Final Draft

Accuracy Assessment Procedures

USGS/NPS Vegetation Mapping Program

November 1994

Prepared for:

United States Department of Interior
U.S. Geological Survey and National Park Service

Prepared by:

Environmental Systems Research Institute
380 New York Street
Redlands, California 92373

National Center for Geographic Information and Analysis
University of California
3510 Phelps Hall
Santa Barbara, California 93106

The Nature Conservancy
1815 N. Lynn Street
Arlington, Virginia 22209

ESRI, ARC/INFO, PC ARC/INFO, ArcView, and ArcCAD are registered trademarks of Environmental Systems Research Institute, Inc. ARC/INFO COGO, ARC/INFO NETWORK, ARC/INFO TIN, ARC/INFO GRID, ARC/INFO LIBRARIAN, ARCSHELL, ARCEDIT, ARCPLOT, ARC Macro Language (AML), Simple Macro Language (SML), DATABASE INTEGRATOR, IMAGE INTEGRATOR, WorkStation ARC/INFO, ArcBrowser, ArcCensus, *ARC News*, ArcKits, ARCware, ArcCity, ArcDoc, ArcExpress, ArcFrame, ArcScan, ArcScene, ArcSchool, ArcSdl, ArcStorm, ArcTools, ArcUSA, ArcWorld, Avenue, FormEdit, Geographic User Interface (GUI), Geographic User System (GUS), Geographic Table of Contents (GTC), ARC Development Framework (ADF), PC ARCEDIT, PC ARCPLOT, PC ARCSHELL, PC OVERLAY, PC NETWORK, PC DATA CONVERSION, PC STARTER KIT, TABLES, University LAB KIT, the ESRI corporate logo, the ARC/INFO logo, the PC ARC/INFO logo, the ArcView logo, the ArcCAD logo, the ArcData logo, ESRI—Team GIS, and ESRI—The GIS People are trademarks of Environmental Systems Research Institute, Inc. ARCMAIL, ArcData, and Rent-a-Tech are service marks of Environmental Systems Research Institute, Inc.

Other companies and products herein are trademarks or registered trademarks of their respective trademark owners. The information contained in any associated brochures is subject to change without notice.

Table of Contents

Section **Page**

Executive Summary **1-1**

1.0 Introduction **1-4**
1.1 Objectives of the Report 1-4
1.2 Scope of the Document 1-4
1.3 Relationship to Other Reports in this Series 1-5
1.4 Structure of the Document 1-5
1.5 Introducing the Proposed Method 1-6

2.0 Theoretical Aspects of Error Measurement in Vegetation Mapping **2-1**
2.1 Modeling Natural Properties 2-1
2.1.1 Homogeneity of Classes 2-1
2.1.2 Distribution of Classes 2-2
2.1.3 Boundary Definition in Transition Zones 2-3
2.1.4 Relationship between Smallest Mapping Unit and Smallest Thematic Unit 2-4
2.1.5 Spatial Patterning of Thematic Units 2-4
2.1.6 Thematic Implications of Spatial Errors 2-5
2.1.7 Cartographic Abstraction 2-5
2.1.8 Effects of Scale 2-6

2.2 Error Detection 2-6
2.2.1 Definitions of Error and Accuracy 2-6
2.2.2 Considerations that Affect the Selection of a Source of Higher Accuracy 2-7

2.3 Sources of Error 2-7
2.3.1 Sources of Thematic Error 2-7
2.3.2 Sources of Positional Error 2-9

3.0 Practical Aspects of Error Measurement in Vegetation Mapping **3-1**
3.1 Quality of Sources of Higher Accuracy 3-1

3.2 Operational Constraints 3-2
3.2.1 Dealing with Accessibility Problems that Affect Randomness of Sampling 3-2
3.2.2 Dealing with Temporal Effects, Season, and Seasonal Variability 3-3

4.0 Sample Collection for Accuracy Assessment 4-1
4.1 Sampling Design 4-1
4.1.1 Sampling Design and Data Collection Objectives 4-1
4.1.2 Sampling Design for Purposes of Classification System Development 4-1
4.1.3 Sampling Design for Purposes of Accuracy Assessment 4-2

4.2 Sampling Methods 4-3
4.2.1 Simple Random Sampling 4-3
4.2.2 Stratified Random Sampling 4-4
4.2.3 Systematic Sampling 4-4
4.2.4 Two-Stage Sampling 4-4
4.2.5 Recommended Sampling Methods 4-4

4.3 Sample Site Selection 4-5
4.3.1 Sample Site Location in the Field 4-6
4.3.2 Sampling for Thematic Accuracy Assessment 4-7
4.3.3 Sampling for Positional Accuracy Assessment 4-9

4.4 Determining Sample Size 4-10
4.4.1 Determining Acceptable Levels of Error and Confidence 4-10
4.4.2 Sampling of Rare and Abundant Classes 4-12

4.5 Data Collection Methods 4-14

5.0 Computational Issues 5-1
5.1 Definitions of Precision and Accuracy 5-1
5.2 Statistical Methods for Thematic Accuracy 5-2
5.3 Statistical Methods for Positional Accuracy 5-5
5.4 Confidence Intervals and Sample Size 5-7
5.5 Hypothesis Testing 5-9
5.5.1 Determining Conformance to NMAS for Positional Accuracy 5-10
5.5.2 Determining Conformance to an 80% Accuracy Requirement for Thematic Accuracy 5-10

5.6 Summary of Recommendations Regarding Confidence Levels and Hypothesis Tests 5-13

6.0 Error Management Issues 6-1
6.1 Accumulation and Tracing of Errors in Output Products 6-1
6.2 Differentiation of Regional Error Rates 6-2
6.3 Weighting of Errors 6-3
6.4 Cost and Benefit Considerations 6-4
6.5 Capturing and Incorporating Uncertainty as Part of the Data 6-6
6.6 Quality Assurance 6-9
6.7 Effective Use of Local Expertise 6-9
6.8 Process Design 6-10

7.0 The Accuracy Standard for the USGS/NPS Vegetation Mapping Program 7-1
7.1 Establishing Thematic and Positional Control 7-1
7.2 Thematic Accuracy Requirements 7-4
7.3 Positional Accuracy Requirements 7-5
7.4 Dealing with Problem Areas and Classes 7-7

8.0 Implementation 8-1
8.1 Overall Plan and Operational Scenarios 8-1
8.2 Sampling and Data Collection 8-9
8.2.1 Number of Samples 8-9
8.2.1.1 Total Number of Sample Sites 8-10
8.2.1.2 Number of Sites per Class 8-11
8.2.2 Sample Site Selection 8-12
8.2.2.1 Stratification 8-13
8.2.2.2 Initial Positioning of Sample Sites 8-13
8.2.2.3 Sampling between and along Gradients 8-13
8.2.3 Sample Collection 8-14
8.2.4 Reporting Results 8-14
8.2.5 Field Assessment Form 8-14
8.3 Statistical Methods 8-15
8.3.1 Positional Accuracy 8-15
8.3.2 Attribute Accuracy 8-16
8.3.3 Reporting Results 8-16
8.3.4 Labeling of Products 8-17
8.3.5 Corrective Measures 8-18

8.4 Alternative Methods 8-18
8.4.1 Subsampling Strategies 8-19
8.4.2 Observational Strategies 8-20
8.4.3 Strategies for Using Alternatives 8-22

8.5 Local Options 8-24

9.0 Operational Testing 9-1
9.1 Testing the Proposed Method 9-1
9.2 Refining the Process 9-2

10.0 Glossary 10-1

11.0 Literature Cited 11-1

12.0 List of Authors and Contributors 12-1

List of Tables

Table 1: Relationship between Sample Size, Confidence Intervals, and Acceptable Error 4-11

Table 2: Sample Misclassification Matrix for Five Classes 5-2

Table 3: Users' and Producers' Accuracy for Five Classes 5-4

Table 4: Aggregation of Two Easily Confused Classes from Table 2 5-5

Table 5: Confidence Intervals for Given Sample Sizes and Confidence Levels 5-8

Table 6: Number of Sites per Class 8-11

List of Figures

Figure 1: Accuracy Assessment Procedures Process Flow 8-3

Executive Summary

The aim of the National Park Service/U.S. Geological Survey (NPS/USGS) Vegetation Mapping Project is to map 237 Inventory and Monitoring Park units using a single, standard vegetation classification. The primary data to be utilized for the vegetation mapping is aerial photography combined with ancillary field data. The final output product will be a digital map of the vegetation. An essential part of this ambitious mapping project will be an accuracy assessment of the final database. Accuracy assessment is important because estimates of thematic and positional errors in the data will allow users of the data to assess data suitability for a particular application. At the same time, data producers will be able to learn more about the nature of errors in the data and improve the mapping process.

Accuracy requirements for the NPS/USGS Vegetation Mapping Project specify an 80% accuracy for each vegetation class that is mapped. In terms of positional accuracy, the data are expected to meet National Map Accuracy Standards (NMAS). At a minimum, accuracy assessment procedures should be able to determine, with sufficient precision, whether the vegetation map meets these requirements.

The objectives of this report are to provide the theoretical framework for accuracy assessment and to make initial recommendations as well as alternative procedures for accuracy assessment of the NPS/USGS Vegetation Mapping Project. The procedures must be scientifically sound and yet practical to implement. The core recommendation is to utilize a basic procedure that is statistically rigorous and consistent with traditional methodologies. However, it is recognized that operational constraints may preclude a complete implementation. For this reason, some alternatives for implementing the preferred scenario are provided with the expectation that operational testing during the prototype phase of the project will be necessary to define the utility and reliability of the alternative procedures.

The report reviews the problem of accuracy assessment for vegetation mapping in terms of theoretical and operational constraints, such as the interpretive nature of vegetation mapping, relative abundance of various classes, temporal change, sample site location, and accessibility problems during field sampling. These constraints ultimately impact how errors are interpreted and the degree of precision with which (traditional) accuracy estimates can be derived for vegetation map data. The usefulness of traditional statistical analysis methods is seen as somewhat limiting for accuracy assessment of vegetation data, because definitions of "error" and "accuracy," within the framework of traditional statistical methods, leave little room to account for interpretive judgment calls that are an important part of vegetation mapping. Despite these limitations, the recommended accuracy assessment procedures

utilize traditional statistical methods to estimate thematic and positional accuracy. At the same time, the authors strongly recommend prototyping of experimental methods which take into account the special characteristics of vegetation data.

Key recommendations made in terms of accuracy assessment for the NPS/USGS Vegetation Mapping Project include the following:

- The accuracy assessment should be independent from the mapping process itself. Since the vegetation database will be compiled as a series of park-specific projects, a separate accuracy assessment should be conducted for each park.

- The accuracy assessment should be point based with an observational unit equivalent or larger than the minimum mapping unit.

- The recommended number of samples per class should reflect the abundance of each class within the project area. Rare classes should be sampled with less frequency than abundant classes. As such, the accuracy of rare classes will be stated with less precision than that of abundant classes.

- Thematic accuracy should be expressed using contingency matrices adjusted for chance agreement with a Kappa index. It is recommended to report users' and producers' accuracy for each class. These accuracies should be expressed as a percentage with a 90% confidence interval. Positional accuracy should be expressed as a root mean square error. It is recommended users be provided with the actual accuracy estimates rather than stating whether or not the product meets a specific accuracy standard.

- Ideally, accuracy assessment should capture all components of uncertainty associated with vegetation mapping. Recognizing that operationally this may not be feasible, it is nevertheless recommended to test experimental methods measuring, for example, within polygon variation, or the uncertainty in the position of polygon boundaries.

- Prior to implementing any of the accuracy assessment procedures, they should be tested operationally during the prototyping phase of the project. It is anticipated that the methodology will need to be refined as a result of the testing.

Arguments in support of these recommendations can be found in the relevant sections of this report.

1.0 Introduction

1.1 Objectives of the Report

Accuracy assessment in spatial data has three primary objectives:

1. To allow users to assess the data's suitability for a particular application

2. To allow producers of the spatial data to learn more about errors in the data and improve the mapping process

3. To verify conformance to production standards

The primary objective of this document is to define suitable procedures for assessing the accuracy of the digital data acquired for the NPS/USGS Vegetation Mapping Project. Procedures will be evaluated as suitable if they are scientifically sound and justifiable and yet are practical to implement (feasible in view of cost and time constraints).

1.2 Scope of the Document

It is generally accepted among mappers of vegetation data that no single accuracy assessment procedure is optimal for application to all vegetation classes or for all areas included in a particular project. Also, it is recognized that any chosen method should be operationally tested prior to becoming the standard method of assessment (Stoms et al. 1994).

