in B.C. Forest Service field guides to site identification and in the Silviculture Prescriptions Field Methods Book (URL: http://www.for.gov.bc.ca/hfd/pubs/docs/sil/sil411/httoc.htm).

3.5.5 Assessment using plot sampling

Sample plots need to be randomly or systematically located within each polygon. A suggested procedure is to locate samples on a square grid, starting at a random location within the polygon. Plots are installed at the grid intersections. The spacing between the sample plots depends on the number of plots to be installed in the polygon and can be determined by the following formula: grid spacing in metres = sqrt{(10,000* Polygon Area in ha)/(no of plots)}. You need only the starting point and the grid spacing.
Another procedure is to plot a transect through the polygon, and for each required sample, randomly select a distance along the transect, a direction off the transect (right or left), and a distance perpendicular to the transect.

If planned in the office, the field stage is fairly straightforward, the key being to locate accurately the start of the grid or transect and to correct distances for slope.

At each sample plot, data to identify the mapping entities (site series and possibly site modifiers and structural stage) is collected using the Ground Inspection Form, following the standards in Field Manual for Describing Terrestrial Ecosystems[4]. Data are collected in a 10 m radius circular plot (330 sq. m.) and the mapping entities and their proportions in each plot are recorded.

See Table 3 for a list of attributes to be collected on each site. If the protocol is being implemented to determine the reliability of secondary interpretations, attributes reflecting those interpretations should also be collected.

TABLE 3. Attributes to be Collected for Plot Samples using Ground Inspection Form

1. Air photo number	19. Depth to and type of restricting layer (if applicable)
2. Date	20. Coarse fragment content
3. Project ID	21. Terrain texture, surficial material, and surface expression
4. Surveyor(s)	22. BCG unit
5. Mapsheet	23. Ecosection
6. Plot no. (polygon + sample)	24. Site series
7. Polygon no.	25. Site modifiers
8. Lat./Long. or UTM	26. Structural stage
9. Aspect	27. Crown closure
10. Elevation	28. Using Ecosystem Polygon Summary section, record each site series in plot and their proportions (% of plot))
11. Slope	29. Total % cover by stratum
12. Soil moisture regime	30. Dominant/indicator plant species
13. Soil nutrient regime	31. % cover of dominant/indicator species
14. Meso slope position	32. Complete or partial
15. Drainage – mineral or organic soils	33. Notes
16. Mineral or organic soil texture	
17. Surface organic horizon thickness	
18. Humus form (to order level)	

[4] Province of B.C. 1998. Field manual for describing terrestrial ecosystems. B.C. Ministry of Forests, and B.C. Ministry of Environment, Lands, and Parks. Land Management Handbook No. 25.

3.6 Summarizing Field Data

After fieldwork, compile and summarize the data and record for each polygon the map entities and their proportions from sampling and from the map polygon data. For the plot sampling, summarize the proportion of mapping units in each plot and compile for the polygon. In all cases, record the polygon proportions to the nearest 10% (or decile). Table 4 presents an example.

Similar tables should be prepared for structural stage or site modifiers, if they are mapped. Polygon areas will be required to calculate the proportional area of each map entity.

TABLE 4: Example of Compiled Field Sample and Map Polygon Site Series Data for a TEM Project

	Map polygon data									Sample polygon summary								
TEM Poly #	Zone	Subzone	Vrt	Dec. 1	SS 1	Dec. 2	SS 2	Dec. 3	SS 3	Zone	Subzone	Vrt	Dec. 1	SS 1	Dec. 2	SS 2	Dec. 3	SS 3
2184	CWH	vm	1	10	AB					CWH	Vm	1	5	AS	5	AB		
2181	CWH	vm	1	7	AB	3	HS			CWH	Vm	1	6	AB	2	AS	2	AF
2205	CWH	vm	1	8	AF	2	SD			CWH	Vm	1	8	AB	2	AS		
2218	CWH	vm	1	10	AB					CWH	Vm	1	9	AB	1	AS		
2183	CWH	vm	1	8	AB	2	AS			CWH	Vm	1	7	AS	3	HD		
2172	CWH	vm	1	10	AB					CWH	Vm	1	8	AB	2	HD		
2163	CWH	vm	1	7	AB	3	HS			CWH	Vm	1	10	AB				
2176	CWH	vm	1	6	AB	4	AF			CWH	Vm	1	8	AS	2	AF		

