

Problem Analysis on Reliability, Quality Control and Validation of Predictive Ecosystem Mapping (PEM)

– FINAL REPORT –

June 29, 1999

Prepared for:
**Terrestrial Ecosystem Mapping Alternatives Task Force
Resource Inventory Committee (RIC)**

Prepared by:
Richard A. Sims and Jeff Matheson

**578569BC Ltd. dba
R.A. Sims and Associates
201-1600 West 6th Avenue
Vancouver, British Columbia V6J 1R3
Tel: (604) 733-5220 / Fax: (604) 733-5204
rasims@home.com**

Executive Summary

Predictive Ecosystem Mapping (PEM) is being advocated, under the right combination of circumstances, as a cost-effective and potentially superior approach to the intermediate-level (e.g., 1:20,000 scale) mapping of ecosystem resources in BC. In establishing the working principles for PEM, it is recognized that reliability assessment is a significant issue. Mapping of ecological relationships can be a complex and multidimensional task, and PEM needs an effective set of tools and approaches for reliability assessment, so that accurate and useful PEM products can be developed.

PEMs involve the production of spatial models and GIS-based algorithms. PEM outputs are typically not derived from a process of manual air photo-interpretation of linework and attributes, and don't result in a set of document photos with annotations that can then be used as a basis for reliability checking. As a result, PEMs must instead be assessed in relation to independently-collected reference datasets.

This report reviews issues related to the development of a set of reliability assessment methodologies for PEM. Spatial (topological) accuracy versus thematic accuracy is generally considered, as well as the need to focus reliability assessment on the intended end-uses for individual PEM projects. Consideration of thematic elements is the most important aspect for reliability assessment, and this involves the characterization of PEM project outputs in relation to their thematic accuracy, precision and reliability. Following the review of background concepts and aspects influencing reliability assessment (Sections 2 and 3, in particular), a formalized framework for reliability assessment is presented (Section 4). A pilot study that follows this framework is summarized in Section 5.

Several recommendations for future steps are made (Section 6), and these should be addressed through either a strategic planning and testing effort, or via some directed pilot projects. 1) The needs of resource managers / planners must be firmly established, so that PEM outputs, in particular interpretive products, are properly developed. 2) The need for reliability assessment increases as PEMs become more removed from primary data sources. This necessitates that some boundaries be established – at some point – as to what is acceptable and not as a PEM tool or technology. 3) The expectations for spatial accuracy of PEM need to be more firmly established, as there are some interpretations that may not be able to achieve the accuracy (and precision) that is needed for detailed decision-making. 4) The PEM reliability assessment system that evolves from this exercise should attempt to partition the errors in mapping, so that systematic problems can be identified, and so mitigation efforts can be focused. This may be a difficult thing to put into practice, but a set of tools that effectively partition errors would provide the best solution. 5) PEM reliability assessment metadata must be incorporated into project metadata standards. A mechanism for this needs to be defined and described, ideally, at the outset. 6) A mechanism is needed for PEM change management, given the potential for rapid technological change with PEM-related tools. Only when the products and interpretations from PEM and the needs of resource planners are clearly understood, can alternatives to robust tests of reliability be considered. 7) Future considerations about PEM map reliability should also be considered, as part of a larger initiative involving ecosystem map reliability as a whole in BC, so that the right ecosystem mapping systems are applied under the appropriate circumstances. One component of this is monitoring the evolving needs for new types of interpretive products or intended applications for PEMs.

Implementation of PEM as an operational tool presents a number of challenges. The development of an effective approach to reliability assessment is fundamental in the implementation process.

Acknowledgments

We are pleased to acknowledge Dave Clark and members of the TEM Alternatives Task Force (Del Meidinger, Bobby Love, Bill Beese, Rob Stewart, Frits Nijholt and Fern Schultz) for their ongoing support for and interest in this project. We also thank Keith Jones and David Moon for sharing their ideas and concepts regarding PEM reliability assessment and related issues. Keith Jones, Del Meidinger and Dave Clark provided very useful review comments on an initial draft of this report.

Table of Contents

Executive Summary	ii
Acknowledgements	iii
Table of Contents	iv
List of Figures	v
List of Tables.....	v
1.0 INTRODUCTION.....	1
1.1 GENERAL CONTEXT	1
1.2 PURPOSE.....	1
1.3 CHARACTERISTICS OF PEM	2
1.4 THE NEED FOR PEM RELIABILITY ASSESSMENT	2
1.5 RELATION TO PEM QUALITY ASSURANCE/QUALITY CONTROL.....	3
1.6 RELATION TO TEM RELIABILITY ASSESSMENT	5
2.0 ISSUES AND ASPECTS OF MAP ACCURACY.....	6
2.1 GENERAL.....	6
2.2 THEMATIC ACCURACY VERSUS RELIABILITY	6
2.3 SOURCES OF THEMATIC ERRORS	7
2.4 THEMATIC PRECISION.....	8
2.5 SPATIAL ACCURACY – POSITIONAL AND TOPOLOGICAL.....	8
2.6 APPROACHES TO ACCURACY ASSESSMENT	9
2.6.1 Error Matrices.....	9
2.6.2 Map Error Evaluation.....	10
2.6.3 Fuzzy Set Theory	10
2.7 THE COLLECTION OF REFERENCE DATA.....	10
2.7.1 Sources of Reference Data	10
2.7.2 Sampling Strategies.....	12
2.7.3 Operational Constraints.....	12
2.7.4 Requirements for Statistical Rigor	13
3.0 ISSUES AND ASPECTS OF PEM RELIABILITY ASSESSMENT	15
3.1 APPLYING TRADITIONAL ACCURACY ASSESSMENTS TO PEM.....	15
3.2 TERMS OF THEMATIC ACCURACY AND RELIABILITY FOR PEM	15
3.3 TERMS OF POSITIONAL ACCURACY FOR PEM	16
3.4 TERMS OF TOPOLOGICAL ACCURACY FOR PEM	16
3.5 COLLECTION OF REFERENCE DATA	17
3.6 INTEGRATION WITH MODEL CALIBRATION/VALIDATION PROCEDURES	17
3.7 INTEGRATION WITH QA/QC PROCEDURES	18
3.8 TIMING OF PEM RELIABILITY ASSESSMENT	18
3.9 COST OF PEM RELIABILITY ASSESSMENT.....	18
3.10ROLE OF RELIABILITY ASSESSMENT IN THE DATA LIFE CYCLE.....	20
4.0 PEM RELIABILITY ASSESSMENT FRAMEWORK	22
4.1 REQUIRED COMPONENTS	22
4.2 CHARACTERISTICS OF PEM REFERENCE DATA.....	22
4.3 ACCURACY ASSESSMENT.....	24
4.3.1 Spatial Accuracy	24

4.3.2 *Thematic Accuracy*.....25
4.3.3 *Calculating and Reporting Thematic Accuracy*26
4.4 RELIABILITY ASSESSMENT26
4.4.1 *Defining Reliability Requirements*26
4.4.2 *Calculating and Reporting Reliability*27
5.0 AN EXAMPLE CASE STUDY– THE BABINE EFMPP TEM/PEM COMPARISON.....29
6.0 CONCLUSIONS AND RECOMMENDATIONS32
6.1 GENERAL CONCLUSIONS.....32
6.2 RECOMMENDED NEXT STEPS FOR RELIABILITY ASSESSMENT OF PEMs32
7.0 GLOSSARY34
8.0 REFERENCES35

List of Figures

Figure 1. Information on reliability can satisfy the needs of producers, primary users, secondary users, technical oversight agencies, and funding agencies.....3
Figure 2. The relationship between quality assurance/control and reliability assessment for a PEM project. Arrows from left-to-right represent errors or uncertainties that arise out of various components and/or processes. Reliability assessment has a limited ability to identify the origin of errors, and this emphasizes the need to conduct QA/QC during developmental phases of a PEM project.....4
Figure 3. Ecosystem mapping that is more removed from primary data sources may require more effort toward reliability assessment. This creates a paradox where the cost of conducting reliability assessment may be higher for PEMs that generate complex outputs. 19
Figure 4. The life-cycle of ecosystem map databases (adapted from Goodchild, 1996).21
Figure 5. A framework for PEM reliability assessment. This is a generic framework that could be used for any ecosystem map.....23
Figure 6. Ecosystem map comparative analysis process for the Babine EFMPP. Several analytical methods are being utilized to compare and assess reliability of four types of ecosystem maps of the same area.31

List of Tables

Table 1. Relationship between sample size, confidence level and acceptable error (adapted from NBS/NPS, 1994). 13
Table 2. A simplified linguistic scale that describes the level of difficulty to delineate a polygon.25
Table 3. A possible site series accuracy scale for a generic PEM map using both point (ground-based) and polygon (photo-based) reference data.25
Table 4. Preliminary accuracy requirements for the determination of minimum acceptable reliability levels for a range of PEM interpretive products or intended applications.27
Table 5. Example reliability ratings for an intended PEM application.28

1.0 INTRODUCTION

1.1 General Context

Predictive Ecosystem Mapping (PEM) is a generic term that refers to GIS-based tools and procedures that can be used to automate – in part or whole – the spatial classification and mapping of ecosystem units and/or derivative units, such that they can then be directly used for a range of resource management and planning purposes (Clark *et al.* 1998, TEM Alternatives Task Force 1998; also, Section 7 of this report contains a glossary of selected terms).

For the most part, PEM outputs emulate the types of spatial outputs from Terrestrial Ecosystem Mapping (TEM) projects. TEM maps and related interpretations are created using a standard set of procedures that are documented and approved by BC's Resources Inventory Committee (RIC 1998). The current effort to evaluate PEM was conducted for the TEM Alternatives Task Force, and funded/administrated by the Ministry of Environment, Lands and Parks' Resource Inventory Branch.

The current investigation deals with the role and application of PEM. Some basic procedures and conventions need to be established at the outset for PEM so that there is not "transitional chaos" as a shift from standard TEM procedures occurs. Key interests with respect to PEM include:

- Their continued but more structured evaluation;
- Their future role in the TEM inventory program; and,
- The formation of "TEM-PEM Principles & Transitional RIC Standards" around the use of these methods.

The current project is one of three detailed studies that are being undertaken concurrently, in an attempt to provide a better framework for moving ahead with PEM:

Detailed Study B1 – Situation Analysis for Knowledge-Based Systems;

Detailed Study B2 – Problem Analysis on Data Quality Assessment Issues for PEM; and,

Detailed Study B3 – Problem Analysis on Reliability, Quality Control and Validation of PEM (this report).

1.2 Purpose

The purpose of this report is three-fold:

1. To identify (update), describe and review alternative methods for assessing land resource inventory map reliability in relation to quality assurance/quality control procedures and with respect to field and knowledge base calibration/validation methods used in both PEMs and conventional TEM.
2. To undertake a comparative analysis of findings: advantages, disadvantages, trade-offs and practical considerations (e.g., access, cost, rate of production), objectivity, sampling-stratification considerations, comparability of results, risk associated with incorrect interpretations or decisions being made, etc.; and, to identify any unresolved issues relative to PEM.

3. To establish a set of alternative map reliability procedures, including a standard terminology, that addresses different classes of interpretive requirements (decision risk), recognizes practical field logistics, and incorporates these activities such that they are integral to the entire mapping, sampling, database development and reporting (tabular and map) process.

1.3 Characteristics of PEM

In British Columbia, TEM has been, for several year, providing detailed resource inventories for a variety of resource management purposes and applications (RIC 1998). Alternative approaches to TEM that can provide similar information at the same (or, under suitable circumstances, better) quality and a lower price are being explored. Some of these alternatives may deliver the same product, or a functional part of the content, cheaper and more quickly than TEM. These attributes may be useful in addressing specific ecologically-based management issues in the absence of full TEM products (Biggs *et al.* 1997, Clark *et al.* 1998).