This document is intended to provide the theoretical framework for the measurement of thematic and positional accuracy associated with vegetation data. As such, the main focus of the document is to define appropriate assessment procedures and justify the selection based on theoretical considerations, especially in cases where several alternative methods exist. In terms of accuracy assessment procedures, the focus of this report is to make recommendations on the following:

- Sampling strategy in terms of sample size, sample location, and the minimum sampling unit that yields both thematic and positional accuracy

- The implementation of the sampling strategy
- Alternative methods for accuracy assessment, including any decision rules that are different when selecting alternative methods to be used
- The analytical methods to be used for the analysis of the samples
- The interpretation of accuracy statements

The recommended methods will need to be validated by being tested operationally during the prototyping phase of the project. The purpose of such operational testing will be to refine those aspects of the accuracy assessment procedure that cannot be finalized from theoretical considerations alone. Operational testing of the accuracy assessment procedure should therefore be seen as an integral part of accuracy assessment procedures development.

1.3 Relationship to Other Reports in this Series

This report is the third of a set of three documents being completed to delineate the methods proposed for application to the NPS/USGS Vegetation Mapping Project. The first of the three studies proposes the vegetation system for use as the classification and mapping standard to achieve the objectives of the project. The second report describes the methodology required to implement the vegetation classification and mapping of the parks. This, the third document, outlines the accuracy assessment procedures that will be utilized to verify the thematic and positional accuracy in the vegetation maps.

1.4 Structure of the Document

This document describes the theoretical framework behind accuracy assessment for vegetation mapping in general, the accuracy requirements of the NPS/USGS Vegetation Mapping Project in particular, and the implementation of assessment procedures recommended for thematic and positional accuracy assessment. Section 1 is this introduction. Sections 2 through 6 describe the error model for vegetation mapping. They include definitions of error and accuracy in the context of vegetation mapping, methods for error detection and measurement, computational methods, methods for error management, and process design. Section 7 describes the accuracy requirements for positional and thematic accuracy for the NPS/USGS Vegetation Mapping Project.

Section 8 deals with the implementation of the recommended accuracy assessment procedure in terms of data collection and computational methods. Section 9 briefly addresses operational testing of the proposed method. Finally, Section 10 provides a glossary of terms and definitions.

1.5 Introducing the Proposed Method

The proposed accuracy assessment procedure assumes that the NPS/USGS Vegetation Mapping Database will be compiled as a series of park-specific projects. The principal method of data compilation will be based on aerial photographs from which initial delineations of vegetation boundaries will be made. The mapping of the vegetation classification will then be refined by using field sampling methods. The minimum mapping unit (MMU) for the project is defined as 0.5 hectare.

This methodology assumes that the classification system itself permits an initial mapping based on aerial photographs. Since thematic accuracy assessment will be affected to some extent by the characteristics of the classification system (especially its stated resolution, class aggregation methods, and minimum mapping unit size), some procedural aspects of the accuracy assessment will evolve with the development of the classification system. However, bearing these constraints in mind, the recommended procedure can be summarized as follows:

1. The accuracy assessment will be park specific and will be concerned with verifying the accuracy of the vegetation mapping.

2. Accuracy will be assessed at the conclusion of the mapping process for a specific park, and will therefore be independent of the mapping process.

3. The most desirable, though not exclusive, source of higher accuracy against which the classified data will be compared is field checking.

4. Both thematic and positional accuracy will be assessed.

5. For thematic accuracy assessment, a stratified random sampling approach which covers the entire park area is recommended. It is recognized that logistics and/or costs may prevent the entire park from being sampled, in which case, some variation of two-stage or nested sampling may have to be considered. The theory is discussed in Section 4 and

the operational aspects of this alternative are discussed in Section 8.

For either situation, five class-specific scenarios have been identified. The procedures or each scenario vary primarily in terms of sample size and in terms of the decision rules regarding sample site location and sample site availability. It is expected that this approach is most appropriate for medium-size parks (2,500–50,000 hectares). Very small and very large parks may require modification of the basic procedures as addressed in Section 8.

The core thematic accuracy assessment procedures are defined as follows:

- The number of sample sites per class will vary from five to thirty, depending on the frequency and abundance of the class. Classes that consist of fewer than five polygons and cover less than 50 hectares of total park area will be observed in their entirety.

- Multiple sample sites per polygon will be allowed; however, the method by which sample sites will be allocated to polygons will be weighted to give preference to larger polygons. Because of the cost associated with visiting sample sites, a larger than needed number of sample sites identified for the given scenario will be selected before the field visit. If sample sites prove inaccessible during the field visit, they will be replaced in the field if possible, or during post-field analysis. In the latter case, a second visit to the field will be necessary.

- Sample sites will be randomly allocated within a class, but they will be positioned away from the polygon boundaries by an amount equivalent to the positional error in the data and the error associated with locating the sample site in the field. At a minimum, the sample site will be positioned far enough from the boundary to be wholly contained within one community.

- Sample sites will be located by means of point samples (i.e., random coordinate locations). For each point location, an area equivalent to or larger than the MMU will be observed. The basic guideline is to observe a circular area about the point location, provided the circular area is wholly contained within a single community

and can be conveniently staked out. Investigators will be given interpretive freedom to vary the shape of the sampling unit to accommodate special community types such as riparian zones or areas where circular sampling units are difficult to use (e.g., in forested areas). These types of decisions will need to be documented by field personnel.

■ Sample sites Sample sites will be located in the field by using Global Positioning Satellite (GPS) surveying methods. Depending on the method chosen, this will permit positioning of the sample site with an accuracy of from 100 meters (using uncorrected observations) to 1 to 5 meters (using real-time differential GPS). Utilizing methods that permit positioning of the sites without post-field processing will be preferable, but may not always be practical. If post-processing is required, investigators will need to utilize collateral data in conjunction with GPS to locate the field site as accurately as possible.

■ All pertinent information, including the name of the investigator(s), the date and time of the visit, the method of ground control, field conditions, the method of observation, special decision rules in analyzing the field site, and the vegetation class observed for the area, will be recorded on a field assessment form.

6. Positional accuracy assessment will be limited to estimating how well the data can be tied to its control. The sample points used will be different from those used for thematic assessment, because positional accuracy assessment depends on the availability of well-defined points. Candidate points will initially be selected from collateral sources such as USGS quad sheets or aerial photographs. Sample points will be distributed throughout the area where possible, in order to derive an estimate that is applicable to the entire mapped area. If sample points cannot be well distributed, the positional accuracy derived will be recorded with a statement as to the area to which the estimate applies. Sample points will be located in the field using differential GPS surveying. The coordinate locations of the sample points will be post-processed before they are used in other computations.

7. Since vegetation class assignment and delineations of vegetation boundaries are subject to much interpretive judgment, it is recommended that alternative methods that capture the uncertainty in the position of a class boundary and the variation within "single class

polygons" be considered for testing. These methods include the following:

- Estimating epsilon bands, which give an indication of the uncertainty inherent in the position of the lines
- Capturing the uncertainty of individual lines as estimated by the investigator in the field or during the aerial photointerpretation as an attribute in the database
- Estimating the mixing ratio of classes during the field survey and capturing this information as an attribute in the database
- Estimating the size of unmapped inclusions within polygons
- Allowing multiple class memberships for individual polygons

Information derived from the first two methods is boundary driven and may be useful to those users interested in determining the width of transition zones. The latter three methods are polygon driven and are an indication of the variation to be expected within a single class.

8. Errors will be analyzed as follows:

   - Thematic errors will be reported in the form of a contingency matrix. Overall accuracy will be reported as "percent classified correctly" and adjusted for chance agreement with a Kappa index. Users' and producers' accuracy will be computed for each class with 90% confidence limits. A t-hypothesis test will be used to determine whether estimated accuracies conform to an 80% per-class accuracy.
   - Positional errors will be tabulated for each point. Positional accuracy will be computed as an RMSE in the x- and y-coordinate directions, and as a circular error with a 90% confidence level. A c^2 hypothesis test will be used to determine whether the estimated accuracy conforms to NMAS.

9. It is recommended that procedures be tested operationally during the prototyping phase of

the project. The objective of the operational testing will be to determine if the specified accuracy requirements can be met for the given classification scheme, and to determine if the proposed method is practical to implement in terms of time and cost. Changes or refinements of the procedure based on the results of the testing may be necessary.

2.0 Theoretical Aspects of Error Measurement in Vegetation Mapping

2.1 Modeling Natural Properties

2.1.1 Homogeneity of Classes

At the most fundamental level, geographical data can be defined as a collection of facts about places. The atomic unit of spatial data is the tuple <x,y,z>, where x and y define a place, and z is some fact about the place, such as vegetation cover class. A convenient way to define the difference between spatial and geographical data is to insist that for geographical data, the coordinates (x,y) be defined by some system of measurement on the Earth, latitude/longitude, for example, or in the UTM (Universal Transverse Mercator) coordinate system.

For many types of geographical facts, it is appropriate to conceive of z as a measurement at a point, or infinitely small area. For example, ground elevation can be measured at a point, as can air temperature, atmospheric pressure, or rainfall. Although a rain gauge will average rainfall over a finite radius, that radius will be much larger than the positional error in the location of the rain gauge, and will be irrelevant in any likely analysis.

For other types of geographical facts, the definition of the value at a point requires an observation over a significant distance around the point. Vegetation cover classes, for example, can only be defined over discrete areas that substantially exceed the area covered by any single plant. The spatial resolution of the vegetation cover class should not be confused with the positional accuracy of the point, or with measures of the size of the study area. In a typical example, the positional accuracy of the point might be 30 meters using GPS, while the spatial resolution of the vegetation cover class might be 100 meters, implying that it is necessary to observe an area of 100 meters on one side (1 hectare) to determine the cover class recorded at its central point.

When vegetation cover class A is recorded at a point, there is no implication that vegetation is uniformly of class A everywhere within an observational area around the point. An observational area of 1 hectare classified as Ponderosa pine may include many species besides Ponderosa pine. The recorded observation merely asserts that Ponderosa pine is sufficiently abundant relative to other species to satisfy the formal classification requirements of the mapping exercise. It follows

that if this form of mapping is to be successful, and if its accuracy is to be assessed on any objective basis, the rules that allow a heterogeneous mix of species to be represented as a single class must be laid down as objectively as possible, and have results that are reasonably replicable from one observer to another.

In addition to the problems associated with assigning a class to a point, vegetation mapping requires that points of like class be aggregated into polygons. Here again there is no implication that an observer would assign the same class to every point within the polygon. Instead, it is generally accepted that the polygon will include areas that are not of the same class, and that such areas are generally more likely to be close to the polygon's mapped edges. Moreover, the inclusion of any class that is smaller in area than the specified minimum mapping unit (MMU) will be ignored. The area of the minimum mapping unit will be larger than the spatial resolution of the data, which was defined earlier as the minimum area that must be observed for a point to be assigned to a class.

2.1.2 Distribution of Classes

The rules of classification used in a vegetation mapping exercise must be sufficiently precise to allow an observer to determine class at any point by observing some defined area around the point. The classification rules must result in the assignment of one and only one class to each point, no matter what the proportions of the species present, and it must be possible to apply the rules to both the mapping and to the subsequent accuracy assessment. In theory, a small number of classes will tend to produce a small number of large polygons, and a large number of classes will tend to produce a large number of small polygons. In practice, the polygons within any one class tend to vary greatly in size. The variation tends to exhibit positive skew, with a relatively small number of very large polygons and a large number of small ones; the mean polygon area is generally greater than the median area. Between classes, it is generally observed that some classes tend to be characterized by small polygons, and some by relatively large polygons.

The number and size of the polygons in each class tend to vary widely, and the accuracy assessment must be conducted accordingly. If an equal number of samples is taken 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 class area, or to the number of polygons in the class, then since a small number of classes are likely to account for

a large proportion of the total area, some rare classes may be missed altogether. In practice, a compromise is often used in which common classes receive more samples than uncommon classes, but not proportionately more.

For example, suppose a mapping exercise has three classes, occupying 70%, 27%, and 3% of the area respectively. The budget for accuracy assessment is sufficient to allow for 300 sample point observations to be made on the ground. To permit equally confident statements to be made about the accuracy of each class, 100_samples would be allocated to each class. On the other hand, to permit equally confident statements to be made about accuracy anywhere on the map, 210, 81, and 9 samples would be allocated to the three classes respectively, leading to relatively unreliable conclusions about the mapping of the third, rare class. A compromise might be to allocate 150, 100, and 50 samples respectively.

2.1.3 Boundary Definition in Transition Zones

Some vegetation class boundaries are well demarcated on the ground or from the air, and can be mapped precisely and with confidence. This is true in areas where vegetation is harvested and replanted, subjected to intensive silviculture, or is affected by sharp topographic features like ridges or tree lines. In other cases boundaries are much less clearly defined, as when, for example, transitions occur in ecotones, or the mixing of species or classes is particularly heterogeneous. In some cases the transition of an ecotone may be mapped as several classes, particularly when the vegetation classification scheme includes specific mixtures. In other cases a mapper may simply "eyeball" a medial line.

In a digital environment it is comparatively easy to ask the mapper to capture a measure of boundary uncertainty, or transition zone width, as standard practice, and to insert it into the database as an attribute of each boundary arc. In practice, however, most mapping still reflects the conventions of the predigital era, and boundary arcs are rarely distinguished in this way. It would be useful if the NPS Vegetation Mapping Project could experiment with this concept, and conduct trials to evaluate its costs and potential benefits. The inclusion of such data in the vegetation database might permit the development of (a) improved mapping techniques that would display cartographic representations of transition zones, (b) confidence limits on the areas of classes bounded by transition zones, and (c) improved methods of overlay for vegetation and other data.