3.7 Use of Data (or 'Scoring')

Data from the assessment will be used to provide map users with information on the accuracy of the main mapping entities and to compare the accuracy of different mapping methods (e.g., TEM and PEM). In order to meet these objectives, one statistical test and several 'scores' will be calculated and summarized. These are as follows:

1. Chi-squared statistical test of whether map entity proportions determined from map predictions (the expected values) are different from those determined by sampling (observed values);
2. The proportion of sample polygons where the dominant entity mapped is the same as that determined by sampling;
3. The 'percent overlap' in the entities mapped versus those determined by sampling;
4. The 'percent acceptable overlap' where "acceptable" mapped entities are compared to those determined by sampling; and
5. A 'confusion matrix' where mapped entities are compared to sampled values.

Details on these statistics are presented below.

3.7.1 Testing ecosystem unit proportions

To perform this test, first calculate the relative area of each map entity from the map and from the polygon samples. In both cases, the proportional area of each map entity within a polygon is used to calculate the proportion of each map entity over the map. The area of each polygon is required to determine the proportional area of each polygon covered by a map entity. The proportions are then presented as a frequency by standardizing to the total number of polygons sampled. The same total number of polygon samples is used for the map (expected) frequencies as for the sample (observed) frequencies, in order to allow for comparison and calculation of the chi-squared statistic.

The frequencies are then compared by a Chi-squared test with the null hypothesis being that there is no difference in the map entity proportions. We suggest an alpha value of 0.1. The sample size should be sufficiently large to ensure that most of the expected frequencies are >5. If 'zero' frequencies occur, pool with other low-frequency map entities so that all expected (map) frequencies are greater than zero.

In the example in Table 5, the null hypothesis is rejected, so the map and sample proportions are significantly different.

TABLE 5: Example of Test of Site Series Proportions

Map entities	**Sample (observed)**			**Map (expected)**				
ESSF	**Area (ha)**	**Prop.**	**Freq.**	**Area (ha)**	**Prop.**	**Freq.**		**(Observed-expected)^2/Expected**
mc/01	1074	0.43	87	783	0.28	55.93		17.26190476
mc/02	28	0.01	2	126	0.05	9.00		5.444444444
mc/03	114	0.05	9	174	0.06	12.43		0.945812808
mc/04	261	0.11	21	400	0.14	28.57		2.006428571
mc/05	64	0.03	5	112	0.04	8.00		1.125
mc/06	409	0.17	33	565	0.20	40.36		1.341213654
mc/07	100	0.04	8	200	0.07	14.29		2.765714286
mc/08	295	0.12	24	200	0.07	14.29		6.605714286
mc/09	34	0.01	3	100	0.04	7.14		2.402857143
mc/10	78	0.03	6	40	0.01	2.86		3.457142857
mc/TA	20	0.01	2	100	0.04	7.14		3.702857143
	2477	1.00	200	2800	1.00	200.00	Chi-square =	47.06 with 10 df[1]
							p-value[2] =	9.21188E-07

[1] The degrees of freedom = number of entities (after pooling) - 1
[2] p-value can be calculated using 'CHIDIST' function in Microsoft Excel

3.7.2 Determining proportion of dominant entity "correct"

Using the compiled data (e.g., Table 4), score each polygon as to whether the dominant entity from the map corresponds to the dominant entity from the polygon sampling. See Table 6 for an example. In the case of a tie (e.g., 50:50 or 40:40:20), consider the mapping correct half the time, incorrect the other half.

Report the statistic as the '% dominant correct' to the nearest percent, +/-95% confidence interval (CI) around the estimate (using the binomial distribution).

Using Table 6 results, the '% dominant correct' is 5 out of 8 (x 100) = 62 %. The single tie (first row) was scored 'correct'; if another tie was found, it would be scored 'incorrect.' The confidence interval can be calculated using the 'CRITBINOM' function in Microsoft Excel. In the example above, where the sample size is only 8, the lower 95% confidence interval value is 25% and the upper 87%. If 100 polygons were sampled, the lower value would be 56% and the upper 69%.