PEM approaches are methods of stratifying the landscape into ecologically-oriented map units on the basis of existing mapped features, point information and knowledge about ecosystem-landscape relationships and their distribution pattern (Clark *et al.* 1998). Jones *et al.* (1999) further define PEM as “a computer, GIS and knowledge based method to stratifying landscapes into ecologically-oriented map units based on the overlaying of existing mapped themes and the processing of the resultant attributes by normally automated inferencing software with a formalized knowledge base comprising ecological–landscape relationships”.

Each PEM approach uses a unique suite of input data to delineate and classify ecosystems and there are differences in the products that are produced. Some approaches may be suitable for only certain applications. In order for resource managers to select among ecosystem mapping methodologies, the characteristics and reliability of ecosystem maps produced using these various methods needs to be known.

1.4 The Need for PEM Reliability Assessment

There is widespread recognition of the importance of reliability assessment of thematic maps. However, few agencies have infrastructures designed to respond to this need. In BC, the emergence of new approaches to resource inventories has created an increased need for objective evaluations and comparisons. Resource managers are now faced with a variety of inventory tools, all of which have slightly different products that have varying degrees of appropriateness for various applications. *A priori* information on the appropriateness of TEM and PEMs are required, and reliability must be robustly documented, in order to select PEM approaches for particular applications. Figure 1 illustrates how information on map reliability is important to many groups.

In order to meet this objective, there are three basic requirements:

1. A clear understanding of the interpretative needs, priorities and methods of resources managers (TEM Alternatives Task Force 1998);
2. An understanding of the PEM products; and,
3. An evaluation of their appropriateness for various interpretations/applications.

Furthermore, the need for reliability must be guided by the consequences or risk of an incorrect interpretation or decision being made (TEM Alternatives Task Force 1998).

To assist resource managers in the selection of an inventory method, decision tools are in development. A draft decision tool has been established by the TEM Alternative Task Force (1998), and a draft inventory program decision tool has been tested in the Prince Rupert Forest Region (Osborn 1997). It will be important for the results of reliability assessments to directly and continuously feed into the development of inventory decision tools, especially for new methodologies.

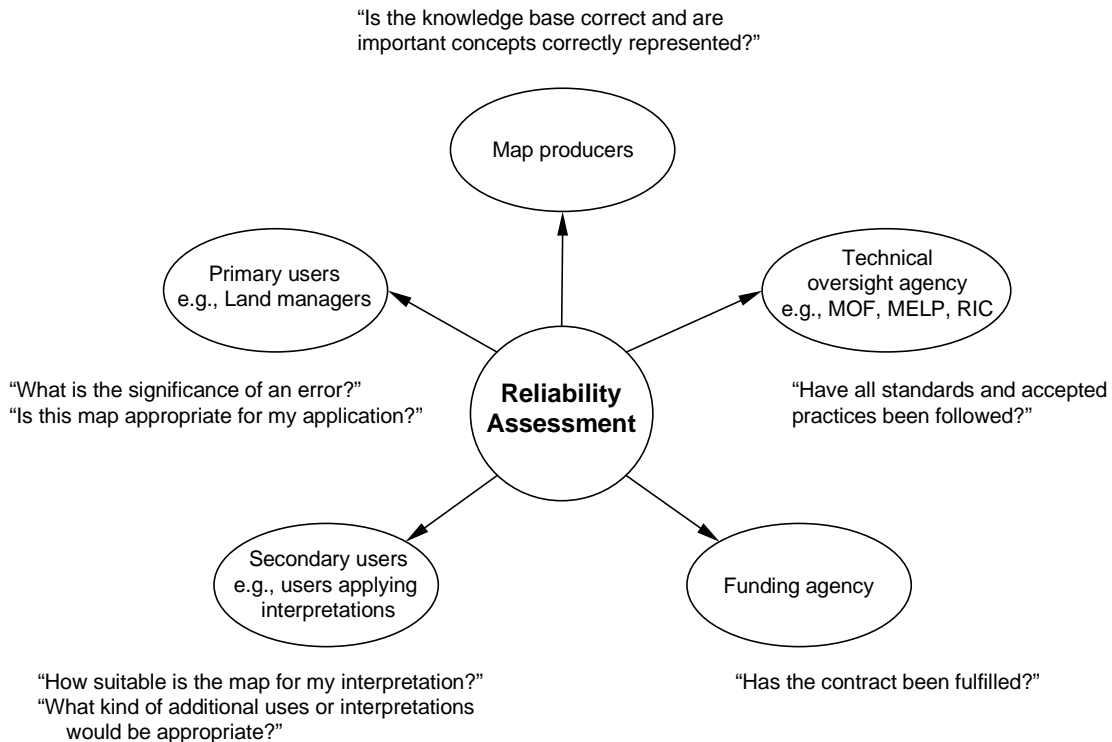


Figure 1. Information on reliability can satisfy the needs of producers, primary users, secondary users, technical oversight agencies, and funding agencies.

1.5 Relation to PEM Quality Assurance/Quality Control

Reliability assessment is one component of a suite of three procedures that are necessary to ensure a reliable PEM product (Figure 2). Input database and knowledge base quality assurance/quality control (QA/QC) should normally be conducted as precursors to an assessment of the product, in part because their assessment should be part of the decision-making process as to whether to proceed with a PEM or not in the first instance. An early QA/QC of input databases and the knowledge base also ensures that the tools used to predict ecosystems are sound.

These two phases are important for screening errors before a reliability assessment is done. If input database screening and knowledge base QA/QC have not been conducted, then the entire burden of a

PEM project's QA/QC lies on the reliability assessment phase. This is probably beyond the capabilities of a test of map reliability and, in any case, would probably occur too late to have a meaningful influence on the development of the PEM.

However, with certain limitations, a reliability assessment is able to identify errors originating with the input databases or with the development of the knowledge base, provided the assessment is concurrent with the mapping. There are other issues associated with this relationship and these are discussed more in Section 3.7.

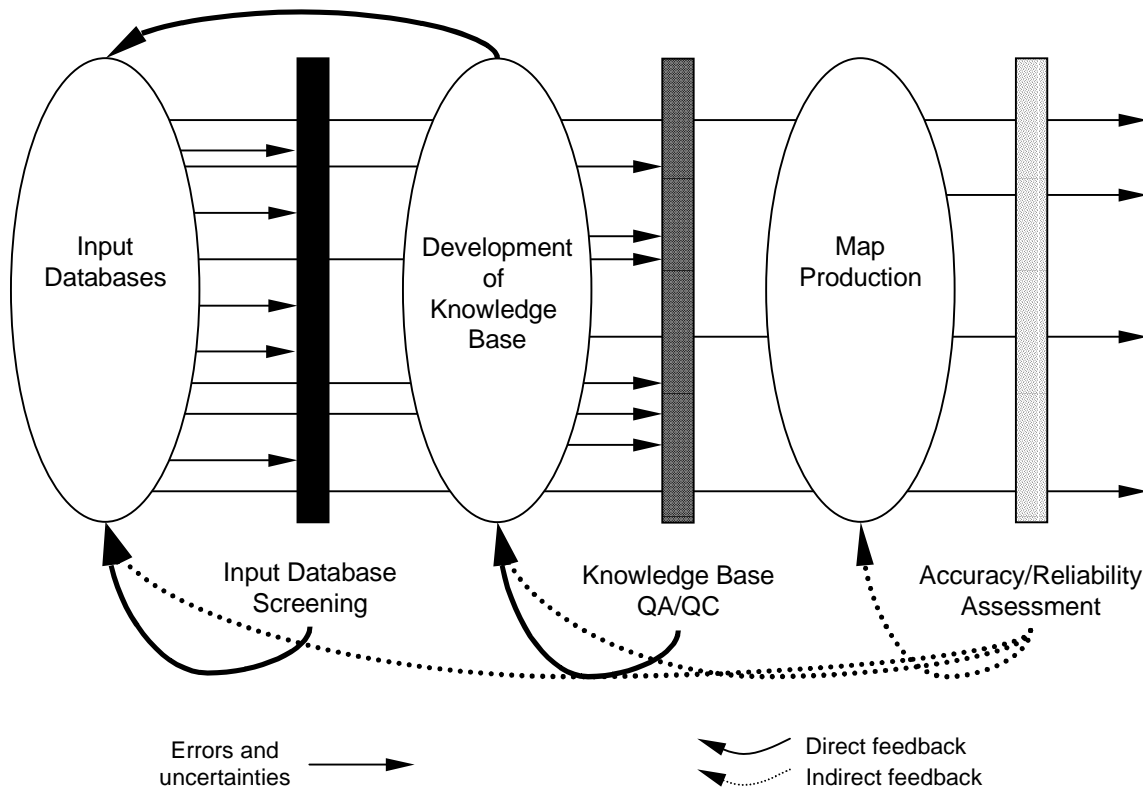


Figure 2. The relationship between quality assurance/control and reliability assessment for a PEM project. Arrows from left-to-right represent errors or uncertainties that arise out of various components and/or processes. Reliability assessment has a limited ability to identify the origin of errors, and this emphasizes the need to conduct QA/QC during developmental phases of a PEM project.

1.6 Relation to TEM Reliability Assessment

Standards for reliability assessment of TEM are currently in development. Since PEM is being advocated as an alternative to TEM (when certain criteria are met), it is recognized that it would be desirable to have complementary reliability assessment procedures for PEM and TEM. If this can be achieved in practice, it will allow some standardization of approaches and hopefully the ability to compare reliability issues among different mapping outputs.

There are RIC-approved standards for TEM (RIC 1998), guidelines for TEM QA/QC (Anon. 1998), a formal mechanism for the review/approval of TEM field projects and map products, and provincial-level coordination of the overall TEM program. Individual map projects are screened into, and then archived within a provincial TEM data clearinghouse.

The distinction between PEM and TEM is somewhat artificial, given that all ecosystem mapping approaches – by their very nature – are predictive. For the purposes of this report, however, the key distinction is that TEM projects follow established RIC procedures that include the conduct of a field data collection phase and air photo interpretation to delineate polygons and annotate their contents (RIC 1998, Sims 1997). PEMs, on the other hand, involve a much greater emphasis on GIS automation, mainly via spatial data overlays and syntheses.

TEM outputs provide annotations and labels that can be directly checked against photo-interpreted materials. The “hard copy” records that result from the TEM process include the photo-based evidence that illustrates the means by which the photo-interpretation “mental model” was applied. Such records do not necessarily exist for PEMs, as the “GIS automation” is directed towards saving time and effort, and increasing precision, typically by improving upon the laborious process of manual interpretation from photos. Consequently, PEM lines and polygon contents may be “derived entities” which need to be subjected to a larger amount and/or a different form of independent checking. The implications of this are discussed in Section 2.7.

2.0 ISSUES AND ASPECTS OF MAP ACCURACY

2.1 General

In this document, the term “reliability assessment” (RA) is used broadly to refer to all aspects of the assessment of map accuracy and reliability. However, some key distinctions should be made at this point regarding the component elements of data quality. In his review of “data input quality issues” for PEM, Moon (1999) defined basic map elements as being either **thematic** (i.e., the information contained in the the map and any associated documents) or **spatial** (i.e., the placement of points lines or polygons within x, y, and (sometimes) z space (**positional**), and/or their relations to each other (**topological**)). The same distinctions are employed in this report.