2.1.4 Relationship between Smallest Mapping Unit and Smallest Thematic Unit

As noted earlier, a minimum mapping unit (MMU), defined as an area measure, is used as a criterion for vegetation mapping; any polygon smaller than this area (defined as an inclusion) is dissolved and merged with its surrounding polygon. When an inclusion has more than one neighboring class, rules are required to determine how it should be merged, and such rules should be made explicit. In principle, an inclusion should be merged with the neighbor of the most similar class. During the accuracy assessment, the same rules will have to be used to resolve ambiguities that may occur due to location of a sample site within an inclusion.

The MMU should also be defined explicitly. It may be desirable for the MMU to vary by class. For example, the MMU for a rare class might be smaller than one for a common class. In addition, it may be difficult to define the MMU for long, thin polygons like riparian zones, where it may be more appropriate to define a minimum width than a minimum area. However, in no case may the MMU be smaller than the minimum area that must be observed in order to establish class (i.e., smaller than the observational unit or spatial resolution discussed earlier).

The refinement of the classification scheme leads to smaller polygons, as noted earlier. It may also lead to a reduction in the size of the observational unit, or improve spatial resolution, since it is likely that a more refined classification will be determined through the observation of smaller minimum areas.

2.1.5 Spatial Patterning of Thematic Units

Vegetation maps tend to exhibit certain characteristic patterns. By definition, polygons must not overlap, and they must exhaust the space they cover, since there must be exactly one class at any point. The boundary network formed by polygon boundaries generally contains only junctions of three arcs; junctions of four or more are exceedingly rare. The smallest polygons, those close to the MMU, tend to be circular, whereas larger polygons, particularly those representing riparian zones, are frequently elongated and aligned with surface streams.

For accuracy assessment purposes, it may be desirable to avoid the edges of polygons, which are likely to be transition zones. In addition, the accuracy with which an observer is able to position a sample point is limited by positional accuracy. Finally, it is desirable to ensure that the observational unit lies completely within the sampled polygon. All three of these arguments

suggest that accuracy assessment should be conducted away from polygon boundaries, but give different estimates of the appropriate minimum distance.

Finally, some polygons may be so small that it is possible to observe the entire polygon, in which case the issue of avoiding boundaries is moot. The entire polygon will be observed if the polygon is close to the MMU, particularly if the MMU is smaller for rare classes, and for long thin riparian zones.

2.1.6 Thematic Implications of Spatial Errors

An error occurs in a vegetation map when the class mapped at some location <x,y> is not the same as the class observed at that location in the field. Of course it is necessary to assume that the processes of mapping and accuracy assessment (i.e., the application of the classification system) are identical, so that a false error is not detected because of some procedural difference.

Two types of errors in spatial data are often distinguished: thematic errors, that is, errors in the class assigned to a point, and positional errors, or errors in the position of the point. In vegetation mapping this distinction is often moot, since a disagreement over the class at <x,y> could be interpreted as being due to either error in the point's position, or error in the class at the point. Where a vegetation boundary is sufficiently well defined, such as where the boundary follows a road, it may be possible to make an independent estimate of positional accuracy and thematic accuracy. Such situations tend to be rare, however. For purposes of this discussion, positional accuracy is interpreted in the narrower sense as being a consequence of the accuracy with which a given position can be located in the field, and is therefore independent of the accuracy of vegetation mapping.

2.1.7 Cartographic Abstraction

The model of vegetation mapping being discussed here includes several spatial parameters: the minimum mapping unit area, the observational unit, and the estimated width of each boundary arc. In addition, the cartographic process imposes smoothness constraints on boundaries that are determined by map scale.

Cartographic abstraction refers to the degree of generalization inherent in a vegetation map, and to various processes by which it can be changed. For example, if the classification scheme is

generalized, a set of rules can be invoked to determine how polygons are merged with neighbors, and these rules are likely to be identical to those used to determine merging for polygons smaller than the MMU. If the scale of the map is reduced, it may be appropriate to further smooth polygon boundaries, and also to generalize the classification scheme. If the MMU is enlarged, rules must be invoked to determine which arcs to dissolve and which polygons to merge.

2.1.8 Effects of Scale

In addition to the relationships between scale, thematic classification, and boundary smoothness already discussed, the intended mapping scale will have an effect on the MMU and observational unit. The smallest object normally included on a map is about 0.5 millimeter across at map scale, since smaller objects are difficult to draw and are likely to be seriously mispositioned. It follows that the MMU should not be less than the area of a square 0.5 millimeter on a side at map scale, and that the observational unit should not be greater than this amount. With an intended map scale of 1:24,000, this implies that the observational unit should not be less than 12 meters on a side.

2.2 Error Detection

2.2.1 Definitions of Error and Accuracy

Vegetation maps are used for a variety of purposes, only some of which may be known at the time the maps are created. Nevertheless, it is important for the operational definitions of error and accuracy to match the expectations of the user community. Therefore, an error is defined here as a difference between the vegetation class specified at point <x,y> by the database, and the class observed at that location in the field. Accuracy is a measure of the absence of such errors (the more frequent the errors, the less the degree of accuracy).

Unless clearly advised otherwise, the average user will expect a definition of accuracy to apply to the entire database. For that reason, if the producer of the data believes that the accuracy of the data is significantly lower near polygon boundaries (or elsewhere), that information should be conveyed clearly to the user, who will otherwise assume that the published accuracy statistic applies everywhere.

Therefore, an accuracy assessment procedure in vegetation mapping is concerned by necessity

with the detection and measurement of both thematic and positional errors. A thematic error results if field observation determines that the class at some selected sample point <x,y> is not the same as the database class at that point. A positional error results if the location of a well-defined point (x,y) in the database is not the same as its coordinate location in the field. An accuracy assessment procedure can be performed by comparing a number of sample sites in the spatial data to the equivalent locations in an appropriate source of higher accuracy. The results of this comparison are then evaluated statistically in order to derive thematic and positional accuracy statements that are applicable to the data as a whole.

2.2.2 Considerations that Affect the Selection of a Source of Higher Accuracy

The choice of an appropriate source of higher accuracy depends on such factors as data availability, the scale of the data to be assessed, and the compilation methods used in the original data. In general, a source of higher accuracy should be independent and should be sufficiently more accurate for its errors to be negligible compared to the errors that are to be measured. In addition, data content and representation in both data sources should be governed by the same decision rules. For example, if a higher accuracy map were to be used for the assessment of thematic accuracy of a vegetation classification, then the source of higher accuracy should have been classified using the same classification method.

2.3 Sources of Error

2.3.1 Sources of Thematic Error

Thematic error can arise for a number of reasons:

- Limited ability to locate <x,y> in the field, because of limited positional accuracy. This is most likely when <x,y> is close to a polygon boundary, in a transition zone, or close to an unmapped inclusion.

- Misinterpretation of the class at <x,y> by the mapper. This is most likely when the observed and true classes are easily confused from the air.

- Location of <x,y> in a transition zone, or in an unmapped inclusion (an inclusion smaller than the MMU).

- Incorrect application of classification rules by the mapper. It is assumed that the accuracy assessment always applies the classification rules correctly, since an incorrect application by the assessor can result in false errors due to procedural differences.

- Errors due to distortions introduced by data processing, such as digitizing error and map registration error.

- Inability to sample randomly. Sampling theory assumes that all locations have an equal probability of being sampled (Stoms et al. 1994). Although samples may initially be selected randomly, many areas in national parks are remote and not easily accessible to sampling. Even areas that are accessible may be reachable on foot only, making field checks time consuming and expensive. Therefore, some samples may be replaced with more accessible sites. If so, some degree of nonrandomness in the final sample will be unavoidable, and limiting sampling to the more accessible areas may bias the sample, since the distribution of vegetation classes is not random; accessible areas tend to be more heavily influenced by ecotonal variation.

- Discrepancies in classification due to temporal effects, season, and seasonal variability. The sampling of vegetation communities is complicated by temporal change, season, and seasonal variability. Each of these effects can alter the appearance of the vegetation more rapidly than the time frame within which the mapping of the vegetation is likely to occur. This will impact the results and the interpretation of the accuracy assessment, because the ground reference against which the data are compared is itself subject to variation. While a discrepancy in vegetation class assignment due to temporal effects is an error within the context of this study, it is not an error due to misclassification on the part of the interpreter, but is simply indicative of the fact that at this point, the database needs to be updated. Such discrepancies may occur for the following reasons:

    - Accuracy assessment sampling is conducted in a different season than the original data collection. For example, original data were collected in spring (early growing season), and the accuracy assessment was conducted in the fall.

    - Accuracy assessment sampling is conducted in the appropriate season, but the current season is abnormal (e.g., an unusually cool spring delays the start of the growing season).

- Some temporal change has occurred that has altered the vegetation cover in the area. For example, a fire has destroyed some old-growth forest after the original data were collected, and the area is now covered by a fire-following community.

2.3.2 Sources of Positional Error

As noted earlier, positional error as a practical matter only has relevance for well-defined features, where it is possible to compare locations. In general, l, polygon boundaries are not well defined, except where they are known to follow other features. However, positional accuracy is an important measure in determining the ability of accuracy assessors to locate sample points in the field, and is determined by the method used to locate such points. Since points must be located respect to the mapped vegetation polygons, it is necessary to know the accuracy of both the sample point location and the polygon boundary location. In view of the absence of well-defined points in the latter case, it is assumed that the positional accuracy of polygon boundary locations is roughly the same as that of the points. For a map of scale 1:24,000 that meets National Map Accuracy Standards (NMAS), this implies a positional accuracy of roughly 0.5 millimeter at map scale, or 12 meters on the ground, assuming a Class 1 product.

Well-defined points may be located in the field by using several methods:

- Differential GPS (DGPS): If a base station accurately positioned with respect to the geodetic control network can be used, DGPS will locate a point within approximately 1 meter of its true position with respect to the Earth frame.

- Absolute GPS: Without geodetic control, 90% of GPS locations will fall within 100 meters of their true location.

- Location by mapped topographic feature or air photo: The ability to locate accurate positions from a topographic map or air photo is at best equal to the map accuracy standard, or about 0.5 millimeter at map scale. In areas lacking in well-defined features, positional accuracy can deteriorate dramatically, and GPS is recommended.

- Location by orthophoto quad or remote sensing image: It seems likely that increasing use will be made of digital orthophoto quads for field positioning. These products have a spatial resolution of 1 meter, but their positional control is roughly 6 meters. Again,

this is a best accuracy for well-defined features, and actual accuracy may be much worse. Remotely sensed images can be registered to have a positional accuracy of less than 1 pixel, but in the absence of well-defined features positional accuracy may be much worse.

In addition to the issues connected with field positioning, positional accuracy tends to be corrupted significantly by the various stages in the creation of a digital database. Except for major blunders, however, these are unlikely to contribute more than the 0.5 millimeter assumed to be the approximate positional error of map products.

3.0 Practical Aspects of Error Measurement in Vegetation Mapping

The successful implementation of an accuracy assessment procedure depends to a large extent on the quality of the higher accuracy source, a clear understanding of the operational constraints that may be encountered in the field, and a statistically valid sampling design.

3.1 Quality of Sources of Higher Accuracy

Potential sources of higher accuracy that might be used include the following:

- Ground checking. The spatial data are compared to ground truth through field checking. This is the recommended approach for the NPS/USGS Vegetation Mapping Project. Actual field checking is recommended because of the large scale of the data, and because the original mapping relied at least partially on field sampling methods.

- Sources that can substitute for ground checking (especially in inaccessible areas), such as supplementary photography and airborne video (Goodchild et al. 1994). Multispectral airborne video data are usually acquired by flying transects. This permits relatively large areas to be observed in a short period of time. This is a promising (but experimental) approach, especially for the assessment of small-scale spatial data, such as may be derived from satellite images. For further discussion, see Section 8.4.2.

- Maps or other sources of spatial data that are of higher accuracy. These would themselves be abstract representations of the data to be assessed, but would be of sufficiently higher accuracy for error to be considered negligible by comparison with the error in the data to be assessed. These sources are considered inadequate for the assessment of thematic accuracy, because the dynamic nature of vegetation cover will probably render most available higher accuracy spatial data as out of date. This means that vegetation cover will in all likelihood be different between the spatial data and the source of higher accuracy, simply because of temporal variation. Furthermore, even if no temporal change has occurred, it would be nearly impossible to find a source of

higher accuracy that has been compiled according to the same compilation rules as the data to be tested. However, these data can be used for the assessment of positional accuracy, because the well-defined points that are needed for this type of assessment tend to be much more stable over time.

- Existing vegetation plot data could be reserved for the purpose of accuracy assessment if the site location is adequately recorded and if the information about the vegetation in the plot is sufficient to support the target classification system. The age of the plot data must also be evaluated to determine if it has relevance to current conditions. This type of data will have to be evaluated on a case-by-case basis but could result in considerable cost savings.