TABLE 6: Example of TEM Data and Evaluation of Dominant Site Series

	Map polygon data									Sample polygon summary									Score
TEM Poly #	Zone	Subzone	Vrt	Dec. 1	SS 1	Dec. 2	SS 2	Dec. 3	SS 3	Zone	Subzone	Vrt	Dec. 1	SS 1	Dec. 2	SS 2	Dec. 3	SS 3	Dominant correct
2184	CWH	vm	1	10	AB					CWH	vm	1	5	AS	5	AB			Y
2181	CWH	vm	1	7	AB	3	HS			CWH	vm	1	6	AB	2	AS	2	AF	Y
2205	CWH	vm	1	8	AF	2	HD			CWH	vm	1	8	AB	2	AS			N
2218	CWH	vm	1	10	AB					CWH	vm	1	9	AB	1	AS			Y
2183	CWH	vm	1	8	AB	2	AS			CWH	vm	1	7	AS	3	HD			N
2172	CWH	vm	1	10	AB					CWH	vm	1	8	AB	2	HD			Y
2163	CWH	vm	1	7	AB	3	HS			CWH	vm	1	10	AB					Y
2176	CWH	vm	1	6	AB	4	AF			CWH	vm	1	8	AS	2	AF			N

3.7.3 Determining percent overlap

Using the compiled data (e.g., Table 4), assess the '% overlap' by comparing the map entity proportions to the sample polygon proportions. Assume that the sample is correct. Present the score as the mean of the % overlap comparisons, to the nearest percent.

Table 7 presents an example. In the first polygon, the polygon was labeled 100% AB and the sample indicated that the polygon was 50% AB. The overlap would be scored as 50% (i.e., the mapped site series had 50% of the sampled site series for the polygon). The mean '% overlap' score for the eight polygons is 390/8 = 49%.

TABLE 7: Example of TEM Data and Evaluation of Percent Overlap

	Map polygon data									Sample polygon summary									Score
TEM Poly #	Zone	Subzone	Vrt	Dec. 1	SS 1	Dec. 2	SS 2	Dec. 3	SS 3	Zone	Subzone	Vrt	Dec. 1	SS 1	Dec. 2	SS 2	Dec. 3	SS 3	% Overlap
2184	CWH	vm	1	10	AB					CWH	vm	1	5	AS	5	AB			50
2181	CWH	vm	1	7	AB	3	HS			CWH	vm	1	6	AB	2	AS	2	AF	60
2205	CWH	vm	1	8	AF	2	HD			CWH	vm	1	8	AB	2	AS			0
2218	CWH	vm	1	10	AB					CWH	vm	1	9	AB	1	AS			90
2183	CWH	vm	1	8	AB	2	AS			CWH	vm	1	7	AS	3	HD			20
2172	CWH	vm	1	10	AB					CWH	vm	1	8	AB	2	HD			80
2163	CWH	vm	1	7	AB	3	HS			CWH	vm	1	10	AB					70
2176	CWH	vm	1	6	AB	4	AF			CWH	vm	1	8	AS	2	AF			20

Where all map units are simple (i.e., one map entity), it is still possible to do this assessment a the sample polygon assessments may not be of one map entity.

3.7.4 Determining percent acceptable overlap

Using the compiled data (e.g., Table 4), the '% acceptable overlap' can be assessed by comparing the map entity proportions to the sample polygon proportions, as in Section 3.7.3, but by scoring half the overlap score for adjacent classes to the correct one. In the case of site series, this would be adjacent site series on the edatopic grid. In the case of structural stage, this would be one class younger or older. It is best to prepare a matrix of "acceptable adjacent" site series to assist in scoring as not all site series that are adjacent on an edatopic grid should be acceptable. For example, a rich floodplain site series may be adjacent to a poor swamp ecosystem on the grid, but they should not be confused in mapping and have very different interpretations, so a partial score should not be given for mapping one in place of the other. The intent of this score is to allow some flexibility in evaluation of mapping where similar site series are difficult to differentiate accurately.

Present the score as the mean of the % acceptable overlap comparisons, to the nearest percent.

Table 8 presents an example. In the first polygon, the map label was 100% AB; the sample indicated that the polygon was 50% AB, 50% AS. The overlap score would be 50% for the AB portion, and zero for the AS portion, as the AS site series is diagonal on the edatopic grid for the CWHvm1 but is not adjacent. In the second polygon in the table, the overlap score is 60% with an additional 10% (one-half of 20%) added for the 'acceptable' comparison between the AF and AB site series. The AF site series is adjacent to the AB site series on the edatopic grid for the CWHvm1. The mean '% acceptable overlap' score for the eight polygons is 500/8 = 62.5%.