For reliability assessment, the focus is predominantly – but not exclusively – upon thematic issues (see Sections 2.2 to 2.4). The assumption is made that, in most instances, input data sources will have undergone some form of prescreening to ensure an acceptable level of positional and topological accuracy for use within a PEM. While this will certainly not eliminate spatial sources of errors, it should serve to control and minimize them. Errors associated with spatial accuracy are considered further, in relation to PEM, in Section 2.5).

Criteria and techniques for assessing accuracy in thematic mapping have been developed, refined and tested over the past twenty years, especially in the field of remote sensing (e.g., see Goodchild and Gopal 1989, Story and Congalton 1986, Congalton 1991). This section provides a review of some of the issues related to assessing thematic map accuracy, and highlights some historical approaches.

2.2 Thematic Accuracy Versus Reliability

Accuracy is defined as a measure of the absence of error or a measure of how close an estimate is to its true value. Accuracy is typically interpreted in relation to the degree of **precision** (exactness or resolution) with which it is measured. It is important to explicitly state how accuracy is being assessed (i.e., by what measures of precision), given that the application of an imprecise scale can lead to a measure of high accuracy, while a precise scale can yield the opposite. Predictions of higher precision (e.g., can spruce be distinguished from fir?) will be more difficult to make with high accuracy than will be predictions of low precision (e.g., can forest be separated from non-forest?).

Thematic accuracy is concerned with the exactness that is associated with map polygon labels, and within PEM, this means the accuracy with which those attributes that are predicted and generated by the GIS-automation process for each polygon. **Thematic precision**, for PEM, relates to the exactness of definition associated with the sets of classes that are used for each part of the polygon label. Defining thematic accuracy and precision within the context of PEM is not trivial. Additional conditions must be embodied in the definitions for both accuracy and precision; for example, how is precision defined in relation to units that cannot be recognized with equal confidence (e.g., can spruce be as accurately determined as fir?), or how is precision apportioned within complexes (e.g., can a 60:40 spruce:fir mix be accurately differentiated from a 40:60 spruce:fir mix?)?.

Reliability is a measurement of accuracy and precision in relation to a defined purpose. As has already been introduced, and as outlined in further detail in Section 3, a PEM product must be assessed in the context of its application and designed function (e.g., for the development of a defined set of wildlife habitat interpretations). This is quite separate from a measure of a map's raw accuracy. For the purposes

of this document, reliability for PEM is defined as a measure of the degree or probability of the map to satisfy its specified functions. As earlier noted, of primary interest is a PEM's label information content in relation to the stated purpose, or the **thematic reliability**. Because a PEM output may be used for multiple applications or interpretations, a map may have multiple measures of reliability which are specific to each interpretation (function).

A PEM's thematic accuracy and reliability may vary with the chosen level of spatial resolution for a specific application. For example, a map may have low polygon accuracy (i.e., the probability of a given polygon correctly predicting what is actually present is low) but may have a different accuracy when a smaller (coarser) scale is selected. A coarser interpretation of polygon boundaries and contents (i.e., related to thematic precision) may then provide a more acceptable (i.e., more accurate and possibly more reliable) product for a coarser-scaled interpretation like regional planning. One means of representing this is by determining the amount of correctly classified units over a sufficiently large area (e.g., most frequently a 1:20,000 scale mapsheet, but also a Landscape Unit, Timber Supply Area, Forest District, Forest Region, or other defined geographic area).

For the purposes of this document, a distinction is also made between **polygon accuracy/reliability** and **global accuracy/reliability**. It should be noted that the latter is not simply the "sum" of the polygon reliabilities within a mapsheet, since it does exist at a broader, or more generic, level of resolution than individual polygons. Global accuracy/reliability can be represented, for example, as a "map-as-a-whole" accuracy/reliability as opposed to overall map accuracy/reliability. In practice, the two terms can be used to differentiate polygon-level and overall map-level (e.g., 1:20,000 scale) levels of resolution, and both are of interest when a PEM's reliability is being assessed.

2.3 Sources of Thematic Errors

In thematic maps, users are primarily concerned with the probability that a mapped class (e.g., a polygon label and explicit or inferred polygon attributes in a database) "correctly matches" the condition found at that point in the field. If the mapped class does not match the condition on the ground, either a thematic error (a misclassification) has occurred, or a positional error has occurred. Thematic errors can arise for a number of reasons (of which only some would apply to a PEM map), and include:

- Poor air-photo quality;
- A highly heterogeneous landscape in relation to the scale of the map product;
- Classes which are difficult to differentiate during delineation and/or in relation to on-the-ground allocation to a class;
- The incorrect application of classification rules (e.g., due to mapping inexperience) during delineation or with on-the-ground allocation to a class;
- Poor-quality input datasets, including the basemap;
- An insufficient knowledge base;
- Low positional accuracy and errors due to distortions introduced by data processing, such as digitizing errors, transcription errors or map projection/registration errors (see spatial accuracy, Section 2.5);
- Temporal changes in vegetation; and,
- Problems or errors associated with an independently-collected dataset gathered to assess the mapping results.

The significance of thematic errors are often quantified by their frequency and magnitude. Frequency refers to how often an error occurs, and magnitude refers to the “level” at which an error occurs. For example, a confusion at the biogeoclimatic level would be, in general, more significant than a confusion at the site series level, because the latter designation depends upon the former. Knowing how frequently and to what magnitude a particular problem occurs allows assessors to weigh the significance of an error with respect to the overall PEM project. Identifying the source of errors provides assessors with information about which aspects of the thematic mapping process are creating the problem.

For the assessment of reliability, it is also important that errors be assessed in terms of their magnitude and importance to those specifically-defined applications that the PEM will be used for. For example, an engineering-related interpretation might require that a highly-accurate assignment be made of soil drainage class to polygons, and in such a case, the accuracy of the biogeoclimatic level assignment would be of little or no consequence.

2.4 Thematic Precision

Thematic precision refers to the exactness with which mapping entities are defined. Moon (1999) describes that thematic precision may refer to:

1. The exactness with which attributes defining mapping entities are defined;
2. The exactness of the attributes associated with a mapping entity when it is found in a specific polygon as opposed to its general definition;
3. The definitional exactness of the composition and distribution of mapping individuals in a compound map entity; and,
4. The exactness of composition and distribution of mapping individuals in a specific polygon labeled as a compound map entity.

Thematic precision should not be confused with positional precision, which refers to “the exactness with which a location can be determined” (Moon 1999).

2.5 Spatial Accuracy – Positional and Topological

As already noted, an error can occur in a thematic map when the class mapped at some location is not the same as the class observed at that location in the field. It is often difficult to confirm whether the error is due to a thematic or spatial error (i.e., a correctly classified point that is in the wrong spot on the map) error. This is because a disagreement over the class could be interpreted as being due to either an error in the point’s position, or an error in the class at that point (NBS/NPS, 1994).

Where a ground position is sufficiently well-defined, such as where the boundary follows a road, or where high-quality and corrected GPS data is available, it may be possible to make an independent distinction between positional accuracy and thematic accuracy. The increasing use of GPS technologies is making it possible to regularly determine the exact positions of points, polygon boundaries and reference features like cutblock edges, roads and waterbodies. As positional accuracy is more regularly and routinely integrated into resource surveys, including PEM, it is likely that accuracy issues will focus more upon the potentially more problematic issue of thematic accuracy.

Positional accuracy is independent of the accuracy of any particular resource map accuracy, and refers to the degree to which map coordinates correspond to the real world coordinates of features shown on the map.

Topological accuracy, for PEM, refers to the relative spatial placement of a mapped point, line or polygon in relation to other mapped units. It is possible and not uncommon for a map to have high topological accuracy and, for a variety of reasons (e.g., related to technical errors), have low positional accuracy. In such instances, the spatial relationships among mapped units may be generally considered acceptable, but are not precisely positioned. For some applications, this may be acceptable – for example, having reasonable topological accuracy is sufficient in the case where the map unit will ultimately be traversed on-the-ground. However, in those cases where the mapped information represents the entire basis for a decision (i.e., including most modern mapping activities!), then positional accuracy is important.

PEM and TEM products that suffer from poor positional accuracy but have acceptable topological accuracy could still be useful for general planning, but not site-specific planning. However, it should be anticipated that PEM (and TEM) maps will be, in future, frequently overlain and combined with other spatial databases, as a basis for “spatially-explicit” planning and scenario building using GIS. Given this, it is important that knowledge about both the positional and topological spatial accuracy be acquired and maintained as part of the reliability assessment procedure.

2.6 Approaches to Accuracy Assessment

2.6.1 Error Matrices

A common technique used to evaluate the accuracy of thematic maps is the “error (or confusion) matrix”. An error matrix is a square array of numbers set out in rows and columns which express the number of sample sites (i.e., pixels, clusters of pixels, or polygons) assigned to a particular class relative to the actual class as verified on the ground (Congalton 1991). Overall accuracy of a classified map can be assessed, using the error matrix.

The accuracy of individual categories can also be evaluated, either as errors of omission or commission (Story and Congalton 1986). Errors of omission are measured by dividing the total number of incorrectly classified samples for a class (row), by the total number of reference samples for that class (i.e., the column total). The resulting percentage describes the probability that a reference (ground) sample will be incorrectly classified, and where “the act of getting it right” was omitted. Errors of commission are measured by dividing the number of incorrectly classified samples of a class (row total) by the total number of samples classified as that class (column total). The resulting percentage indicates the probability that a reference (ground) sample will be incorrect, and “the act of getting it wrong” was committed.

The error matrix technique is based on the assumption that natural landscapes can be divided into discrete units, and it is best applied to simple thematic maps. It is limited in its functionality for complex mapping projects, because there is a practical limitation to the size of the constructed matrix. For PEMs, it would have particular difficulties dealing with the additional permutations that exist, because compound or complex labels need to be examined.

2.6.2 Map Error Evaluation

A modified approach to the error matrix has been proposed by Moon *et al.* (1987), and tested via various field applications of soil survey mapping. The approach involves initial stratification of map delineations (map units or attributes/variables), and the application of categorical variable statistics to the resulting sets of classes. A field data-collection methodology involves the preparation of an independent set of field observations (i.e., points [sites] installed along random transects installed within the mapped area). Observational data are then summarized to determine site deviations, error values and confidence intervals for the estimated proportions. Statistics are then prepared which generally partition the types and extents of map error, and then identify whether those errors are associated with a small number of badly mapped delineations or a large number of slightly mis-mapped delineations (Moon *et al.* 1987).

2.6.3 Fuzzy Set Theory

Another method for assessing the accuracy of thematic maps is based on fuzzy set theory which addresses the fact that ecological boundaries are generally not discrete, but rather, are continuous in nature (Gopal and Woodcock 1994). The fuzzy analysis technique uses a linguistic membership scale to evaluate how well an observed class on the ground corresponds with each different category found on the map; in other words, it measures the “relative” accuracy of a sample area. The five linguistic values and descriptions in Gopal and Woodcock’s (1994) scale are:

- (1) *Absolutely Wrong*: This answer is absolutely unacceptable. Very Wrong.
- (2) *Understandable but Wrong*: Not a good answer. There is something about the site that makes the answer understandable but there is clearly a better answer. This answer would pose a problem for users of the map. Not right.
- (3) *Reasonable or Acceptable Answer*: May not be the best possible answer, but it is acceptable; this answer does not pose a problem to the user if it is seen on the map. Right.
- (4) *Good Answer*: Would be happy to find this answer given on the map. Very Right.
- (5) *Absolutely Right*: No doubt about the match. Perfect.

For example, a point on a map classified as coniferous forest might be assigned the following values for the following types of map classes: (3) acceptable answer - mixed forest; (1) absolutely wrong - water; and (5) absolutely right - coniferous forest.