3.2 Operational Constraints

3.2.1 Dealing with Accessibility Problems that Affect Randomness of Sampling

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 randomness 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. The alternative site should be in the same mapping polygon. If no such alternative site is possible, the sample site should be dropped. A second visit into the field with a completely new alternative site would only be made if the sample size after the first visit is below the minimum sample size required for the analysis.

Since sample site replacement will, to a considerable degree, be dependent on the judgment of the investigator, it will be 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.

If site accessibility for ground-based sampling is especially limiting, consideration must be given to other sources of higher accuracy, such as airborne videography or observation from a hovering aircraft. Remote or secondary types of sampling will, of course, lower the confidence in the accuracy of the site evaluation, but this may be an acceptable tradeoff to obtain more and better distributed sampling sites. Testing of these "remote procedures" should be undertaken at the prototype stage to assess their impact on the confidence in the accuracy assessment.

3.2.2 Dealing with Temporal Effects, Season, and Seasonal Variability

The recommended approach for dealing with the effects of season, seasonal variability, and temporal change is to conduct field visits where these conditions are as similar as possible to the conditions when the original map was compiled. Since this may not often be feasible, it is recommended to address variations due to these factors through documentation. Sample sites affected by drastic temporal change should be dropped from the assessment, but site conditions and changes in vegetation class should be documented.

Dealing with season, seasonal variability, and temporal change is discussed in more detail below.

- Season. Sampling for accuracy assessment should be conducted in the same season as the sampling for the vegetation classification. From an operational point of view this implies that in many situations there will be at least a one-year delay between original data collection and the accuracy assessment.

- Seasonal variability. Investigators should be given some liberty in choosing the date of the accuracy assessment, because seasons are subject to considerable variation. Climatic conditions should correspond closely to those present in the original sampling or have minimal effect on the appearance of the vegetation. This emphasizes the need for conducting the original vegetation classification as well as the accuracy assessment during a period of relative stability in the appearance and composition of a vegetation community. It also emphasizes the need to carefully document climatic and environmental conditions during the original classification survey and the accuracy assessment.

- Temporal change. Samples 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 remapping, since this type of change will probably affect the position of the boundaries of adjacent vegetation communities.

4.0 Sample Collection for Accuracy Assessment

4.1 Sampling Design

4.1.1 Sampling Design and Data Collection Objectives

The sampling design implemented for accuracy assessment of the NPS/USGS Vegetation Mapping Project should adhere to the scientific principles that govern sampling and statistical analysis and also be practical. Specifically, the assessment methodology should satisfy the following objectives (Stoms et al. 1994):

1. It should be scientifically sound. As such, the method should be repeatable, and the sampling design should permit the adequate representation of the population about which statistical inferences are to be drawn.

2. The methods used should be applicable to all areas that are part of the project. Although there may be some regional variation in the implementation of the accuracy assessment, these variations should still be based on the same theoretical foundation, such that the results of separate assessments are comparable.

3. The method should be economically feasible in view of both time and cost constraints.

In addition, regardless of the sampling design employed, the actual method of sample collection in the field used for accuracy assessment, and that used in the field survey to compile the original data should ideally be the same. In other words, if relevé sampling is utilized to determine the initial vegetation class, relevé sampling should also be used during the accuracy assessment. If it is not practical for the field and accuracy sampling techniques to be identical, then the potential sources of error that may be introduced should be documented as part of the assessment report. It is recognized that for many sampling sites, simple observation will be sufficient to verify class membership.

4.1.2 Sampling Design for Purposes of Classification System Development

The purpose of collecting a sample is to analyze it in order to make inferences about the population the sample represents. If samples are collected in order to initially classify vegetation,

the population about which inferences are to be drawn is the vegetation cover itself. In other words, the sampling design should permit the investigator to capture the variation in vegetation as it exists in the real world and categorize it according to some defined classification scheme. In order to sample for vegetation mapping, it will be important to place sample sites in a manner that permits the investigator to determine where changes occur in vegetation cover. In the NPS/USGS Vegetation Mapping Project, the initial delineation of boundaries will be made by inspecting aerial photos; the delineation will then be refined through field surveys. The emphasis will be on locating samples in a manner that encompasses the diversity of vegetation that is to be mapped, with some emphasis on determining the location of transition zones, since these will be represented as "hard" polygon boundaries on the map. There will be comparatively less emphasis on randomization of sample locations, and more emphasis on covering as large an area as possible during the field survey. Indirectly, the number of samples to be used will be driven by the variability of the vegetation that exists in the area.

4.1.3 Sampling Design for Purposes of Accuracy Assessment

The objectives of collecting samples for the accuracy assessment are somewhat different than those for classification. Here, the sampling is aimed at drawing inferences about the nature and magnitude of discrepancies between the true properties of a point or area and its representation on the map. As such, the purpose of collecting samples to assess the accuracy of data generated as part of the NPS/USGS Vegetation Mapping Project is twofold:

1. To obtain a measure of the probability with which a particular location has been assigned its correct vegetation class (thematic accuracy).

2. To determine how well a particular location is positioned compared to its true position in the field (positional accuracy).

With these types of objectives, the randomness of the sample sites should be emphasized, and the number of samples sites will be heavily influenced by statistical constraints. In other words, the main objective will be to collect a sufficient number of samples to permit statistical inferences to be drawn about the data as a whole.

4.2 Sampling Methods

There are a number of methods for drawing samples from a population. The most common are random, stratified random, systematic, and two-stage (also cluster or nested) sampling. Other approaches exist, but they are essentially variations on the four methods listed here.

The properties of the most common sampling schemes are summarized below (Taylor 1977).

4.2.1 Simple random sampling

Sample sites are selected randomly by choosing random coordinate locations as sample site locations. In terms of initial site selection, each sample location has an equal probability of being selected; sample site selection is independent in that the selection of one site has no bearing on the selection of another. However, as noted earlier, final sample sites used may not be completely random, but may have been selected on the basis of accessibility. Random sampling is probably the most well-known form of sampling. Its property of independence of selection is necessary to the scientific validity of many of the statistical procedures applicable to accuracy assessment.

Despite its validity for purposes of statistical analysis, random sampling has disadvantages for classification accuracy assessment in spatial data:

- Abundant classes are over-represented in the sample. Since an abundant class covers more area, abundant classes have a higher probability of containing sample sites. This will have implications regardless of whether an overall or per-class classification accuracy is required. If an overall classification accuracy statement is required, the accuracy statement will disproportionately represent the error properties of the abundant class(es). If a per-class accuracy assessment is required, simple random sampling will locate a small number of sample sites in the rare classes, reducing confidence in the accuracy of the estimate of the rare class.

- The samples selected may occasionally be overweighted in terms of extreme values, since all values have an equal probability of being selected. Therefore, simple random samples tend to be imprecise.

4.2.2 Stratified Random Sampling

Stratified random sampling is a method for overcoming the shortcomings of random sampling, while retaining the advantages of randomness in terms of the statistical validity of the results of the method. With this approach, the population is stratified into smaller subpopulations from which random samples are drawn. For an accuracy assessment of vegetation classification, the obvious stratification is by vegetation class. An appropriate random sample can then be selected for each vegetation class, regardless of its abundance, and the error properties of each class can be modeled separately from one another. Since stratified samples are more homogeneous than simple random samples, they tend to be more precise.

4.2.3 Systematic Sampling

In systematic sampling, all samples are selected at some type of regular interval. Operationally speaking, this is probably the simplest and least expensive approach to sampling. However, site selections are not independent. Once the initial site has been chosen, the location of all others is predetermined. In addition to being invalid for many statistical purposes, systematic sampling schemes, even if the number of samples is large, do not always result in samples that reflect the population about which inferences are to be drawn, especially if the population is arranged in some regular pattern.

4.2.4 Two-Stage Sampling

In two-stage sampling (also cluster or nested sampling), the population is divided into subpopulations, each of which covers the entire range of expected population properties. In the case of accuracy assessment, each subpopulation would have to contain the entire variation of errors that are to be expected in the entire population. Two-stage sampling is usually inexpensive and can be carried out rapidly. However, if the subpopulations reflect population properties poorly, inferences drawn from such a sample will be unreliable.

4.2.5 Recommended Sampling Methods

For thematic accuracy, since the NPS/USGS Vegetation Mapping Project requires a per-class accuracy statement, the recommended approach is to utilize stratified random sampling, where the strata are vegetation classes. This approach is also recommended because it is by far the most

widely used (and tested) method of sample collection for the thematic accuracy assessment of vegetation classes (Manual of Remote Sensing 1983). In addition, the requirement for class-specific accuracy implies a stratified random sampling approach. For positional accuracy, both simple random sampling and stratified random sampling may be acceptable. If stratified random sampling is used, it is recommended that sample locations be stratified by area (quadrant), if possible, to ensure that the accuracy estimate is valid over the entire area assessed.

Despite the suitability of stratified random sampling for vegetation classification accuracy assessment, the method has logistical constraints. In theory, all locations in a specific vegetation class should have an equal probability of being sampled. However, as noted in Section 2.3.1, some sample sites will almost certainly be so difficult to access that their use as sample sites becomes prohibitively expensive unless checked by a secondary source of higher accuracy. Since field checking is the recommended method for the accuracy assessment, these sites will almost certainly be lost from the sample and replaced by other, more accessible sites.

Although stratified random sampling permits class-specific accuracies to be developed, the derivation of an accuracy for each individual class may be prohibitively expensive, especially if the class is rare, because a statistically adequate number of samples must be collected for each class, regardless of its abundance. For that reason, a much larger number of samples must be collected than would be necessary for simple random sampling. To overcome some of these constraints, it is recommended that the importance of rare classes be de-emphasized during sampling or by pooling classes known to have similar error properties.

4.3 Sample Site Selection

Random site selection has a somewhat different meaning for the assessment of positional and thematic accuracy. Determining positional accuracy requires selecting well-defined points that can be located in the spatial data as well as in the source of higher accuracy. Selecting a random sample of points for positional accuracy assessment therefore implies choosing a random sample among all well-defined points in the spatial data. The selection of sample sites for the assessment of thematic accuracy is not restricted by the same constraint: any coordinate location may be chosen as a sample site as long as the coordinate can be visited in the field.

One special situation that needs to be addressed is the use of photointerpreter training sites as sample sites for thematic accuracy assessment. Whenever a vegetation classification relies on aerial photographs or other remote sensing data, interpreters select training sites in the imagery. A training site can be defined as a representative sample site of a specific (vegetation) cover type (Lillesand and Kiefer 1979). Interpreters then assign areas with the same characteristics as the training site the class represented by the training site. (The validity of the classification of the training site is often verified in the field before photointerpretation begins.) The use of training sites as sample sites for thematic accuracy assessment can potentially introduce bias, if the majority of the accuracy assessment sites are also training sites. Thematic accuracy assessments based on training sites will probably indicate too high an accuracy, since training sites are almost certain to be classified correctly. For these reasons, the use of training sites as thematic accuracy assessment sample sites should be discouraged. However, if the selection of sample sites for thematic accuracy assessment is independent of the selection of sample sites for the mapping process, as is recommended for the NPS/USGS Vegetation Mapping Project, the probability that a training site will be selected as an accuracy assessment site is small. If one is selected as part of the sample, it is unlikely to have undue influence on the accuracy estimate, especially since the aerial extent of the training site is unlikely to correspond exactly to the extent of the sample site.

The distinct requirements for the selection of sites for thematic and positional accuracy assessments imply that a separate set of sample sites must be selected for each type of assessment. Selection criteria in terms of positional and thematic accuracy assessment are further explained in Sections 4.3.2 and 4.3.3 below.

4.3.1 Sample Site Location in the Field

An important operational issue is the type of method the field checkers will use to identify the location of sample sites. The type of method to be used depends on cost and on the accuracy with which sample sites need to be positioned. In general, the more accurate methods will also be the more expensive ones. The accuracy requirements for the NPS/USGS Vegetation Mapping Project (addressed in Section 7.0) almost certainly require one of the higher-precision GPS solutions, particularly for the determination of positional accuracy. The accuracy with which sample sites must be identified for thematic accuracy assessment may be more relaxed.

However, since sample site location requires navigating to preselected sites, users must have access to a method that permits them to correct their position in the field. The issues associated with the use of GPS for sample site selection are further discussed in Section 7.1.

The method recommended for the NPS/USGS Vegetation Mapping Project are real-time DGPS (which permits sample sites to be located within 2 to 3 meters), and DGPS (which is accurate to 100 meters without post-processing, 1 meter with post-processing). If real-time DGPS is not possible (because of communications problems or for other reasons), field investigators will need to use interpretive judgment and collateral data to locate the sample site as closely as possible in the field. This implies that coordinate locations different from those originally selected are permissible, as long as they are within the error radius of uncorrected DGPS data.

4.3.2 Sampling for Thematic Accuracy Assessment

Three issues must be considered when sampling vegetation for classification accuracy assessment:

1. How to deal with ecotonal variation and transition zones

2. The size of the sampling unit to use once the sample site is selected

3. The treatment of classes based on their abundance

First, as noted in Section 2.1.5, vegetation data will be classified into nonoverlapping polygons with smooth boundaries that imply a fixed line along which there will be a distinct change from one vegetation class to another. In reality, these delineations are actually transition zones (ecotones) of varying (but real) width, and the change from one vegetation class to the next will be anything but clear.