TABLE 8: Example of TEM Data and Evaluation of Percent Acceptable Overlap

	Map polygon data									**Sample polygon summary**									**Score**
TEM Poly #	Zone	Subzone	Vrt	Dec. 1	SS 1	Dec. 2	SS 2	Dec. 3	SS 3	Zone	Subzone	Vrt	Dec. 1	SS 1	Dec. 2	SS 2	Dec. 3	SS 3	% Accept. Overlap
2184	CWH	vm	1	10	AB					CWH	vm	1	5	AS	5	AB			50
2181	CWH	vm	1	7	AB	3	HS			CWH	vm	1	6	AB	2	AS	2	AF	70
2205	CWH	vm	1	8	AF	2	HD			CWH	vm	1	8	AB	2	AS			50
2218	CWH	vm	1	10	AB					CWH	vm	1	9	AB	1	AS			90
2183	CWH	vm	1	8	AB	2	AS			CWH	vm	1	7	AS	3	HD			35
2172	CWH	vm	1	10	AB					CWH	vm	1	8	AB	2	HD			90
2163	CWH	vm	1	7	AB	3	HS			CWH	vm	1	10	AB					85
2176	CWH	vm	1	6	AB	4	AF			CWH	vm	1	8	AS	2	AF			30

3.7.5 Creating a confusion matrix

Where map entities are simple (i.e., one per polygon), a confusion matrix can be prepared. Where map entities are complexes of mapping entities, it can be difficult or impossible to prepare a confusion matrix as it may not be clear which ecosystems are being 'confused' with others.

A confusion matrix, also known as an error or contingency matrix, is commonly used in the process of accuracy assessments for resource classifications. The process is well developed for remote sensing and satellite image analysis. A confusion matrix can be used to estimate the overall accuracy of a classification, the average accuracy for all entities in the classification, or the specific accuracy for a given item. The last two estimates require the concept of errors of omission and commission, described below.

To create a "**Confusion Matrix**":

- list the categories of "things" in the classification system in both the row and column headings;
- label either one as "Known to be" (usually from ground-based plots in natural resource mapping);
- label the other as "Classified as" (usually air photo interpretation or digital image classification in natural resource mapping);
- enter each "sample" (usually from ground-based plots in natural resource mapping) in the appropriate cell (i.e., what it was known to be and what it was classified as);and
- sum the rows and columns

- Commission errors are represented by non-diagonal elements in the table where a known "thing" (a ground plot) is into a category in which it does not belong. In other words you committed the act of getting it wrong.
- Omission errors are represented by non-diagonal elements where the known "thing" is not classified into a category in which it does belong. In other words you omitted the act of getting it right.
- Each commission error in a given category is also an omission error for a different category. Therefore the overall error is the sum of the off-diagonal elements divided by the total number of elements.

For example:

	Known to be:				
Classified as:	Bog	Fen	Marsh	total	Omission
Bog	10	1		11	9%
Fen	5	10		15	33%
Marsh	1	1	5	7	28%
Total	16	12	5	33	
Commission	37%	17%	0%		

Overall error (8/33) =24%
Average Omission Error =23%
Average Commission Error =27%

- Commission Errors: Six bogs were misclassified (5 as fens and 1 as a marsh). Two fens were misclassified (1 as a bog and 1 as a marsh). No "known" marshes were misclassified.
- Omission Errors: Of those classified as bogs, 1 was a fen; of those classified as fens, 5 were bogs; and of those classified as marshes, 1 was a bog and 1 was a fen.
- Each Error Is Both: For the "known" bog "classified" as a marsh (lower left cell of the table) we committed the error of not classifying it as a bog. Our omission error was to classify him as a marsh.

In summary, for this example; we always know a marsh when we see one but on rare occasions we think a bog or a fen may be a marsh.

.

From a confusion matrix, the overall 'accuracy' can be determined by assessing the diagonal. In the example above, 25/33 or 76% of the ecosystems were classified correctly. In mapping, a certain number of correct classifications can occur by chance, even in the most uncertain situations. Therefore, a value called the 'Kappa index'[5] is often calculated to express the degree of agreement but taking into account the chance 'correct' classifications. A procedure for calculating the Kappa index is found at http://biology.usgs.gov/npsveg/aa/sect5.html. A statistical test on the degree of agreement (K-statistic) can also be calculated.