Fuzzy functions are used in association with the linguistic values assigned to each sample area to evaluate the nature, frequency, magnitude and source of errors in a thematic mapping project.

2.7 The Collection of Reference Data

2.7.1 Sources of Reference Data

Reference data provides the basis for evaluation of accuracy and reliability of a PEM. Suitable reference data for this purpose may come from a variety of sources:

- Ground-truthing data (i.e., collected solely for the purpose of accuracy assessment);
- Expert photo-interpretation;
- Reliable existing ground plot data; and,

- Maps or other sources of spatial data that are known to be of high accuracy.

For obvious reasons, ground-truthing data represents the most accurate and preferred source of reference data. As described in subsequent sections, the field collection procedure can be aligned with the requirements for reliability assessment, and the data can be collected and interpreted specifically for the purpose at hand.

Under some circumstances, photo-interpreted reference data may be a relatively inexpensive and acceptable alternative. It needs to be conducted by an expert, and the uncertainty and variability of this type of data must be considered and weighed when it is being used as a basis for evaluation. Photo-interpreted data may be useful as a reference database for small-scale mapping when high-quality, large-scale photos are available.

The use of existing ground plot data may also, under some circumstances, be an acceptable source of data for reliability assessment. The main advantage of using existing plot data is that there is a time and effort savings in relation to the alternative of installing and visiting new field plots. Existing ground plot data is only acceptable if the same classification scheme is used, and is unambiguously applied. If the existing data was used in the development or validation of the knowledge base, then it does not represent an independent dataset for a final analysis of reliability, and cannot be used. The positional accuracy of ground plots, the age of the plot data, the suite of measured variables and their methods of measurement, the sampling design and density, and other factors can all limit the usefulness of existing plot data for reliability assessment. As an example, vegetation may have undergone significant structural and successional changes if considerable time has elapsed since the plot was initially described and the PEM was generated.

Other existing mapped data may provide useful sources of reference data, but in many cases it would be dated (e.g., a lack of existing mapped data is frequently the rationale for undertaking new thematic mapping in the first place). Significantly different data collection or aggregation/presentation methodologies may have been used. However, base map data (e.g., from TRIM or NTS) may be one type of existing mapped data that might provide some useful reference data for the assessment of positional accuracy.

Regardless of the type of reference data used, the data should represent – even if it is point data – a geographic extent on the ground. Ideally, the data should be positionally and topologically accurate, and describe an area that is relatively homogeneous. It is also desirable (but not essential) that the reference data corresponds to an area that is equal in size or larger than the “minimum mappable area” for the scale of map being generated. The minimum mappable unit varies among approaches, and depends upon the intended resolution of the map being prepared. For TEM, the minimum polygon size at 1:20,000 scale mapping is 2.0 ha. According to RIC (1998), the smallest possible decile in a complex is 10% (although a lower limit of 20% is observed in many TEM projects), making the smallest possible mapped unit 0.2 ha or approximately 45m x 45m. To achieve the requirement for a homogenous class, the sample point should lie away from transition zones.

Reference data for PEM reliability assessment should exist as a “source of higher accuracy”. The data content and representation should be highly complementary to the products being assessed. For example, if a higher accuracy map is to be used for the assessment of thematic accuracy of PEM products, then the source of higher accuracy should have been classified using essentially the same classification method. Any errors inherent within a source of higher accuracy should be negligible in comparison to the errors that are to be measured

2.7.2 Sampling Strategies

Choosing an appropriate sampling method is important for creating a sample population which is representative of the entire map. Although random sampling is the most unbiased way to select samples, this type of technique is unlikely to create a representative sample population.

Complex landscapes with many small ecosystem units are naturally more difficult to classify than simple landscapes, and therefore more likely to be inaccurate. Random sampling will tend to under-represent small but possibly significant categories unless the sample size is quite large (Congalton 1991). Accessing randomly selected polygons in remote areas can also be a problem. Random sampling could also result in some rare classes being missed altogether.

Consequently, some type of stratification is often required. If an equal number of samples is acquired within each class, the expense of accuracy assessment will be the same whether a class is rare or abundant, and the assessment will be independent of the class's importance to any eventual application. On the other hand, if the number of samples allocated to each class is proportional to the class area, then the reliability will be assessed with greater precision for those classes of greatest significance to interpretations of the whole map. If uncommon ecosystems are most important to the interpretation, then they should be sampled in equal (or nearly so) sample sizes to more common classes.

In practice, a compromise is often used in which common classes receive more samples than uncommon classes, but not proportionately more (NBS/NPS, 1994). To achieve this, sampling is typically undertaken using a check-sheet or simple matrix that is carried into the field, so that an ongoing tally can be maintained while sampling. As more common or accessible units become adequately sampled, attention is redirected towards finding, describing and including less common or more inaccessible units.

One other alternative is that a map can be stratified based on one or a few specific components (e.g., a site series and/or soil moisture regime). A subset of points or polygons is then randomly selected from this stratified sample population.

2.7.3 Operational Constraints

As mentioned above, randomly selected sample points can result in accessibility problems. A recommended approach for dealing with accessibility problems for ground-based sampling is to drop and replace inaccessible sample sites according to a fixed set of decision rules. Sample replacement procedures should be documented as part of the accuracy assessment report.

It is recommended that potentially inaccessible sample sites be anticipated (and replaced) as much as possible prior to the field visit, because this will ensure a greater degree of objectivity in the sample than site replacement in the field. If sample sites are found to be impractical to access during the field visit itself, it is recommended that the site be dropped and that an alternative site be located in the field, if possible. If no such alternative site is possible, the sample site should be dropped (NBS/NPS, 1994).

Since sample site replacement will, to a considerable degree, be dependent on the judgment of the investigator, it is important to provide some guidelines as to what constitutes an accessible sample site. Cost and time constraints are expected to play an important part in determining site accessibility, but care should be taken to minimize the introduction of sampling biases (i.e., towards easier-to-access locations, at the expense of sampling variability within the map area).

If site accessibility for ground-based sampling is especially limiting, consideration must be given to other sources of higher accuracy, such as air-photos, videography or observations from a hovering aircraft (i.e., “air calls”). Such remote or secondary types of sampling need to be documented as such and will, of course, lower the confidence in the accuracy of the site evaluation, but this may be an acceptable trade-off to obtain more and better distributed sampling sites. Some testing of these “remote procedures”, undertaken at the prototype stage, is a useful step, and it would help in the assessment of their impact on the confidence in the accuracy assessment.

Seasonal effects, seasonal variability, and temporal changes can also cause problems with the collection of reference data. Where this is a concern (and it is not normally of concern for PEM mapping), field visits need to be conducted when the conditions are as similar as possible to the conditions when the original map was compiled. From an operational point of view, this might result in situations where there will be at least a one-year delay between original data collection and the accuracy assessment.

Samples that are particularly subject to temporal change, like a forest fire, would ideally be dropped from the accuracy assessment itself. The change should be documented and the area targeted for re-mapping, since this type of change will probably affect the position of the boundaries of adjacent vegetation communities.

2.7.4 Requirements for Statistical Rigor

The appropriate number of sample sites to use for an accuracy assessment is dictated by the level of error that is permissible in the data and the required statistical confidence level. Table 1 shows sample sizes for various combinations of acceptable error. The derivation of these numbers is reviewed in NBS/NPS (1994).

In order to obtain per-class estimates of accuracy, the number of samples in each class must meet these sample size requirements. For maps that require high levels of accuracy and for which statistically rigorous accuracy assessment is performed, a huge number of samples are required.

Even for modest levels of error (0.02) and confidence (90%), seven samples are required for each class. For a map that has 20 sites series, 140 samples would be required. However, the level of confidence in overall accuracy (for all classes cumulatively) would be extremely high. Clearly, a balance between what is statistically sound and what is practically attainable must be found.

Table 1. Relationship between sample size, confidence level and acceptable error (adapted from NBS/NPS, 1994).

Level of Acceptable Error	Confidence Level		
	90%	95%	99%
.01	2,704	4,330	8,656
.05	108	173	346
.10	27	43	87
.20	7	11	22
.25	4	7	14

Congalton (1991) suggests that 30-50 replicates is an acceptable minimum sample size for generating valid statistical results in resource classification mapping; he also, however, acknowledges that “a balance between what is statistically sound and what is practically attainable must be found.”

As earlier noted, there is a relatively high cost involved in reliability assessment field sampling, and this cost needs to be controlled and minimized as much as possible. Practicality is important here, but so is the value of these data against risk of error. The effort should be linked to the interpretive risk, and considered against at what point you just have to get out of the office and truck and take a look on the ground.

3.0 ISSUES AND ASPECTS OF PEM RELIABILITY ASSESSMENT

3.1 Applying Traditional Accuracy Assessments to PEM

Traditional approaches of map accuracy assessment can be problematic when applied to PEMs. Medium scale mapping (1:20,000 and 1:50,000 scale) in British Columbia produces polygon labels that may contain more than one ecosystem (as is the case in TEM) or predict probabilities of occurrence of multiple ecosystems (e.g., PEM “belief values” (cf. Mulder and Corns 1996)).

To illustrate the problem, consider an ecosystem map with polygons containing up to three ecosystems. If there are ten types of ecosystems (e.g., site series), there are 6,939 possible combinations. Even assuming that only a fraction of these might actually be mapped (many combinations might be ecologically impossible), an error matrix would be unwieldy and the likelihood of being able to populate the matrix with ground reference data is practically impossible.

3.2 Terms of Thematic Accuracy and Reliability for PEM

The thematic accuracy of a PEM map label may need to be interpreted in terms of more than a single measure of precision. Some ecosystems are easily confused and thus the expectation of differentiating them may be low. As well, map labels will often contain multiple ecosystems or predict the occurrence of multiple ecosystems with varying probabilities. Relative to a single reference point on the ground, an evaluation of a label’s thematic accuracy may have multiple possibilities. Therefore, thematic accuracy and precision must be carefully established, given the large number of combinations and permutations of conditions that may practically occur within a PEM map label.

Reliability must also be carefully defined for similar reasons. However, reliability has been defined as a map's accuracy relative to its function or application. Therefore, a different set of criteria must be considered. These criteria include:

- The definition of the terms of accuracy and precision required for the PEM’s “primary intended purpose” (as well as, perhaps, some key stated secondary interpretations);
- Implications of thematic errors;
- Implications of positional errors;
- Implications of topological errors; and,
- Implications of global versus polygon accuracy.

It is worth noting that if a map is accurate and done for a stated explicit purpose (interpretation), then the interpretation accuracy (= reliability accuracy) will be accurate too. However, the inverse is necessarily true.

3.3 Terms of Positional Accuracy for PEM

Positional errors can have far-reaching implications for PEMs. They can be problematic for locating features on the ground but also because they cause confusion with polygon classification errors. As discussed in Section 2.5, however, it is possible to make an independent estimate of positional accuracy with well-defined features and appropriate reference data.

Sources of reference data include geo-referenced information that has a higher degree of accuracy. These could include:

- TRIM data;
- Ortho-photos; and,
- Ground-collected (and corrected) GPS data.

Statements about positional accuracy for PEMs should be, wherever possible, based upon well-defined features or points. Since ecological boundaries are interpreted (either by a human or digitally), they frequently do not fit this condition. More useful candidates for positional accuracy are those discrete points that can be identified with certainty on both the map and the source of higher accuracy, such as features that form sharply-defined intersections. Many natural boundaries, such as river-banks, are subject to fluctuation and as such, should be used only with proper cautions.