Regardless of the sampling scheme employed, a given sample location may be located in a transition zone, although its classification in the spatial data implies a clear class membership. Depending on the positional accuracy of the data and how well the sample location can actually be identified in the field, there may be further difficulties in comparing a field sample site to its representation in the spatial data for purposes of accuracy assessment.

There are two ways to deal with ecotonal zones in a sampling design (both encourage sampling away from polygon boundaries):

- Sample away from ecotonal variation by using an arbitrary definition of the width of the transition zone. In other words, select a stratified random sample of points, but restrict the selection of sample points to areas away from the polygon boundary by a fixed distance, deemed to be the "transition zone." This will bias the sample, since class membership away from the transition zone is more pronounced, thereby increasing the chance of agreement between the class assignment in the database and the class assignment made during the accuracy assessment.

- Select sample locations regardless of ecotonal variation. However, place sample locations sufficiently far from polygon boundaries to account for positional errors that are either inherent in the data or are due to the inability to identify the sample site in the field. At a minimum, select sample sites sufficiently far from the polygon boundary to completely contain the sampling unit. This approach is less arbitrary than the first, since it uses a decision rule based on a real source of error in the data.

The second method is preferred, since the results derived are more indicative of the real ambiguity that exists when class membership is assigned to a particular vegetation polygon.

The second issue to consider is the size and shape of the area to be sampled, once the sample location has been determined. The MMU is currently defined as an area of 0.5 hectare, and the basic recommendation is to sample a circular area equivalent to the MMU. However, the best approach will probably be to give field investigators considerable freedom in defining the size and shape of a particular sampling area once they reach the sampling site, although guidelines can be provided for both variables. This flexibility is necessary because both the size and the shape of the sample area will depend to some extent on the properties of the vegetation class. Therefore, the recommended approach is to give the investigator a summary of recommended procedures, but to permit him or her to make some decisions based on field conditions as to the following:

- Sampling area. Ultimately, MMUs may be class specific. If this is the case, the sampling unit should be equivalent to the MMU for that class. Sampling an area of at least the MMU will control to some extent for unwanted effects, such as unmapped

inclusions.

- Sampling area shape. A permitted modification of the recommendation may be to define the sample site as linear and to use a linear rather than an aerial measure (e.g., sample an area at least 40 meters in diameter in one direction), especially for communities expected to occur mainly as elongated polygons (e.g., riparian zones).

- Although a circular MMU is considered optimal, it is recognized that this may be impractical in the field, especially in wooded areas. Investigators will, therefore, be permitted to use an alternate shape where necessary, provided the choice of shape is properly documented.

The third issue considers the sample sizes for each vegetation class or strata. Ideally, all classes would be sampled equally. However, depending on the characteristics and abundance of classes in a particular area and the resources available for sampling, it may be necessary to allocate fewer sample points to rare classes or those more easily identified and mapped. Further discussion is provided in Section 4.4.2.

4.3.3 Sampling for Positional Accuracy Assessment

Positional accuracy assessment depends on the availability of well-defined points that can be established in the field as well as in the data to be assessed. Examples of such well-defined points are road intersections, benchmarks, or the corners of buildings (National Park Service 1993). For the results of the positional accuracy tests to be representative of the entire product area, sample sites should be distributed throughout the area of the product being tested. This recommendation is in keeping with the recommendation of the ASPRS standard (Merchant 1987), which is for test points to be selected in all quadrants of the product.

In practice, this may not always be implementable, since well-defined features tend to be sparse in national parks. Even when available, these points tend to be clustered. The recommended approach is to locate a sufficient number of points in each quadrant of the area to be assessed, whenever possible. Candidate points may well be selected from sources such as USGS 7 1/2' quads, aerial photographs, or orthophotos. It is recommended to use natural features where possible, as opposed to photogrammetric targets; using photogrammetric targets for positional accuracy assessment would be similar to using training sites for thematic accuracy assessment.

If the distribution of test points does not represent the entire area, an accuracy estimate should be computed in the usual manner, but should be reported with a statement regarding the aerial extent for which the estimate is valid.

4.4 Determining Sample Size

4.4.1 Determining Acceptable Levels of Error and Confidence

The appropriate number of sample sites to use for an accuracy assessment depends on the level of error that is permissible in the data, and on the level of confidence one wishes to have in the estimate. Other factors that may come into play are time and cost constraints in conducting the accuracy assessment. The question of sample size (i.e., sample number) has been addressed in many research papers. Since thematic accuracy assessment deals with statistical inferences regarding proportions, the sample size problem is usually resolved using the binomial distribution (e.g., van Genderen and Lock 1977, Aronoff 1985, Ginevan 1979, Goodchild et al. 1994), and this is the approach that will be taken here. The sample size required for a given confidence level and a given acceptable error in the sample can be calculated directly from the binomial distributions as follows (Goodchild et al. 1994):

$$n = \frac{pq}{\left(\frac{E}{z_\alpha}\right)^2}$$

where p is the required accuracy of the data (80% in the case of the NPS/USGS Vegetation Mapping Project), q is (1 - p) or 20%, E is the allowable error, and z_α is drawn from the normal curve for the given level of significance α. Note that this formula ignores the finite population corrective. This is satisfactory for most applications and is recommended here, since population size will, in most cases, not be known. If population size is known, it can be corrected as follows (Snedecor and Cochran 1967):

where n' is the adjusted sample size, n the original sample size, and N the size of the population.

Assuming an acceptable level of error of 0.1, a significance level of 0.10 (90% confidence level), and a required accuracy of 80%, the minimum number of samples is:

$$n = \frac{0.80)(0.20)}{\left(\frac{0.10}{1.3}\right)^2} = 27$$

The variation of sample size with level of confidence and level of acceptable error is illustrated in Table 1 (derived from Goodchild et al. 1994).

Table 1: Relationship between Sample Size, Confidence Intervals[1], and Acceptable Error

Level of acceptable error	α = 0.10 (90% confidence)	α = 0.05 (95% confidence)	α = 0.01 (99% confidence)
	Number of samples		
0.01	2,704	4,330	8,656
0.02	676	1,082	2,164
0.03	300	481	962
0.04	169	271	541
0.05	108	173	346
0.07	55	88	176
0.10	27	43	87
0.15	12	19	38
0.20	7	11	22
0.25	4	7	14

[1] Confidence intervals are defined in Section 5.4.

In general, required sample size decreases as the level of confidence required in the estimate decreases and the level of acceptable error in the sample increases. The highest number of samples is required in studies in which investigators need to be able to state the degree of similarity between two quantities (here the required and the estimated accuracy) with a high level of confidence.

What are acceptable levels of error for map accuracy assessment, and with what level of confidence should investigators be able to state it? Cost and time constraints and the nature of vegetation mapping greatly influence the answer to these questions. A per-class accuracy assessment implies an acceptable level of confidence and error for each individual class. If a given park contains 100 classes, a confidence level of 90% and an acceptable level of error of 10% for each class implies an overall sample size of 2,700 (27 sample points times 100 samples). This number is probably prohibitively large for an accuracy assessment based on field checking, even for a large park. At the same time, the per-class accuracy estimates derived are not very precise. However, the level of confidence in an overall accuracy assessment based on this sample size would be very high: approximately 99% with a sampling error of approximately 0.02.

Given that vegetation mapping is necessarily interpretive, it is recommended that relaxed requirements be used in terms of acceptable levels of error as well as confidence levels in the estimate. Otherwise, regardless how carefully the mapping process is carried out, it is unlikely that accuracy requirements will be met. Using a sample size that ensures a confidence level of 90% with an acceptable sample error of 10% therefore seems reasonable.

4.4.2 Sampling of Rare and Abundant Classes

Given the high cost of ground sampling, some decisions must be made if rare classes are to be sampled as frequently as abundant classes. It is anticipated that in most cases, a few abundant classes will cover most of the area in a park. The remainder of the park area will be occupied by a considerable number of rare classes. Therefore, if all classes are sampled to the same precision, most accuracy assessment costs will be incurred by rare classes. In reality, sampling all rare classes to the same level of precision as the abundant ones will probably far exceed the budget available for accuracy assessment. It is therefore recommended that some weighting factor be applied that allocates a larger number of samples to the abundant classes. Two approaches that may be taken toward the weighting are as follows:

1. Sample size can be made proportional to the abundance and frequency of the class. With this approach, the rarest classes would probably never be sampled. This is unacceptable,

since it is desirable to be able to associate at least a point accuracy estimate with each class.

2. Maximum and minimum sample sizes can be established, taking into account statistical as well as cost constraints and probable class abundance and frequency.

If the second approach is taken, it is recommended that 30 samples be specified as the maximum sample size for abundant classes, and that 5 samples be specified as the sample size for the rarest classes. Any class too rare for 5 sample sites to be selected should be observed in its entirety. Since a number of classes are intermediate in abundance between abundant and rare, five scenarios are based on class abundance and frequency have been defined:

Scenario A: The class is abundant. It covers more than 50 hectares of the total area and consists of at least 30 polygons. In this case, the recommended sample size is 30.

Scenario B: The class is relatively abundant. It covers more than 50 hectares of the total area but consists of fewer than 30 polygons. In this case, the recommended sample size is 20. The rationale for reducing the sample size for this type of class is that sample sites are more difficult to find because of the lower frequency of the class.

Scenario C: The class is relatively rare. It covers less than 50 hectares of the total area but consists of more than 30 polygons. In this case, the recommended sample size is 20. The rationale for reducing the sample size is that the class occupies a small area. At the same time, however, the class consists of a considerable number of distinct polygons that are possibly widely distributed. The number of samples therefore remains relatively high because of the high frequency of the class.

Scenario D: The class is rare. It has more than 5 but fewer than 30 polygons and covers less than 50 hectares of the area. In this case, the recommended number of samples is 5. The rationale for reducing the sample size is that the class consists of small polygons and the frequency of the polygons is low.

Specifying more than 5 sample sites will therefore probably result in multiple sample sites within the same (small) polygon. Collecting 5 sample sites will allow an accuracy estimate to be computed, although it will not be very precise.

Scenario E: The class is very rare. It has fewer than 5 polygons and occupies less than 50_hectares of the total area. In this case, it is recommended that the existence of the class be confirmed by a visit to each sample site. The rationale for the recommendation is that with fewer than 5 sample sites (assuming 1 site per polygon), no estimate of level of confidence can be established for the sample (the existence of the class can only be confirmed through field checking).

The recommendations above take into account both the statistical and operational aspects of sampling. Since the numerical values associated with each scenario above are somewhat arbitrary, the percentages and frequencies that should be associated with the definitions of "abundant" classes or "rare" classes should be tested in the field. It is quite possible that these percentages will have to be regionally adjusted.

If sample size is based on the abundance of a class, it follows that the accuracy estimate associated with rare classes cannot be stated with the same level of confidence as that associated with more abundant classes. For example, with a sample size of 5, the level of error in the estimate is closer to 25% at a 90% confidence level, as opposed to 10% with a sample size of 27. This will have implications for our ability to accept a given point estimate as meeting accuracy requirements. Whether or not a given accuracy estimate is accepted as meeting requirements depends on the width of the confidence interval associated with the point estimate and the outcome of a hypothesis test that determines if a given point estimate is equivalent to or exceeds requirements. The computation of the confidence intervals and statistics associated with hypothesis tests are discussed in Sections 5.3 and 5.4.

4.5 Data Collection Methods

Accuracy assessment for thematic data may be point or polygon based. With a point-based approach, the aim of the accuracy assessment is to verify the classification of a particular location on the Earth's surface. This assessment can be conducted quite independently from any

(arbitrary) boundaries the mapper may have used to delineate vegetation boundaries in the spatial data. Typically, this type of assessment requires the random selection of a number of coordinates. Field investigators then visit the site and observe a prespecified area around the site to determine its classification.

With a polygon-based approach, on the other hand, the aim of the accuracy assessment is to randomly select a number of polygons, each of which will then be sampled to determine whether the polygon as a whole has been correctly classified. In most cases, point-based approaches permit the observation of a "point area" in its entirety (i.e., without further subsampling to determine its identity). Polygon-based approaches, on the other hand, are for the most part too large to be observed in their entirety. To field-check the classification of an entire polygon therefore requires utilizing some sort of sampling technique, such as plot or transect sampling. (Note that point-based approaches may also require sampling; however, if the size of the point is chosen to correspond to some unit, such as the MMU for a given class, the need for sampling can be minimized.)

Plot sampling involves randomly placing a number of plots of fixed (manageable) size within a polygon. The size of the plots will vary according to the community type. Common guidelines are 25 m^2 for bushes and shrubs and 100 m^2 to 1600 m^2 for forest trees. Another approach is to choose the average stand size (relevé) as the plot size, where average stand size is defined as an area of sufficient size to contain a representative species distribution for a given community. Vegetation data are collected for each plot. The information about the plots is then summarized, and the original classification rules are used to make a decision about whether or not the polygon has been correctly classified.