4.0 REPORTING ASSESSMENT STATISTICS

The following statistics will be presented for each assessment, whether a QA check or an accuracy assessment:

---

[5] Foody, G. M. 1992. On the Compensation for Chance Agreement in Image Classification Accuracy Assessment. *Photogrammetric Engineering and Remote Sensing,* 58 (10): 1459-1460.

1. Chi-squared test of proportions [e.g., insignificant difference in ecosystem unit proportions].
2. Percent dominant correct [e.g., 62%], and 95% confidence interval [e.g., +/- 6%] for each map entity (i.e., site series, structural stage, etc.).
3. Percent overlap [e.g., 49%] for each map entity.
4. Percent acceptable overlap, if assessed [optional].
5. If all "simple" map units, presentation of a confusion matrix [optional].

Understanding the results of the assessment is critical to the correct use of the data. The interpretation of each is discussed below.

4.1 Chi-squared Test of Proportions

This test is an unbiased assessment of whether the area of each entity from the map matches that determined from the polygon samples. Areal extent of a site series (e.g., is a common interpretive use of an ecosystem map). Testing whether the map areas are the same as that from the sampling provides an indication of the accuracy of the mapping.

4.2 Percent Dominant Correct

Determining the percent of the polygons where the dominant map entity is correct is a useful statistic. If the value is high, it is one indication that a high proportion of the map is accurate.

4.3 Percent Overlap

Determining the percent overlap between the mapped entities and the polygon assessments is another overall indicator of the accuracy of the mapping. This value has been used in TEM QA where it is often referred to as the "Penfold" statistic.

4.4 Percent Acceptable Overlap

The percent acceptable overlap can indicate how well the mapping was correct or close to correct. If the percent overlap value is not as high as preferred, then the calculation of this value can assist in determining if the mapping is acceptable, as it is "close" to correct. This value has been used in TEM QA where it has been referred to as the "Sutherland" statistic.

4.5 Confusion Matrix

A confusion matrix is very useful, so where it can be prepared, it should be. Knowing where errors are occurring in the classification and of what kind is very useful to users of a map and can also assist in upgrading maps or mapping models.

5.0 CONCLUSION

Assessing the accuracy of mapping in complex thematic maps is not a simple matter. The approach outlined in this protocol provides a means of obtaining some overall, statistically valid 'scores' of the accuracy of TEM, PEM or other ecosystem maps. The results can be used for quality assurance or for presenting the statistics on the accuracy of mapping.

The assessment focuses on the thematic content of polygons. Polygon boundaries are not assessed. The results are also "non-spatial," in that they do not show explicitly where the errors can be found in the whole map. However, information is provided about overall accuracy of selected ecosystem map entities.

This information is important and useful because it provides potential users of the map or project with a level of confidence in the results.

Accuracy assessments are often not conducted due to the cost of sampling. Although costly, accuracy assessments are critically important to determining appropriate uses for the mapping. In other cases, the sample design is changed to a design that is considered more efficient and therefore less costly. This is often some sort of systematic or two-stage sample design. Unfortunately, a systematic sample is not unbiased and is therefore not recommended. A two-stage sample design is possible, but requires different statistical analyses that have not been discussed in this protocol.

APPENDIX 1: Sample Sizes for Overall Chi-squared Test

To calculate the sample sizes using this approach (i.e., chi-squared goodness-of-fit test where the map proportions are assumed to be correct under the null hypothesis), the map and "true" proportional areas have to be specified. The table below presents a set of choices (at the top of the table) where the map and sample areas are fairly similar and therefore a larger sample size would be required to reject the null hypothesis (that they are the same), than if the differences were greater. These values were selected for determining the sample size. The assumption that some site series are common, others uncommon to rare, fits the distribution normally found in areas that have been mapped.

Site series	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
True proportional area	0.35	0.25	0.15	0.10	0.05	0.04	0.03	0.02	0.01
Map proportional area	0.30	0.20	0.20	0.20	0.02	0.02	0.02	0.02	0.02

Sample sizes required to test null hypothesis that map proportional areas (above) are correct when true areas are the same as those given above.

	Alpha =0.05			**Alpha = 0.1**		
Minimum power	0.80	0.90	0.95	0.80	0.90	0.95
	Sample size					
Ordinary Pearson chi-square	95	121	144	78	102	124
Likelihood ratio chi-square	99	126	150	82	107	129

Example: If the sample size is 95 there is 80% chance that the null hypothesis will be rejected at the alpha = 5% level of significance.