In order for the reference data to be considered a source of higher accuracy, its own positional accuracy must also be known. For TRIM data, positional accuracy standards are established and, for many applications of PEM, these may be acceptable standards. For ortho-photos and GPS data, accuracy is known or can be calculated. For other data sources, it may not be possible to establish accuracy easily. In some cases, a separate test of positional accuracy may not be required if the map utilizes the source of higher accuracy as data input. For a example, if features from TRIM are used to delineate polygons, those polygons may have the same accuracy as TRIM and the measurement of accuracy is either simplified or moot.

Mapping consistently using recognizable and predictive positional features is a challenge. Points, lines and polygons produced by a PEM must ultimately be considered as a “positional reality” in that they were put down and labeled with interpretive implications that will be used by a manager. Where a high level of positional accuracy is not achieved, there may still be considerable value in the product if the topological accuracy is still high, that is, that the relationships among up/down and adjacent units are still valid (See Section 3.4). However, when such situations are encountered, the map user then needs to “get onto the ground” because the positional accuracy of the units doesn’t support the level of resolution that your decision requires. Conversely, poor positional accuracy and good topological accuracy can readily support PEM applications and interpretations that are to be used for general-level planning requirements.

3.4 Terms of Topological Accuracy for PEM

For ecosystem maps that do not generate linework from photo-interpreted sources, there is a need for a test of the accuracy or "reasonableness" of the polygon linework in PEMs. (There is certainly a need to assess photo-interpreted linework as well, however, qualitative assessments are much easier to perform, as, e.g., in the current protocol for TEM QA). There are two components of topological reliability that can be characterized:

1. The appropriateness and accuracy of line placement; and,

2. The spatial distribution pattern and characteristics of polygons, as determined using descriptive statistics (e.g., measures of size and shape). This approach can be especially useful for comparative analyses, and is further discussed in Section 5.

The first item refers to the ability of the PEM procedure to appropriately and accurately delineate ecological features. Many features are relatively uniform and contain unambiguous characteristics from which a polygon can be defined and a classification can be made (for example, a river, clearcut or cliff may be easy to delineate). Other ecological features may be considerably more complex or more subtle, and a degree of interpretation is required to appropriately place the line. In these cases, there may be a range of appropriate locations for a line, and the appropriateness decreases with distance from the location of the feature on the ground. An example of this might be a subtle slope break or a change in surficial materials. A measure of the ability of the PEM to appropriately delineate these types of polygons in terms of ecologically meaningful units on the ground is important. It may be especially important where a PEM generates polygons and their contents in a manner that is removed from ground and photo-interpretation.

3.5 Collection of Reference Data

Many of the issues relating to the collection of reference data for PEM reliability assessment apply to all reliability assessment programs associated with resource mapping:

- Sources of reference data;
- Access issues;
- Polygon and point reference data types;
- Sampling design;
- Location of sample points;
- Requirements of a sample point; and,
- The need to balance cost/effort with result.

Another important factor for ecosystem map reliability is that many mapping approaches create polygons that are complex (i.e., polygons that contain more than one ecosystem). For TEM and possibly some PEMs, a complex label indicates that there are multiple ecosystem units present in the polygon. For PEMs that generate labels that are “belief values”, a complex label may simply identify the probability of occurrence of each ecosystem unit at that location. In either case, point observations alone may be inadequate for assessing complex ecosystem units, depending on the objectives of the reliability test. For example, line transects, with a series of points along the line, may provide a better characterization of a polygon’s contents (Wang 1982), although other methods of placing points within polygons to get unbiased sampling may also be considered. However, increasing the number of samples within a polygon decreases the number of polygons that could be visited or increases the cost of reliability assessment.

3.6 Integration with Model Calibration/Validation Procedures

Accuracy and reliability assessment have a potentially large effect on the model development process. Therefore, there is an opportunity for ongoing accuracy and reliability assessment to feed into PEM development. However, it must be remembered that in order to achieve an independent test of a PEM’s reliability, a reference dataset that has not been used to calibrate/validate the model must be used.

3.7 Integration with QA/QC Procedures

Reliability assessment procedures are related to QA/QC procedures in two main ways. First, good QA/QC is required so that the burden of project review does not fall entirely on the reliability assessment (already discussed in Section 1.5 and illustrated in Figure 1). Second, reliability assessment can be used to focus QA/QC procedures. For example, reliability assessment may identify a problem with a particular process in the generation of a map.

If the feedback loop is properly established, the QA/QC procedure can then be revisited so that it responds with closer monitoring, more explicit documentation of procedures, training and certification, or any number of possibilities that might ensure the creation of more accurate maps.

3.8 Timing of PEM Reliability Assessment

Within a given PEM's life cycle, reliability assessment is very time-sensitive, particularly in comparison to the reliability assessment procedures for TEM, which may be (but probably shouldn't be, for the same of the same reasons) conducted long after the mapping project is completed. Funding agencies must know that a PEM product is reliable, in order to determine contract fulfillment. Resource managers need to know what the magnitude and significance of errors are, and where they are propagated within the PEM. Specialists applying interpretations must know the suitability of the map for their interpretations, while the map producers need to be able to use the results of reliability assessment at the proper stage in the PEM life cycle, so that they can refine the map in the most effective manner. In order for these things to occur both during and upon completion of the project, reliability assessment procedures need to be in place at the inception of the project and then conducted concurrently with the PEM.

If reliability assessment is not conducted in a timely fashion, there may be delays in application of the product, the product may become operational before it has been properly assessed, there may be delays in contractor payment, and/or the relationship between reliability assessment and model development cannot be realized. One of the benefits of automated approaches to ecosystem mapping is that they can occur over a very short time frame, making PEM a very "collapsed" process. Spatial models can often be iteratively re-run and adjusted in short order, for example, based upon the findings of reliability assessment.

The capability to rapidly respond or adjust in the middle or near the end of the map generation phases makes PEM, as a procedure, quite unique from TEM. This characteristic also means that an organized, efficient and time-sensitive approach to reliability assessment be established for individual PEM projects.

3.9 Cost of PEM Reliability Assessment

PEM reliability assessments need to balance the cost and effort of reliability assessment with the needs of the reliability assessment results. There is a general expectation that PEMs, because they involve more automation, will be a more cost-effective inventory method than TEM. Reliability assessment components of PEMs will likewise need to be as cost-effective as possible. Because of this reality, statistically rigorous procedures may be very difficult to achieve. The greatest costs will typically be associated with the collection of ground-reference data.

PEM outputs that are “more removed” from primary data sources may require more effort toward reliability assessment (as well as toward input data screening and knowledge base management). The extent of reliability assessment effort may be considerably greater for PEMs that result in more complex or abstract spatial models, or more intricate or detailed versions of polygon boundaries and contents. This situation, in practice, can create a paradox where the cost of conducting reliability assessment potentially offsets those savings that may be realized by using an automated approach (Figure 3).

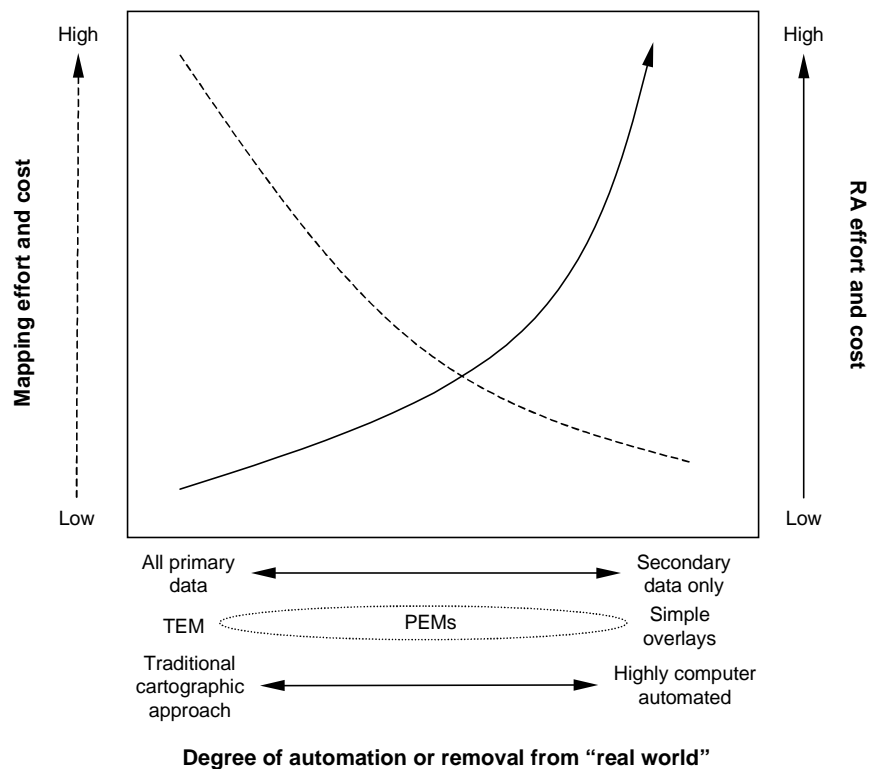


Figure 3. Ecosystem mapping that is more removed from primary data sources may require more effort toward reliability assessment. This creates a paradox where the cost of conducting reliability assessment may be higher for PEMs that generate complex outputs.

There are several ways that this paradox may be dealt with. For example:

1. There may be an intermediate point (e.g., the intersection of the two lines on Figure 3) where the combined costs of PEM map production and reliability assessment are less than at any other point along the spectrum. This point of intersection, if it can be established at the outset of a PEM project, represents an optimal condition where the PEM level of effort and the reliability assessment effort are “balanced”.
2. For established mapping protocols (e.g., those with RIC Standards), more effort toward QA/QC could reduce the need for reliability assessment (assuming reliability assessment is more expensive).

3. For established mapping protocols, not all portions of a PEM project may need to be evaluated. Reliability assessment, as an overall component of PEM, might be scaled back so that it is performed on a selected subset of phases within individual projects, or perhaps only on a randomly selected set of PEM projects.
4. Additional means of reducing the cost associated with the collection of ground-reference data could be considered (e.g., try to use, where it exists, existing ground-reference data (that has not been used for knowledge base validation), or photo-interpreted reference data as a replacement for ground-reference data collection, etc...).

Initially, for new or evolving PEM approaches, robust tests of reliability assessment will likely be needed. Only when the products and interpretations from PEM and the needs of resource planners are clearly understood can alternatives to robust tests of reliability be considered.

3.10 Role of Reliability Assessment in the Data Life Cycle

One of the challenges of reliability assessment is determining its position within the data life cycle and ensuring its incorporation into project metadata. Figure 4 shows the life-cycle of data in TEM and a hypothetical PEM, demonstrating that the two can be fundamentally different. TEM follows a linear process, more or less, with initial phases of field data collection, air photo interpretation and cartographic transcription, and digital data transfer, all with clearly-defined aspects of quality assurance/quality control. The TEM products are archived, but reliability assessment is then undertaken, usually, prior to the development of interpretations and applications for particular purposes.

The data life cycle of PEM is more iterative, in that the main development phase is a collapsed set of steps that involve PEM algorithm derivation, testing, implementation and verification. Distinct roles for QA/QC are more focused than for TEM. Reliability assessment – rather than an “add-on at the end” as it is in TEM – needs to be undertaken in concert with the iterative derivation of PEM outputs.

Custodians of the process are more conventional for TEM (Figure 4). For PEM, a small, focused, but multi-disciplinary team (mainly an ecologist working in concert with GIS database staff) conducts the spatial model development. The team’s expertise needs to include specific abilities in GIS and the PEM modeling approach being used, as well as vegetation ecology, soil and terrain science, forestry, wildlife habitat ecology, and possibly other disciplines as dictated by an individual PEM project.