Since plot sampling is very time consuming, an alternative method, transect sampling, can be used. With this approach, one or several transects are oriented within the polygon. A number of samples are then placed along the transect(s), usually utilizing a random starting point and a fixed sampling interval. Vegetation data are collected at each sample point along the transect (usually within a fixed width). As in plot sampling, the information collected for each point is then summarized, and the original classification rules are used to make a determination as to whether the polygon is correctly classified.

As has already been implied in Section 4.1.3, the recommendation for the NPS/USGS Vegetation Mapping Project is to choose a point-based approach.

A point-based approach as opposed to a polygon-based approach is preferable for accuracy assessment from both a user perspective and from the perspective of data collection strategy.

A point-based approach is preferable from a user perspective because although data representation may be in the form of polygons, many data uses involve raster data. (For the NPS/USGS Vegetation Mapping Project, an example would be to overlay the polygon data with satellite imagery to conduct some sort of change analysis.) If rasterized, a point-based accuracy assessment is inherently more meaningful to the user than one based on the original mapping polygons. Also, users tend to have implicit "point-based" views of the meaning of an accuracy estimate. As such, accuracy figures tend to be interpreted as the probability of encountering that class when visiting a particular spot on the Earth's surface, not when visiting a particular area.

A point-based approach is preferable from the perspective of data collection because it is difficult to draw inferences about polygons as a whole, based on plot or transect collection methods, without essentially sampling at the same resolution as for the original classification. In the case of plots, such a scheme would almost certainly be prohibitively expensive. The use of transects, although more cost-effective, is problematic, because it will be impossible to position the transect in a way that ensures its representing the polygon as a whole. This is so because different polygons have different levels of homogeneity even if they belong to the same class (Goodchild et al. 1994). Since a transect represents only a very small area of a polygon, it may not be a very good representation of the entire polygon. Therefore, a substantial level of doubt is inherent in assigning a class to a polygon on the basis of the results of transect sampling.

Assuming the use of the point-based approach, the minimum area of observation around the selected point should be at least equal to the minimum mapping unit (MMU) of 0.5 hectares. This assures that small inclusions of nonrepresentative vegetation do not unduly influence the classification of an area. This does assume that the MMU of 0.5 hectares is sufficiently large to contain a full sample stand, where a full sample stand meets the following requirements (Mueller-Dombois and Ellenberg 1974):

1. All species belonging to the plant community are present within the sampling area.

2. The habitat should be uniform within the stand.

3. The plant cover should be as homogeneous as possible.

It is possible that some classes of vegetation will require an MMU greater than 0.5 hectares to satisfy the above requirements. As a practical matter, this can be addressed in the identification keys that will need to be developed to guide both the field reconnaisance and accuracy assessment teams. See discussion in Section 5 of Field Sampling Methodology.

5.0 Computational Issues

5.1 Definitions of Precision and Accuracy

Precision and accuracy are related and, in a sense, complementary concepts. Precision can be defined as the degree of conformity among a set of observations to a random variable (Mikhail and Ackerman 1976). As such, precision is a measure of dispersion of the probability distribution associated with a measurement (repeatability of a measurement). Accuracy, on the other hand, can be defined as how close an estimate is to its true value. Ideally, estimates should be both precise and accurate. In other words, a point estimate (parameter) should be as close as possible to its true value. At the same time, the degree of variability in the point value, as derived from repeat measurements (sampling), should be small. However, it is possible to have highly precise estimates that are inaccurate, and highly accurate estimates that are imprecise. The difference between precision and accuracy lies in the possible presence of systematic effects. A precise but inaccurate estimate is considered to be affected by systematic error.

If an estimate is precise, it has a narrow confidence interval. Since the width of a confidence interval is affected by sample size, it follows that high-precision estimates require larger sample sizes than low-precision ones. Although the degree of precision in an estimate can, to some extent, be determined by sample number, the degree of accuracy in the estimate is determined by comparing the estimate to its expected (required) value through hypothesis testing. Whether or not the estimate is considered to be affected by systematic effects (i.e., is significantly different from the expected or required value), will again depend on the precision of the estimate. With a high-precision estimate, even small deviations from the required value will be considered significant. With a low-precision estimate, comparatively large deviations from the expected value will probably be interpreted simply as random effects.

In terms of accuracy assessment, both the degree of precision and the accuracy of the estimates derived from the sample will determine whether accuracy requirements can be considered to have been met. Because of the proposed number of samples, the precision of the per-class estimates for the NPS/USGS Vegetation Mapping data will be quite low. A low-precision estimate will be advantageous if accuracy requirements are considered to have been met if the true value is contained within the (comparatively wide) confidence interval. However, if the accuracy of the estimate is required to be equal to or greater than the required value, a high-precision (narrow confidence interval) estimate will be preferable, because, with this requirement, the true value must lie at the lower edge of or outside the confidence interval.

5.2 Statistical Methods for Thematics Accuracy

Many of the techniques for measuring uncertainty in mapped classes were developed for remote sensing to provide interpreters with ways to assess the accuracy of remotely sensed land classifications. In this kind of accuracy assessment, it is common practice to select a sample of locations and to compare the class assigned to each location with some source of higher locations and to compare the class assigned to each location with some source of higher accuracy, usually ground truth obtained by direct observation in the field. The results are then tabulated in the form of an error or misclassification matrix (also referred to as a contingency or confusion matrix), such as the one shown in Table 2 (Congalton and Mead 1983; Congalton, Oderwald, and Mead 1983; Story and Congalton 1986; Rosenfield and Fitzpatrick-Lins 1986; and Congalton 1991).

Table 2: Sample Misclassification Matrix for Five Classes

Sample Data	Reference data					
	A	B	C	D	E	Total
A	80	4	0	15	7	106
B	2	17	0	9	2	30
C	12	5	9	4	8	38
D	7	8	0	65	0	80
E	3	2	1	6	38	50
Total	104	36	10	99	55	304

In this table, the columns define the classes in the reference data, and the rows define the classes in the data being evaluated for accuracy. The values in the cells in the table indicate how well the classified data agree with the reference data.

The diagonal elements of the matrix indicate correct classifications. Therefore, a crude overall measure of accuracy is the percentage of cases that lie on the diagonal, in this case 209/304, or 68.8%. In other words, overall accuracy can be expressed as follows:

$$\frac{\Sigma(\textit{items along diagonal})}{\Sigma(\textit{all row or column entries})}$$

Unfortunately, this can be misleading as an index, since a certain number of correct classifications will occur by chance, even in the most uncertain situations. A preferred index is the Kappa index, which has a maximum of 1 and a minimum of zero, the latter expected under

maximum uncertainty. Kappa is computed as follows (Foody 1992):

$$\hat{k} = \frac{P_{correct} - P_{chance}}{1 - P_{chance}}$$

where $P_{correct}$ is the proportion of correctly classified entries and P_{chance} is the proportion of samples that could be expected to be classified correctly by chance. P_{chance} is computed as follows (Foody 1992):

$$P_{chance} = \sum_{i=1}^{n} P_{row(i)} P_{column(i)}$$

where $P_{row(i)}$ is the proportion of total entries that are in row i, $P_{column\ (i)}$ is the proportion of all entries that are in column i, and n is the total number of rows or columns.

For Table 2, the Kappa index is 58.3%, which is somewhat lower than the percent correctly classified. Some evidence exists that the Kappa index in fact underestimates true accuracy in the data, and modifications to account for this exist (Foody 1992). However, it is recommended that the unmodified Kappa index be utilized to report overall classification accuracy for the NPS/USGS Vegetation Mapping Project, primarily because it is documented and well known within the vegetation mapping community.

An overall statement of accuracy is useful, but it does not say anything about the accuracy of individual classes. Per-class accuracies can be extracted from the contingency matrix, but they can be misleading if they have not been differentiated into producers' and users' accuracy (also referred to as errors of omission and errors of commission). Errors of omission calculate the probability that a reference sample has been classified correctly; this is often the only type of accuracy that is reported. This quantity is computed by dividing the number of samples that have been classified correctly by the total number of reference samples in that class. Errors of commission calculate the probability that a sample from the classified data actually represents that category on the ground. This type of error is computed by dividing the number of correctly classified samples by the total number of samples that were classified as belonging to that category. The users' and producers' accuracies for the example in Table 2, above, are summarized in Table 3.

Table 3: Users' and Producers' Accuracy for Five Classes

Class	Users' accuracy		Producers' accuracy	
A	75.5%	(80/106)	76.9%	(80/104)
B	56.7%	(17/30)	47.2%	(17/36)
C	23.7%	(9/38)	90.0%	(9/10)
D	81.2%	(65/80)	65.7%	(65/99)
E	76.0%	(38/50)	69.1%	(38/55)

Producers' accuracy is an important measure because the producers of spatial data are interested in how well a particular area on the Earth's surface can be mapped. Users' accuracy, on the other hand, is important for users of spatial data, because users are principally interested in knowing how well spatial data actually represent what can be found on the ground (Story and Congalton 1986). Ideally, both users' and producers' accuracy should be similar for all classes. However, as illustrated in Table 3, users' and producers' accuracy may differ considerably among classes. For example, Class C has a producers' accuracy of 90%, meaning that 90% of the reference samples were also found to be classified as Class C. Users' accuracy, on the other hand, is only 23%. This means that only 23% of the polygons classified as Class C in the data can be expected to be Class C when visited on the ground. Only reporting the producers' accuracy as part of the accuracy assessment would give a completely misleading picture of data accuracy to users, who would interpret the 90% figure to mean that 90% of the polygons classified as class C in the data also have this class on the ground.

Ideally, accuracy requirements for classification should address both users' and producers' accuracy, and it is recommended that thematic accuracy always be reported in terms of users' and producers' accuracy in order to give as complete a description of the error properties of individual classes as possible for both the users and the producers of spatial data. The accuracy standard envisioned for the NPS/USGS Vegetation Mapping Project is an 80% per-class accuracy for all thematic classes, and this implies a users' and producers' accuracy of 80%. However, depending on the degree of confusion that occurs between classes, meeting this requirement for both users' and producers' accuracy may not always be possible. In addition, it may be difficult to obtain an adequate number of samples in the case of rare classes. Furthermore, the accuracy of a classification is confounded by the classification scheme itself. If two classes are difficult to differentiate, it might seem reasonable to expect large numbers of cases in the appropriate off-diagonal cells of the misclassification matrix. For example, one might conclude from Table 2 that Class D is easily confused with Class A, since 15 of the 99 objects of Class D were wrongly classified as A. If A and D are lumped together, forming the misclassification matrix shown in

Table 4, the percentage of samples classified correctly rises to 76.0%, indicating a higher accuracy because of the simpler classification scheme. The Kappa index, on the other hand, drops from 58.3% to 56.0%, since chance agreement is now easier to achieve.

Table 4: Aggregation of Two Easily Confused Classes from Table 2

Sample data	Reference data				
	A + D	B	C	E	Total
A + D	167	12	0	7	186
B	11	17	0	2	30
C	16	5	9	8	38
E	9	2	1	38	50
Total	203	36	10	55	304

Because misclassifications must be estimated from a single set of samples, it is common to calculate one misclassification matrix for an entire data set. But in reality, the chances of misclassification vary not only by class, but also across the map, because misclassification errors are almost certainly more likely in some areas than others. As more is learned about the nature of errors, and why errors occur, it will be possible to produce more refined models of error calculation that take such spatial heterogeneity into account.

With a misclassification matrix, or similar estimate of the probabilities that a feature belongs to each class, it is possible to simulate error using Monte Carlo simulation, and to propagate error into products obtained from the data set. Goodchild, Sun, and Yang (1992) demonstrate this with a soil map example.

5.3 Statistical Methods for Positional Accuracy

Positional accuracy is best discussed within the framework of measurement. Position can be thought of as a combination of two measurements, x and y, representing the easting and northing of a pair of UTM coordinates, or longitude and latitude, respectively. Each measurement is subject to error. For example, if 100 people were asked to measure the coordinates of a road intersection on a topographic map, the results would show variation in both coordinates. In practice, the variation in both coordinates is likely to follow a normal distribution or bell curve, with most measurements clustered, and a few extremes in both positive and negative directions. The amount of variation is likely to be similar in both coordinate directions. Moreover, errors are likely to be uncorrelated, in the sense that the direction and amount of error in one coordinate

is independent of the direction and amount of error in the other. Errors are also likely to be unbiased, in the sense that the average of all 100 measurements will be very close to the true location. If these assumptions are true, and there is no obvious reason why they should not be, the errors in both coordinates can be visualized as a three-dimensional bell curve, or circular normal distribution.

Several statistics of the circular normal distribution are in common use to describe positional accuracy. Perhaps the most commonly used is the Circular Map Accuracy Standard, or CMAS, defined as the 90th percentile of the circular normal distribution, or 2.146 times its standard deviation. More graphically, it forms a circle about the true location of the point, within which the observed location is expected to lie 90% of the time. Using the example of the topographic map, it might turn out that 90% of the 100 people determined the road intersection's coordinates to within 0.5 millimeter of their true locations at the scale of the map, leaving 10 people with positional errors of more than 0.5 millimeter.