Goodchild (1996) noted that it is critical that well-documented metadata accompanies resource inventory projects that are digitally archived. The lack of clear “hand-off protocols” for project responsibilities among the various custodians or “handlers of the data”, throughout the project life cycle, invariably leads to a critical loss or dilution of the quality of metadata (Goodchild 1996). As part of PEM reliability assessment, metadata that is as complete as possible needs to be prepared and reviewed by the custodians/data handlers of each phase of a project.

Standards for PEM project metadata need to be firmly established at the outset of a project. One clear advantage of PEMs, in this regard, is that the protocols used can be well documented and are quite amenable to defining these metadata since the whole process is quite explicit in terms of data used, knowledge base algorithms, and verification accuracies. In terms of repeatable protocols TEMs are not able to be as consistently applied in reality and mental interpretive models are seldom explicitly documented nor shared.

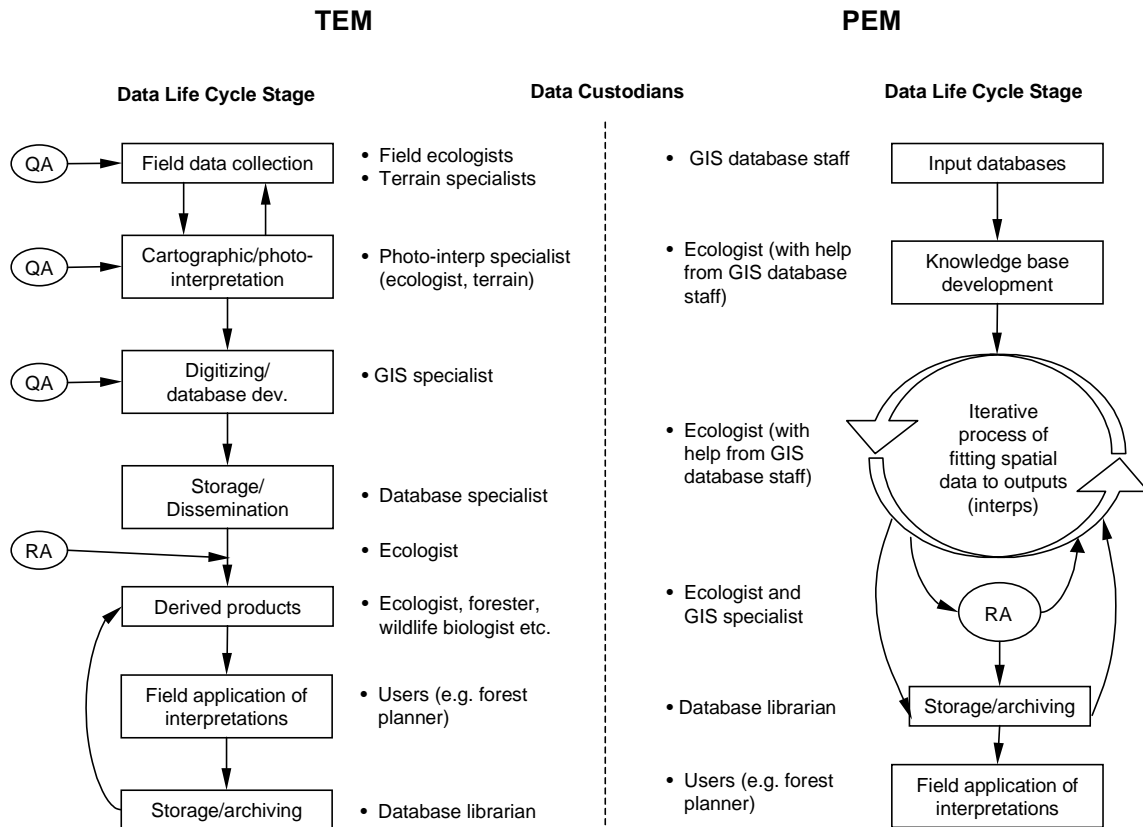


Figure 4. The life-cycle of ecosystem map databases (adapted from Goodchild, 1996).

4.0 PEM RELIABILITY ASSESSMENT FRAMEWORK

4.1 Required Components

A framework for PEM reliability assessment requires the following components:

1. The process needs to recognize reliability assessment in terms of the intended applications of a PEM project's outputs.
2. Universality – the reliability assessment process should be generic, such that it can be applied to any PEM product; at the same time, the method should also be adaptable to, or at least compatible with, reliability assessment processes developed for TEM projects.
3. Terms of accuracy and reliability should be readily understood, and therefore as simple and unambiguous as possible. Terms of reliability, in particular, need to be developed and defined – on a project-specific basis – in relation to explicitly-stated intended applications and interpretive products that will be generated, and delivered to primary and secondary users.
4. Independent point-based or polygon-based reference data need to be obtained and used. When reliability is tested at a site (point), there is a need to determine both the reliability of the input parameters used to classify the site and the allocation to a class. This then allows the assessor to determine if the classification is wrong, and whether it is due to the data or the interpretation of the data to allocate to a class.
5. Tests of accuracy and reliability must be cost-efficient, but sufficiently robust to provide unambiguous results.
6. The approach must be flexible enough to conform to new and different approaches to mapping, and to various ways of presenting/depicting results. Similarly, the process needs to be capable of interfacing directly with evolving decision tools (such that, as new technologies or new PEM approaches appear, they can be evaluated and integrated).

Figure 5 shows the major components of an approach to reliability assessment that satisfies the above-noted requirements. On the left-hand side of this flow-chart, a PEM product is initially created, consisting of points, lines and polygons, with associated attributes (e.g., in Figure 5, we list some of these as site series, modifiers, structural stages, and spatial attributes). As a basis for the RA process, reference data (either points or polygons) is then collected and directed towards accuracy assessment (which is defined in terms of both spatial and thematic elements), and reliability assessment (which is defined in terms of reliability, which in itself is based upon the interpretive products and intended applications). The output (e.g., an “inventory decision tool” is what we have indicated, in this instance, to represent an intended application) has then traveled through this assessment process.

Some operational mechanics of the framework shown in Figure 5 are discussed in the following sections.

4.2 Characteristics of PEM Reference Data

An independent set of point-based reference data (as per Figure 5) should be selected using stratified random sampling. The stratification used should very much be dictated by what statements about particular strata reliability (a single strata and in comparison with other strata) that is being aimed for. A carefully chosen stratification strategy could provide more useful information for subsequent map-field effort and/or for where a user should have greatest/least confidence.

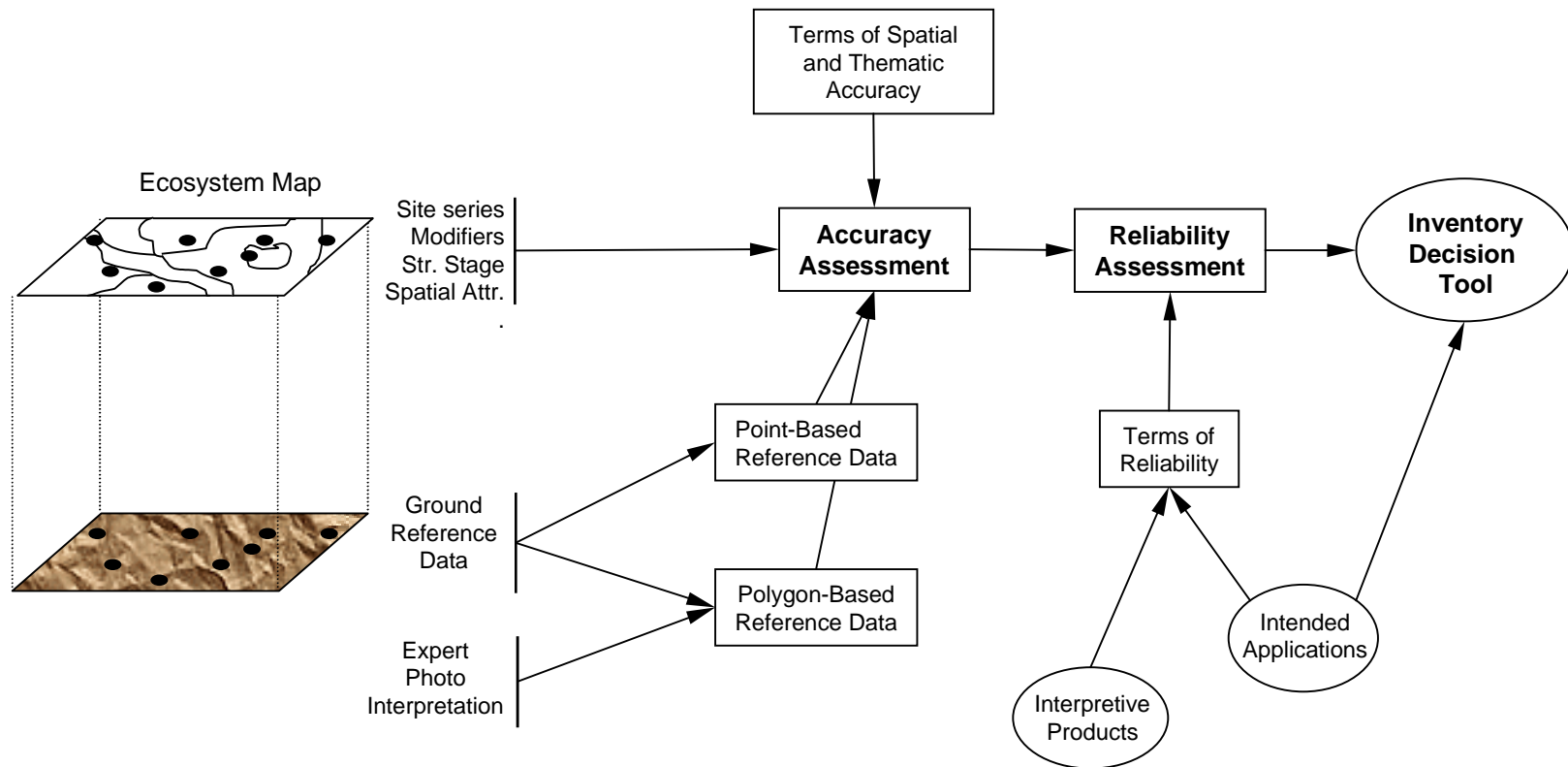


Figure 5. A framework for PEM reliability assessment. This is a generic framework that could be used for any ecosystem map.

The study area should be stratified in terms of:

- **Access** – The level of access within a study area needs to be initially assessed. A buffer zone should be established for sampling, based upon the selection of an acceptable pre-defined distance from a point of access (e.g., a road, trail or helicopter landing area). Width of the buffer should be dependent upon the ruggedness of the terrain, the level of effort and time that can be devoted to field sampling, and the degree of accessibility that exists throughout the study area.
- **BGC Units** – Measures should be taken to ensure that adequate numbers of field sample locations are located in each BGC subzone/variant that exists within the study area.
- **Ecosystem Units** – As for BCG units, stratification needs to be undertaken to ensure that a representative combination of site series, modifiers, structural stage, and/or other map attributes/features are included in the field sampling. As appropriate, the stratification process could be guided by the “importance” of individual taxon within the area being examined, as well as the potential risks associated with misinterpretations.
- **Special Considerations** – Depending upon a PEM project’s specific objectives, additional stratification or emphases may also be appropriate. For example, within a given study area that is being partially examined using polygon-based reference data, there may occur some sub-areas where photo-interpreted data alone would not be adequate, and so measures to include field sampling locations might be required.