CMAS is a well-defined statistic, and it forms the basis of the current National Map Accuracy Standard for positional accuracy (Bureau of the Budget 1947). However, map accuracy standards are currently being revised, and the standard is likely to become the National Cartographic Standards for Spatial Accuracy, as proposed by the American Society for Photogrammetry and Remote Sensing (ASPRS). With this standard, positional accuracy is reported as a standard error in each coordinate direction; it is a measure of variability that includes the effects of both bias and random error. It is computed as follows (Merchant 1985):

$$RMSE_x = \left[\frac{1}{N} \sum_{i=1}^{n} (\delta X_i - \delta X_c)^2 \right]^{1/2}$$

where $RMSE_x$ is the standard error in the x-coordinate direction, N is the sample size, δX_i is the actual (measured) coordinate location, and δX_c is the true coordinate location as determined in the source of higher accuracy.

Using standard errors to report positional accuracy is not in conflict with the CMAS, because standard errors can be related to CMAS, provided that the assumption that errors follow circular normal distributions holds. Since it is common practice to state positional errors at a 90% confidence level, it is recommended that both the standard error in each coordinate direction, and the circular error derived from these errors, be reported. The relationship between standard error and circular error is as follows (ASPRS 1989):

$$CMAS = 2.146 * \sigma_x \text{; or } CMAS = 2.146 * \sigma_y$$

where σ_x and σ_y are the standard errors in each coordinate direction.

CMAS and circular errors are useful measures when positional accuracy is assessed from a sample of points. For general purposes, however, it is necessary to use a less rigorous approach. Unless otherwise specified in this section of the report, the term "positional accuracy" should be interpreted as a linear measure approximately equal to the CMAS, and based on the same assumptions, but not implying the same rigor of definition.

CMAS can also be used to describe the positional accuracy of points relative to each other. The accuracy with which two points can be positioned relative to each other depends on the positional accuracy of both. If it is possible to assume that errors in the two points are independent, a simple calculation can be used to determine the error in relative position by taking the square root of the sum of the squares of the two positional accuracies. For example, if one point has a positional accuracy of 1 millimeter and a second point has a positional accuracy of 2 millimeters, a rough estimate of the error in their relative positions is:

$$(1^2 + 2^2)^{1/2} = 2.24 \text{ millimeter}$$

or 54 meters on the ground for a 1:24,000 product.

5.4 Confidence Intervals and Sample Size

Since accuracy estimates are based on samples, a point estimate (e.g., of a mean or variance) requires an associated confidence interval to be truly useful. A confidence interval is an interval within which we have a specific level of confidence the true value of an estimate lies. The width of a confidence interval is affected by the sample size used to derive the point estimate, and by the confidence level itself. Larger sample sizes will result in a more narrow confidence interval, as will lower confidence levels. Smaller sample sizes and higher confidence levels will widen the confidence interval. Confidence levels are often held fixed for a given study (i.e., all values are reported to a predetermined level of confidence), where conventionally used confidence levels are 90%, 95%, or 99%. Therefore, the width of the confidence interval will vary with changing sample size. If, as is recommended for the NPS/USGS Vegetation Mapping Project, sample size varies with class abundance, confidence intervals for rare classes will be wider than those for abundant classes. An alternative approach would be to hold the width of the confidence interval at a fixed level. With a given fixed confidence interval, our certainty (confidence level) that the true value of the estimate falls within the interval would decrease for a smaller sample size. Rather than report that an estimate falls within a variable-width interval with a fixed certainty, it is considered preferable to report that an estimate falls within a fixed interval with varying certainty based on the sample size used to derive the estimate.

Confidence intervals may be two-sided or one-sided (also referred to as two-tailed or one-tailed). Two-sided confidence limits for an estimate express the probability that a point estimate falls within the interval. One-sided limits express the probability that the true value of the estimate lies outside the limit of the interval at one specified end. For accuracy assessment purposes, investigators are typically interested in either a two-sided estimate or a one-sided estimate that expresses whether or not true value is at least equal to or exceeds the estimated value.

The confidence interval for a binomial distribution is obtained from the following equation (Snedecor and Cochran 1967, p. 211):

$$\hat{p} \pm \left\{ z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} + \frac{1}{(2n)} \right\}$$

where z_α comes from a table of the z-distribution at the significance level a (this value is 1.645 for a two-sided limit, and 1.282 for a one-sided limit with a 90% confidence interval. For a 95% significance level, the value changes to 1.645 for a one-sided limit and 1.960 for a two-sided limit), is the accuracy estimate, and n is the sample size. The term 1/(2n) is the correction for continuity. The correction should be applied to account for the fact the binomial distribution describes discrete populations. For large sample sizes, the correction will become very small, as should be expected, because for large populations, the normal distribution is a good approximation of the binomial one.

Table 5 below shows the appropriate z-values and two-tailed and one-tailed confidence limits for the recommended sample sizes of 5, 20, and 30.

Table 5: Confidence Intervals for Given Sample Sizes and Confidence Levels

Sample size	One-sided				Two-sided			
	z-value		Confidence interval		z-value		Confidence interval	
	90%	95%	90%	95%	90%	95%	90%	95%
5	1.282	1.645	0.33	0.39	1.645	1.960	0.39	0.45
20	1.282	1.645	0.14	0.17	1.645	1.960	0.17	0.20
30	1.282	1.645	0.11	0.14	1.645	1.960	0.14	0.16

The confidence intervals above can be converted from proportions to percentages by multiplying each proportion by 100. For the 90% confidence level, confidence intervals (in percent) are between 33% and 39% for the smallest sample size and decrease to between 11% and 14% for a sample size of 30. If an 80% accuracy estimate is assumed for a given class, a two-sided interpretation of the confidence interval with a sample size of 20 and a confidence level of 90% implies that with 90% certainty, the true accuracy value lies between 80% ± 17%. With a one-sided interpretation, it can be concluded that with 90% certainty, the estimated value is equal to or greater than 66% (80%–14%).

Confirming whether or not the estimate conforms to a previously established accuracy requirement is solved through hypothesis testing, which is further explained in the section below.

5.5 Hypothesis Testing

The purpose of hypothesis testing is to determine whether an estimation function supports the assumption (hypothesis) that a sample has been drawn from a population with specified parameter values, such as a normal distribution with a given standard deviation (Mikhail and Ackerman 1976). Therefore, a statistical hypothesis is a statement about the probability distribution of a random variable. In general, any statistical test requires the formulation of a null hypothesis (H_0) (theoretical expectation of a random variable) and an alternative hypothesis (H_A). The null hypothesis always specifies the set of theoretical distribution parameters of a random variable against which the parameters estimated from the sample are to be compared. The outcome of this comparison determines whether the sample parameters are in sufficient agreement with the null hypothesis. If this is the case, one concludes that there is sufficient evidence to accept the null hypothesis as true. Otherwise, the null hypothesis is rejected as false and the alternative hypothesis accepted as true.

In terms of accuracy assessment for the NPS/USGS Vegetation Mapping Project, two parameters are of interest:

1. The estimate obtained for positional accuracy. Here, the objective is to determine whether or not the estimate is equal to or exceeds National Map Accuracy Standards.

2. The estimates obtained for thematic accuracy, both overall and on a per-class basis. Here the objective is to determine whether or not the estimate exceeds 80%.

5.5.1 Determining Conformance to NMAS for Positional Accuracy

It is assumed that the National Cartographic Standards for Spatial Accuracy will become the new national map accuracy standards, and conformance to NMAS requires that 1:24,000 products have a standard error of no more than 6.0 meters for Class 1 products (12.0 meters for Class 2 products). In order to determine whether the estimated standard error in either the x- or y-coordinate direction meets this accuracy requirement, the null hypothesis and alternative hypothesis are formulated as follows:

H_0: The theoretical (required) standard error and the estimated standard error for a given coordinate direction are the same. Under these conditions, the random variable defined below:

$$\chi^2 = \left[\frac{(n-1)}{\sigma_x} RMSE_x \right]$$

follows a χ^2 distribution with n-1 degrees of freedom. σ_x is the required accuracy standard, and $RMSE_x$ is the estimated standard error.

H_A: The estimated standard error is different from the required standard error.
The data are accepted as meeting the accuracy requirement (i.e., H_0 is accepted) if the following is true:

$$\left|\chi^2\right| = \chi^2_{n-1,\alpha}$$

or, if the computed χ^2 test statistic is less than or equal to the theoretical χ^2 statistic at confidence level a with n-1 degrees of freedom. For a 95% confidence level (5% significance level) and a sample size of 30 points, the maximum value the computed χ^2 statistic may assume is 42.557.

5.5.2 Determining Conformance to an 80% Accuracy Requirement for Thematic_Accuracy

The way in which hypothesis tests are applied to thematic data depends on the way accuracy is defined. Three approaches are possible (Goodchild et al. 1994):

1. The hypothesis that 80% accuracy has been met will be accepted unless the sample map accuracy is low enough so that the conclusion that rejection is appropriate can be drawn with some predetermined degree of certainty.

2. The hypothesis that 80% accuracy has been met will be rejected unless the sample accuracy is sufficiently high so that the existence of 80% accuracy can be confirmed with some degree of certainty.

3. A given sample map is assessed for accuracy, and the probability is that the actual map accuracy exceeds the standard of 80%. This approach is not really different from (1) and (2); it simply rephrases the question of accuracy in probabilistic terms.

The first two approaches toward hypothesis testing can best be illustrated by example. Suppose classification accuracy for a specific class has been estimated from a sample of 20 points, 15 of which (75%) have been properly identified. With the first approach, the aim of the hypothesis test is to determine whether or not 75% is sufficiently low to consider it to be significantly different from the required accuracy of 80%. Whether or not 75% is sufficiently low depends on the required level of confidence for the estimate. Since the objective is to determine whether or not the estimate is sufficiently different from the required one, a two-sided confidence limit is in order. Table 5 above indicates that the two-sided 90% confidence interval for a sample size of 20 points is 15.88 and therefore ranges from 59.12% to 90.88%. Since the required map accuracy lies within this interval, it can be concluded that the claimed and estimated accuracy are the same. Therefore, the estimated classification accuracy meets requirements.

In formal terms, the null and alternative hypotheses would be formulated as follows:

Required level of confidence: 90%
Sample size: 20
Estimated accuracy: 75%
Required accuracy: 80%
Confidence interval: 75% ± 15.88

H_0: The estimated accuracy and the claimed accuracy are the same. Under these conditions, the random variable comparing the estimated and claimed accuracy,

$$t = \frac{\hat{p} - p}{\sqrt{\frac{p(1 - p)}{n}}}$$

will follow a t- distribution with n-1 degrees of freedom. In the equation, n is the sample size, p is the required accuracy, and $\hat{p}$ is the estimated accuracy.

H_A: The estimated and claimed accuracies are different.

If the calculated value for t exceeds the threshold of 1.72 set for a two-tailed hypothesis test at a 90% confidence level, the null hypothesis should be rejected. Otherwise, the null hypothesis should be accepted. For the given example, t is calculated as 1.123. The claimed and estimated values should therefore be accepted as the same.

With this approach to hypothesis testing, it is relatively easy to meet accuracy requirements, especially with small sample sizes (and correspondingly wide confidence intervals), because the sample map accuracy must be low before the assertion of 80% accuracy can be rejected. This approach encourages the use of small sample sizes (Goodchild et al. 1994). To limit the width of the confidence interval, a low level of significance should be chosen (0.10 or larger), in order to limit the acceptable error in the sample.

A much stricter interpretation of accuracy is the requirement that the required classification accuracy be met or exceeded at a given level of confidence. This approach is also best illustrated by example. Suppose that classification accuracy for a given class has been estimated at 90% with a sample size of 20. We wish to test the hypothesis that the 90% accuracy estimate meets or exceeds the required 80%. In formal terms

Required level of confidence: 90%
Sample size: 20
Estimated accuracy: 90%
Required accuracy: 80% or greater
Confidence interval: 90% ± 12.2%

H_0: The estimated accuracy is less than the required accuracy.
H_A: The estimated accuracy is equal to or greater than the required accuracy.

If the calculated value for t exceeds the 90% threshold (1.32) for this one-sided test, the null hypothesis should be rejected, and the alternative accepted. However, the t-statistic under these conditions is 1.118. Therefore, the alternative cannot be accepted. Even though the estimated accuracy value was considerably higher, the width of the confidence interval prevents a conclusion with sufficient certainty that the estimated value meets or exceeds the required one. Clearly this approach favors large sample sizes, in order to narrow confidence intervals (and therefore increase the chance that a given point estimate lies above the threshold established by the upper limit of the confidence interval). For example, with a sample size of 30, the t-statistic increases to 1.34. This exceeds the threshold of 1.32 at a 90% confidence level and would therefore permit the alternative hypothesis to be accepted. Again it is recommended that a low significance level (90% or lower) be used, because this will make it easier to achieve the required accuracy standard.