Any supplementary polygon-based reference data (i.e., via expert photo interpretation; Figure 5) should also be selected following stratified random sampling. Some of the above-noted contexts should be followed in selecting photo-interpretation locations. It would also be appropriate to stratify the study area so as to identify locations where field access would be poor, such that ground plots could not be installed.

Ideally, sampling should be such that all delineations have a probability of being sampled proportional to their map area. The sampling protocol should attempt to achieve a minimum of 5 samples per ecosystem stratum. Where possible, these should consist of all point-based or all polygon-based reference data, just for consistency of interpretation, however this might be difficult to achieve and a mixture is likely to occur. For important strata with less than 5 occurrences, all occurrences should be sampled.

4.3 Accuracy Assessment

4.3.1 Spatial Accuracy

As discussed in Section 3.3, tests of positional accuracy may be difficult for most ecosystem maps if there are no features that qualify as “well-defined”. A test of spatial reliability (as described in the next section) would therefore be very important.

The appropriateness and accuracy of line placement needs to be characterized. Topological accuracy can be quantified in terms of correctly placing hard, intermediate and soft boundaries. Table 2 shows a linguistic scale that might be used to quantify this. The result of this analysis would be a histogram showing frequency of acceptable boundary placement for hard, soft and intermediate line types. A large part of this assessment should be narrative. Both ground-collected and photo-interpreted reference data can be used.

Topological accuracy is better than positional accuracy almost always. One confounding consideration, however, for topological accuracy is that it can exist among polygons (e.g., this site series is north of that

one, east of this one, generally within this area), but it can also occur internally within, in particular, compound units (e.g., where there is an explained internal pattern).

Table 2. A simplified linguistic scale that describes the level of difficulty to delineate a polygon.

Level of Difficulty	Polygon Delineation
1 – Easy	Polygon is defined by easily identified terrain and vegetative features (e.g., water, clearcuts, steep slopes, etc...).
2 – Moderate	Polygon is defined by moderately easily identified terrain and vegetational features (e.g., similar structural stages, etc...) and the polygon lines must be interpreted to a certain degree.
3 – Difficult	Polygon is not defined by easily identified terrain and vegetational features (e.g., drainage, similar surficial materials, etc...) and the polygon line must be interpreted.

4.3.2 Thematic Accuracy

As per the framework requirements, the terms of accuracy should be independent of the application or purpose of the map. The scale should be universal, or at least easily modified for different attributes and purposes. As well, the same value scale needs to be used for both point and polygon-based data.

An approach derived from fuzzy set theory (introduced in Section 2.6.3) can be used to satisfy these requirements. Table 3 lists a hypothetical accuracy scale that could be used for determining site series accuracy relative to both point-based and polygon-based reference data. This example assumes that each polygon label contains up to two belief values (probabilities of occurrence). This scale is only an example. It could contain any number of values or categories of accuracy.

Table 3. A possible site series accuracy scale for a generic PEM map using both point (ground-based) and polygon (photo-based) reference data.

Value	Degree of Accuracy	Point Data	Polygon Data
1a	Correct	Predicted as the 1 st decile.	All predicted units are present in the polygon.
1b	Correct	Predicted as the 2 nd decile.	The first or second predicted unit is present in the polygon but not both.
2	Acceptable	Within one on the edatopic grid for either decile.	The primary predicted unit is within one on the edatopic grid of the primary unit on the ground.
3	Incorrect	Completely wrong site series. Not close to the correct site series.	Completely wrong site series. Not close to the correct site series.

4.3.3 Calculating and Reporting Thematic Accuracy

After scores or values have been assigned to sample point according to the terms of accuracy, the frequency and magnitude of the errors can be calculated. Gopal and Woodcock (1988) used the term “fuzzy functions” to describe these calculations.

As an example, consider the terms of accuracy listed for site series in Table 3. The number of times that the correct site series (or combination of site series) is predicted could be calculated by dividing the number of correct predictions over the total number of samples. However an acceptable prediction might not need to be exact but within a certain degree of error (as defined in the terms of accuracy). The frequency of “close” predictions could then be calculated. This could be done for each degree of accuracy. By using this method, the magnitude of errors can be determined, provided that adequate terms of accuracy have been defined.

If the same accuracy scale is used for both point-based and polygon-based reference data, summary functions could be used on all sources and types of reference data cumulatively.

Qualitative descriptions of the nature of errors also provide valuable information. This allows the evaluator to describe characteristics of the map that are not easily, or worth the time, to quantitatively characterize.

4.4 Reliability Assessment

4.4.1 Defining Reliability Requirements

For the purposes of the framework (Figure 5), reliability has been defined as a measure of the degree or probability of the map satisfying its specified functions. Because an ecosystem map may be used for multiple applications or interpretations, a map may have multiple measures of reliability, and these will be specific to each interpretation that is anticipated from a PEM product.

At this point, terms of reliability, as part of the process shown in Figure 5, cannot be established until the requirements for map interpretations are defined. It is this aspect where the greatest amount of effort will be required before reliability assessment (as described in this document) can become operational. One approach to defining reliability requirements is presented in Table 4. For each interpretive product or intended application (row), each component of accuracy is qualitatively scored (column) with low, medium and high requirement for that component. Quantitative values could also be substituted, in those instances where suitable thresholds, ranges or midpoint values could be established.

As an example of this approach, in order to use a PEM product for “regional summary planning” (Table 4), it would require a moderate polygon accuracy, high global accuracy, low positional accuracy, and medium to high topological accuracy. The consequences of errors would be low, since the PEM polygon data was being rolled-up into a much smaller-scaled interpretation. Precision requirements (the scales by which accuracy are measured) would be dependent upon the variables that are being included for planning; for example, if “general forest cover” were one of these, the precision requirements for this attribute would be low or moderate.

Table 4. Preliminary accuracy requirements for the determination of minimum acceptable reliability levels for a range of PEM interpretive products or intended applications.

Interpretative Product / Intended Application	Polygon Accuracy	Global Accuracy	Positional Accuracy	Topological Accuracy	Consequences of Error	Precision Requirements
Gap Analysis	L	H	L	?	L	BEU, Site series, others
Plant Community Conservation	H	H	H	M	M	Site series and association
LRMP	L	M	L	L	L	BGC subzone variants, BEU
Soil And Site Sensitivity	H	M	H	M	M	Soil drainage, moisture, other soil-related variables
Surface Erosion Potential	H	H	H	M	M-H	Soil drainage, other bioterrain-related variables
Landscape Level Planning	M	H	L	M-H	L-H	Variable, depending upon objectives
Critical Wildlife Habitat	M-H	M-H	L	M-H	M	Site series, aspect, others
Identified Wildlife Habitat	M-H	M-H	L	M-H	M	Site series, structural stage, aspect
Timber Supply Analysis	M	M-H	M	M-H	M	Site series, structural stage, BCG, others
Site-Level Environmental Assessment	H	H	H	H	H	Various, depending upon the assessment's requirements
Regional Summary Reporting	L	M-H	L	M	M	Variable, depending upon objectives
Protected Areas Strategy	L-M	M-H	L-M	M-H	M	Site series, structural stage, BCG, various modifiers
Biodiversity Guidebook (FPC) Representation	L-M	M-H	L-M	M-H	M	Site series, structural stage, BCG, various others

L = low; M = moderate; H = high; N/A = not applicable.

4.4.2 Calculating and Reporting Reliability

It is more difficult to quantify reliability than to determine accuracy. In some cases, the map accuracy will also be its reliability if the map is used for the exact purpose as what it shows. In other cases, reliability for a particular application could be a score (Table 5). For example, in order for a map to be used for plant community conservation it must have high positional and topological accuracy. If a map fails to achieve

high accuracy (for example 70%), then the map would have moderate or low reliability for plant community conservation (e.g., see Table 4).

Reliability is assessed from the outputs of the accuracy assessment (Figure 5). This means that the estimation of accuracy and precision is the only component that requires field work or comparison of predicted classes or values against reference data. Reliability or utility for specific applications can be determined after estimation of accuracy and precision. The same approach could be used for evaluating input data quality of continuous, rank, or categorical variable attributes or taxonomic classes such as soil series, terrain units or ecosystem site series.

In any event, map reliability must be stated in terms of the method by which it was assessed and the application of the map. Even if a scoring approach is used, some of the reporting of map reliability may still need to be provided in a descriptive or narrative form.

Table 5. Example reliability ratings for an intended PEM application.

Reliability	Description
Low	Map is not appropriate for this application.
Moderate	Predictions made using this map for this application have a moderate degree of reliability. If the implications of an incorrect decision are severe, this map should not be used.
High	Predictions made using this map for this application have a high degree of reliability.

5.0 AN EXAMPLE CASE STUDY– THE BABINE EFMPP TEM/PEM COMPARISON

As part of the Babine Forest Products' Enhanced Forest Management Pilot Project (EFMPP) near Burns Lake, BC, Ecological Land Data Acquisition Resource (ELDAR (see Mulder and Corns 1996)), Vegetation Resource Inventory with Estimated Site Series (VRI-ESS) and Virtual Ecosystems (VEC) are being or have been completed for an area where TEM has already been completed. In co-operation with Babine Forest Products, R.A. Sims and Associates is undertaking a reliability assessment and comparative analysis of these four map products.

Although the project has not yet been completed, the development of the framework and preliminary analyses have been completed. The framework is presented here to illustrate a multi-faceted approach that attempts to compare map characteristics and their accuracy and reliability for various applications.

The project utilizes several approaches to compare and assess reliability of each of the ecosystem maps. Figure 6 illustrates the analytical process. There are four basic components, and a separate testing tool is being developed to separately address each component:

- Thematic Reliability;
- Topological/Positional Accuracy;
- Spatial Comparisons; and
- Thematic Comparison.

The approach to thematic reliability and topological accuracy are similar to the methods described in this document. For the spatial comparison, a GIS analysis is calculating and treating various summary statistics, such as:

- Average polygon size;
- Average polygon size for each map category;
- Polygon shape;
- Polygon angularity (another measure of shape);
- Average polygon perimeter;
- Number of polygons per unit area; and,
- Total area of each ecosystem taxon.

These statistics are calculated using Arc/Info GIS and the statistical program FRAGSTATS (McGarigal and Marks 1994).

In general, the approach we are using in the “Babine TEM/PEM Comparison Project” indicates that accuracy and reliability assessment requires that several complementary analyses be undertaken, in order to quantify the various contexts of reliability. As earlier advocated in this document, it is necessary to construct an assessment system that is, to a degree, customized for each PEM project. In constructing the current system, particular consideration was given to:

- the study area's physical characteristics (e.g., in particular, access and the diversity of BCG and ecosystem units);
- manners in which the PEM projects evaluated polygon linework and annotations;
- the PEM project's overall objectives, life cycle and associated custodians / data handlers; and,
- the intended interpretations and those applications that the final PEM product was being used for (mainly, in this case, wildlife interpretations and forest planning objectives).

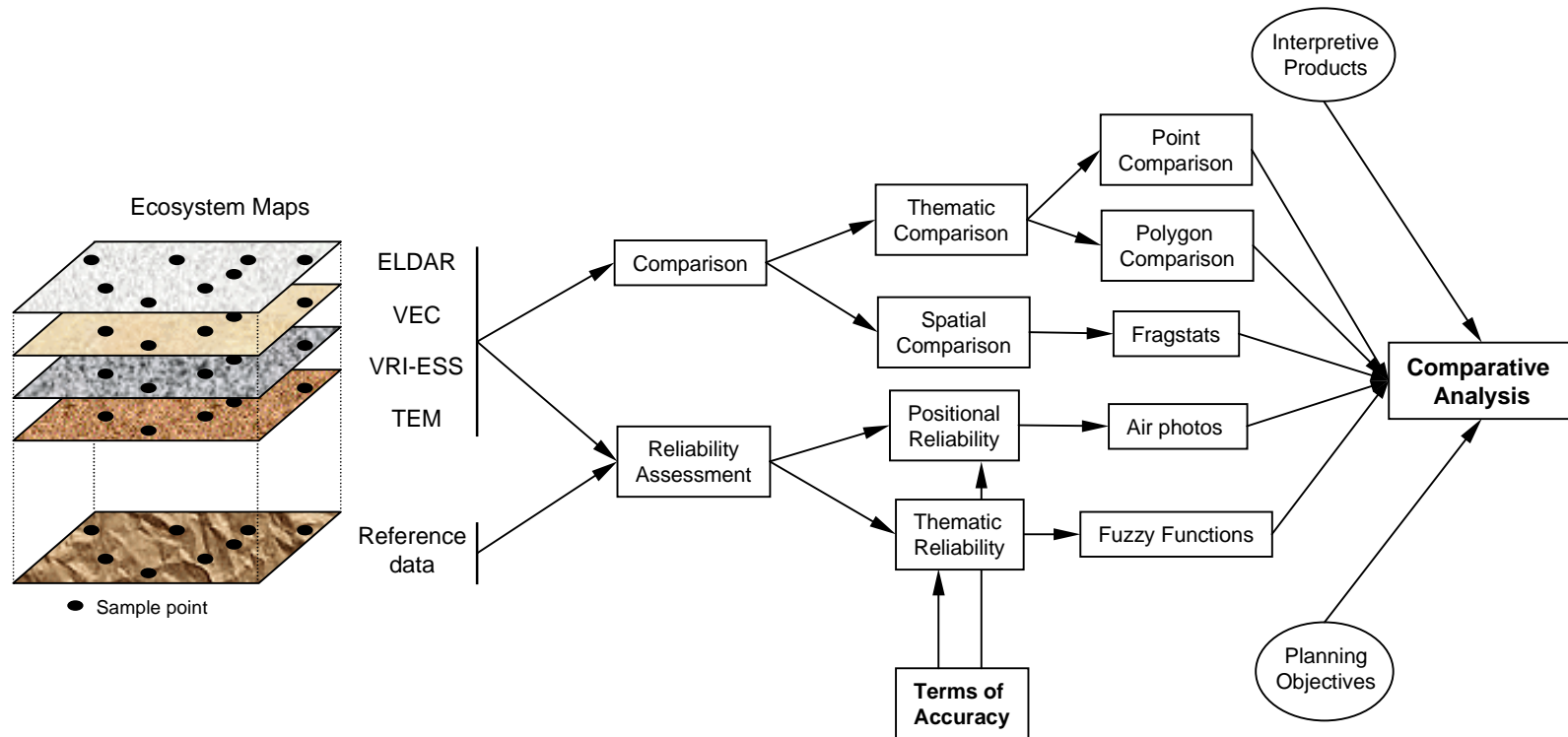


Figure 6. Ecosystem map comparative analysis process for the Babine EFMPP. Several analytical methods are being utilized to compare and assess reliability of four types of ecosystem maps of the same area.

6.0 CONCLUSIONS AND RECOMMENDATIONS

6.1 General Conclusions

Reliability assessment is a significant issue for all resource inventories. It is particularly evident as an issue in ecosystem mapping because the mapping of ecological relationships can be a complex and multidimensional issue. Directed independent checking is required in order to confirm characteristics that are not readily observed on the ground using simple observational principles.

Typically, explicit tests of accuracy and reliability are not conducted for resource inventory maps. Errors in a map are often believed to “cancel each other out” or, in some other manner, to disappear or be negligible in their overall effect upon an individual map or study area. Because overall accuracy and reliability are difficult to accurately determine, they are often assumed to have a low effect on a map’s informational content. A variety of studies have shown, however, that many thematic maps used in natural resource management have very low levels of reliability. Given the challenges that exist for the assessment of existing traditional map products, there are some major obstacles that must be resolved before operational assessments of PEM can be made.

PEM outputs cannot be directly compared to linework on air photos and visually-interpreted annotations, and must instead be considered in relation to abstract spatial models and GIS-based algorithms. Consequently, there is an increased need for objectively-determined measures of accuracy and reliability in PEM projects – certainly more so than for conventional forest inventory, TEM, or other types of resource maps.

6.2 Recommended Next Steps for Reliability Assessment of PEMs

General themes and constraints for reliability assessment of PEM have been considered and discussed in this report. The following steps still need to be individually addressed in more detail, prior to implementing any standard procedure for PEM reliability assessment. These items can be developed through either a strategic planning and testing effort, or via some directed pilot projects.

- The needs of resource managers/planners must be firmly established, so that PEM outputs, in particular interpretive products, are properly developed.
- The need for reliability assessment increases as PEMs become more removed from primary data sources. This necessitates that some boundaries be established – at some point – as to what is acceptable and not as a PEM tool or technology.
- The expectations for spatial accuracy of PEM need to be more firmly established, as there are some interpretations that may not be able to achieve the accuracy (and precision) that is needed for detailed decision-making.
- The PEM reliability assessment system that evolves from this exercise should attempt to partition the errors in mapping, so that systematic problems can be identified, and so mitigation efforts can be focused. This may be a difficult thing to put into practice, but a set of tools that effectively partition errors would provide the best solution.
- PEM reliability assessment metadata must be incorporated into project metadata standards. A mechanism for this needs to be defined and described, ideally, at the outset.
- A mechanism is needed for PEM change management, given the potential for rapid technological change with PEM-related tools. Only when the products and interpretations from PEM and the

needs of resource planners are clearly understood, can alternatives to robust tests of reliability be considered.

- Future considerations about PEM map reliability should also be considered as part of a larger initiative involving ecosystem map reliability as a whole in BC, so that the right ecosystem mapping systems are applied under the appropriate circumstances. One component of this is monitoring the evolving needs for new types of interpretive products or intended applications for PEMs.

Implementation of PEM as an operational tool presents a number of challenges. The development of an effective approach to reliability assessment is fundamental in the implementation process. This report has presented and discussed a range of issues related to PEM reliability assessment. Using these – collectively – as context, we hope that superior tools and methodologies can now be designed, developed and tested for PEM QA/QC, accuracy assessment and reliability assessment.

7.0 GLOSSARY

- Accuracy:*** Measure of the absence of error. Also a measure of how close an estimate is to its true value. Accuracy is typically interpreted in relation to the degree of precision (exactness or resolution) with which it is measured. It is important to explicitly state how accuracy is being assessed (i.e., by what measures of precision).
- Polygon accuracy:*** Accuracy at the level of the polygon.
- Precision:*** A measure of the absence of error or a measure of how close an estimate is to its true value. It is important to explicitly state how accuracy is being assessed (i.e., by what measures of precision).
- Global accuracy:*** Accuracy over a larger area than a polygon; for the purposes of this report, we envisage this at the 1:20,000 mapsheet level, although other levels of generalization are possible.
- Positional Error:*** Discrepancy between the coordinate location of a point in the database and coordinate of the same point in a source of higher accuracy.
- Positional Accuracy:*** The degree of conformity among a set of observations. Expresses the degree of repeatability of a measurement with respect to the positional component of spatial accuracy.
- Predictive Ecosystem Mapping (PEM):***
Methods of stratifying the landscape into ecologically-oriented map units on the basis of combinations of existing mapped features, point information and knowledge about ecosystems-landscape relationships and their distribution pattern. (TEM Alternatives Task Force, 1998).
- Reliability:*** A measure of the degree or probability of the map satisfying its specified functions.
- TEM Alternatives:*** Interim or alternative TEM-like approaches that products similar to conventional TEM. These methods are normally aimed at reducing inventory costs over traditional approaches. (Terrestrial Ecosystem Mapping Alternatives Task Force, 1998).
- Thematic accuracy:*** Discrepancy between the class at a particular point in the database and the class observed at the same coordinate location in the field.
- Topological accuracy:*** The relative accuracy of a mapped unit in relation to other mapped units; their degree of accuracy in terms of their relative juxtapositions on the landscape. Expresses the degree of repeatability of a measurement with respect to the topological component of spatial accuracy.

8.0 REFERENCES

- Anon. 1998. Notes for submissions for QA of TEM Projects. Res. Inventory Br., Min. Environ., Lands and Parks, Victoria, BC. Unpubl. Document (MELP TEM website).
- Biggs, W., P. Eligh, R. Sims, and R.J. Wiart. 1997. A Business Approach to Terrestrial Ecosystem Mapping (TEM) and TEM Alternatives. Resource Inventory Committee, Victoria, BC. 61pp.
- Clark, D., D. Meidinger, and F. Schultz. 1998. Background Information for Detailed Supporting Studies for Evaluation, Role and Development of Standards for Predictive Approaches to Terrestrial Ecosystem Mapping (TEM). Prepared for Terrestrial Ecosystem Mapping Alternatives Task Force, Resource Inventory Committee. Victoria, BC.
- Congalton, R.G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 37: 35-46.
- Goodchild, M. and S. Gopal. 1989. Accuracy of spatial databases. Taylor & Francis, Ltd., London.
- Goodchild, M.F. 1996. Communicating the results of accuracy assessment: Metadata, digital libraries, and assessing fitness for use. National Center for Geographic Information and Analysis, and Department of Geography of the University of California. Unpubl. Rep. No. 13 pp.
- Gopal, S. and C. Woodcock. 1994. Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photo. Eng. & Remote Sens.* 60(2): 181-188.
- McGarigal, K. and B. Marks. 1994. FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure. Version 2.0.
- Moon, D.E. 1999. Problem Analysis on Data Quality Assessment Issues. Draft. Submitted to TEM Alternative Task Force by CDT Core Design Technologies Inc.
- Moon, D.E., J.W. Hall, and C. Selby. 1987. A procedure for determining ground-truth reliability for land resource surveys. Ver. 2. Land Res. Research Centre, Ag. Canada, Vancouver, BC.
- Mulder, J. A. and I.G.W. Corns. 1996. ELDAR: A decision support tool for ecological map creation and field data collection. Alberta Research Council: Calgary, AB. 5 pp.
- National Biological Survey and National Park Service (NBS/NPS). 1994. Accuracy Assessment Procedures. NBS/NPS Vegetation Mapping Program. USDI. Prepared by Environmental Systems Research Institute, National Center for Geographic Information and Analysis.
- Osborn, L. 1997. Resources Inventory Strategy for Northwest British Columbia. RIC, MoF, MoELP, FRBC. 25pp.
- Resources Inventory Committee (RIC). 1998. Standards for Terrestrial Ecosystem Mapping in British Columbia. Ecosystem Working Group, Terrestrial Ecosystems Task Force, Ministry of Environment, Lands and Parks, Victoria, BC. 190 pp.
- Sims, R.A. 1997. Quality Assurance in Terrestrial Ecosystem Mapping: A Quantitative Method for Assessing Accuracy. Geomatics International, Vancouver, BC. Unpubl. Draft.

Story, M. and R.G. Congalton. 1986. Accuracy assessment: a user's perspective. *Photo. Eng. & Remote Sens.* 52(3): 397-399.

TEM Alternatives Task Force. 1998. Decision Matrix for TEM and TEM-Alternatives. Unpublished document. 4 pp.

Wang, C. 1982. Application of a transect method to soil survey problems. *Land Resources Res. Inst. Agri. Can., Ottawa, ON. LRRI Contrib. No. 82-02.* 34 pp.