5.6 Summary of Recommendations Regarding Confidence Levels and Hypothesis Tests

Meeting accuracy requirements at the second, more demanding level will require that the point estimate be much higher than the required accuracy, especially if sample sizes are small. At the same time, it is unlikely that sample sizes on a per-class basis can be increased much beyond 30_samples per class, even for the most abundant classes. In fact, it is strongly recommended that sample sizes of less than 30 be used for rare classes. Since the majority of classes are in fact expected to be rare, sample sizes of 20 or less will be the more common scenario. With these sample sizes, meeting or exceeding the required accuracy will be nearly impossible. On a per-class basis, it is therefore recommended that the less stringent approach be used to confirm accuracy. In other words, the accuracy requirement will be considered met, if it can be concluded with sufficient certainty that the sample accuracy and the required accuracy are in fact the same (two-sided approach). In addition, it is recommended that 90% confidence intervals be computed on a per-class basis.

Although sample sizes on a per-class basis are small, it is likely that overall sample size for thematic accuracy assessment will be sufficiently large to draw more rigorous conclusions regarding overall classification accuracy. It is therefore recommended that conditions be defined for meeting overall accuracy by using the more demanding requirement of either meeting or exceeding the established accuracy standard. Again it is recommended that 90% be used as the required level of confidence.

6.0 Error Management Issues

6.1 Accumulation and Tracing of Errors in Output Products

Errors may be introduced into spatial data at any step during data production and data maintenance. Understanding the types and magnitude of errors that can occur at individual production steps is an important part of error management. An analysis of error rates at specific production steps can identify error-prone processes that can be subsequently modified or targeted for intensified quality assurance checks. Concurrently, quality assurance procedures may be relaxed for processes known to introduce little error.

Accuracy assessment usually only gives a measure of the total error present in the database. Total error in spatial data is usually modeled as the sum of errors introduced during each production step, where "sum" should be interpreted in the sense of "combination." This total error report is of importance for communicating the fitness of a product for use, but has limited utility for evaluating the production process or interim products. To accomplish the latter, it is necessary to trace or partition the errors to the individual production steps. This is, of course, best undertaken as part of in-process quality assurance procedures rather than after a poor accuracy assessment (see Section 6.3.4).

The tracing or partitioning of error propagation in the production process begins with a model of the production process and the known causes of error. Ideally, expected rates of error for each production step are estimated when production and quality assurance procedures are developed for the project. This provides a control or benchmark for further error analysis. For the NPS/USGS Vegetation Mapping Project, estimates should be developed through controlled experiments in the prototyping phase of the project.

For illustration purposes, the main production steps and related sources of error for a project like vegetation classification for national parks can be summarized as follows:

- Data collection. The main data sources are aerial photographs and vegetation data collected through field sampling. Possible errors include uncertainties in the delineation of boundaries on aerial photographs as well as misidentification of vegetation communities in the field.

- Photogrammetric corrections. Depending on the final mapping methodology, the aerial photographs may have to be registered to ground control. Since locating ground control points is not error free, the transformations used in the photo rectification process will have some positional error associated with them.

- Data compilation. The rectified data from the aerial photographs and the data collected in the field will be reconciled and compiled onto stable base Mylar. Apart from interpretive errors in the delineation of boundaries, this will introduce additional errors because of the width of the pen used to delineate the polygon boundaries.

- Data input. Once compiled, the Mylar basemaps will be either scanned or digitized. Possible errors are instrument errors (both scanners and digitizers have limits in terms of registration and resolution), operator error (e.g., poor boundary definition during digitizing), or poor quality vectorization software (to convert the scanned raster data into an appropriate vector-based product).

- Data processing. This involves establishing polygon topology in the data and the attribution of polygon data according to a prespecified database design. Possible errors include inappropriate processing tolerances, incorrect attribution, or transformation/projection errors.

6.2 Differentiation of Regional Error Rates

The NPS/USGS vegetation maps will be compiled as a series of park-specific projects. As such, the types of vegetation communities, as well as the time of the mapping and the personnel contracted to compile the data, will in all likelihood be different. Since all three of these factors may have an effect on error rates, this suggests that thematic as well as positional errors should be reported on a park-specific basis.

It is therefore recommended that a separate accuracy assessment be made for each individual park. Even if two separate parks are contiguous, it is recommended that separate assessments be made, because the parks may have to be mapped by different contractors at different times. This implies that vegetation communities occurring in different regions may have a different accuracy associated with them. Moreover, if the abundance of a community varies from park to park, the confidence with which its accuracy can be stated may also be different.

6.3 Weighting of Errors

Regardless of the classification method employed, some vegetation classes will be more similar to each other than others. Intuitively, it therefore seems reasonable to consider confusion between similar classes a less severe error than confusion between two very different classes. Given the subjectivity inherent in vegetation mapping, it may also be of interest to eliminate errors that may simply be due to interpretive differences among different investigators. Since analyses conducted on contingency matrices do not permit differentiation on the basis of severity, the question arises whether and to what extent other measures might be employed that would classify errors on the basis of severity and permit them to be treated differently in the accuracy assessment.

One possibility would be to simply aggregate easily confused classes in the contingency matrix as part of the accuracy assessment itself (see Section 5.3). The assumption behind this approach is that easily confused classes are also similar. The aggregation will result in a lower number of errors overall, but at the same time, information specific to each class will be lost to the user. From an accuracy assessment point of view, the two classes can no longer be differentiated. In addition, aggregating classes based on between-class confusion following compilation of the contingency matrix requires a considerable amount of judgment on the part of the investigator as to when two classes will be "sufficiently confused" to be considered similar.

Another method, also somewhat experimental in nature, approaches the problem from the perspective of fuzzy sets. With this approach, a particular area can be assigned to multiple classes, with varying degrees of class membership (Gopal and Woodcock 1994). Class membership for a particular area could, for example, be defined on a five-point linguistic membership scale as "absolutely wrong," "understandable but wrong," "reasonable class assignment," "very good answer," or "absolutely right" (Gopal and Woodcock 1994). For each vegetation class in a particular classification system, this would permit an a priori definition as to how class confusion from one class to another will be interpreted. For example, confusion of a given class A with class_B may be classified as "absolutely wrong," while confusion with class C may still be considered a reasonable answer. With this sort of scale, each area might be assigned a score that weighted its correctness, with "absolutely wrong" and "understandable but wrong" classes receiving low scores for correctness. Instead of the traditional confusion matrix, the results of an accuracy assessment based on fuzzy class membership are presented in the form of

tables that analyze the varying degrees of matching and mismatching that may have occurred between classes.

From an operational point of view, implementing this method will require a field investigator to assign one of the five membership classes for each class in the classification scheme to each sample point used for the accuracy assessment. However, for complex classification systems, or systems with many classes (as is to be expected for the NPS/USGS Vegetation Mapping Project), this procedure may be cumbersome and difficult, even if it yields results that are more appropriate to the nature of vegetation mapping than measures based on contingency matrices.

The ability to evaluate errors in terms of their severity is one of the most promising aspects of using accuracy assessment measures based on fuzzy sets. However, this approach is still relatively experimental, and traditional probability measures, such as confidence levels and confidence intervals, cannot be applied to it. While this is not necessarily a theoretical shortcoming of the method, it is an operational one. Since the meaning of accuracy measures based on fuzzy sets may not be understood by many users of spatial data, the use of this type of measure will make it difficult for users to evaluate the data in terms of its fitness for a particular application.

Although both aggregation of similar classes and accuracy assessment based on fuzzy sets may result in an interpretation of errors that is more representative of what constitutes an error in vegetation mapping, it is recommended that classification errors be reported in the form of a contingency matrix showing all classes that were part of the analysis. This approach is also recommended because any weighting of errors is inherently subjective in nature and depends to a great extent on the intended use(s) of the data, most of which may be unknown to the producers of the data. Reporting the full contingency matrix and the methods by which it was derived will then permit users to assign their own weighting schemes to the data if they desire.

6.4 Cost and Benefit Considerations

Cost constraints will be one of the factors that determine to what degree error rates in data can be reduced. Clearly, it is desirable to have as little error as possible in any given product. However, as a practical matter, reducing errors beyond a certain level will be prohibitively expensive. Determining an acceptable error rate in a product is the first step in justifying costs. This can

only be done by knowing the uses or intended uses of the data and the impact of accuracy on the success of the application.

When considering methods and associated costs to reduce error rates, care should be taken to distinguish between precision and accuracy (see Section 5.1). The first step in reducing error rates in a cost-effective manner should be to examine ways in which the accuracy (not the precision) of a measurement can be increased. In vegetation mapping, accuracy may remain comparatively low because of ambiguities inherent in the mapping process. If the accuracy of the mapping process cannot be improved, little will be gained by increasing the precision of the measurements used to compile the product and to determine its error rate.

This is important in determining the "point of diminishing return" when determining the degree of precision required for the mapping of a given product, because from a budgetary point of view a high degree of precision in measurements (including those used to confirm accuracy) will be one of the main contributors to cost. This is true because high-precision results require very large sample sizes, as well as more expensive equipment. As an example, being able to state accuracy based on a sample with an allowable error of 10% typically requires fewer than 100 samples even for 99% confidence levels. Stating the same accuracy with an allowable error of 0.01% requires several thousand samples (see Table 1).
Generally, in situations where a high degree of accuracy is inherently difficult to achieve, sampling to achieve a high level of precision will for the most part not be worth the cost. This implies that for vegetation mapping, with its inherent ambiguities, comparatively little can be gained by collecting many samples. At the same time, the use of high-precision instruments (or extremely conservative processing tolerances) is unwarranted if the accuracy of the position of lines (e.g., vegetation boundaries) has a high level of uncertainty associated with it.

Prior experience with land-use classification and vegetation mapping using remote sensing data indicates that it is unlikely that accuracies greater than 80% can be achieved on a consistent basis for individual vegetation classes, even in a large-scale product that utilizes field sampling methods in conjunction with aerial photography (Goodchild et al, 1994; Congalton and Green, 1993; Congalton and Mead, 1983). Given the comparatively small gains in accuracy that can be achieved at high cost, it seems unwarranted to sample to a level of precision higher than 10%.

The extent to which accuracy can be achieved can only be determined during operational testing of the mapping and accuracy assessment process. The results from these tests can then be used to derive a level of precision that is reasonable in terms of both product accuracy and cost.

6.5 Capturing and Incorporating Uncertainty as Part of the Data

Incorporating information regarding errors and data quality is particularly important for digital spatial databases (as opposed to traditional hard-copy maps) and requires substantial rethinking on how such information ought to be represented (Chrisman 1984). In hard-copy products, scale and the map detail associated with it are fixed. Users typically associate smaller scales with lower accuracy, even if no accuracy statement is explicitly given as part of the map. Small-scale maps, with their limited amount of detail, more or less prevent users from utilizing these data sources for high-accuracy projects. In digital data, the concepts of scale and data detail are much more ambiguous.

Most computer systems permit unlimited enlargement capabilities and coordinate precision that is independent of the accuracy of a given product. This tends to give users a false sense of accuracy (Chrisman 1984). At the same time, one digital data source can be combined with another (be it through overlay or simply edge match operations), regardless of how compatible the data sources are in terms of accuracy. Unless data producers and users explicitly propagate data accuracy and data quality information through these types of processing steps, data accuracy information may be lost from the database. The ability to maintain data accuracy information in static as well as dynamic databases and to propagate it into derived data sets requires some thought in terms of database design. At the same time, it needs to be considered on a project-specific basis what types of data accuracy and data quality measures would be most beneficial to data producers and users, and how this information can be collected.

Issues to consider for the NPS/USGS Vegetation Mapping Project include how to incorporate thematic and positional accuracy statements. Should producers report simply whether the data meet a given standard; should they report the actual accuracies derived with their associated confidence intervals; or should they report the full contingency matrix (table of test point discrepancies in the case of positional errors) used to derive the accuracy estimate? In terms of thematic accuracy, should accuracy information be class based, feature based, or both; and should the accuracy measures differentiate regionally between error rates? Finally, in terms of

vegetation mapping, should the data accuracy information include estimates of less tangible quantities, such as uncertainties in the position of lines due to transition zones and the heterogeneity of vegetation polygons?

Depending on physical database design, errors inherent in spatial data may be incorporated in a database either as part of the database itself, or in the form of metadata. In each case, the information would be represented in the form of additional attributes that may be either feature or layer specific. Although the information incorporated in the data may simply consist of a positional and thematic accuracy statement for a specific park, additional, and possibly more experimental, uncertainty measures may be incorporated into the database as well. This is of particular interest for those types of errors that are not easily quantified using traditional methods. This includes uncertainties in vegetation boundaries in transition zones as well as uncertainties in vegetation class assignment, especially where the species mix is heterogeneous. Collection, incorporation, and maintenance of such data obviously bears a cost, but the expense may be warranted if the information captured is deemed useful and important by data users.

In addition to traditional measures estimating positional and thematic accuracy, the following measures may be of interest for accuracy assessment of the NPS/USGS vegetation database:

- Width of the Perkal ε-band. This measure uses a band of width ε surrounding the line such that the band contains the line with a known certainty (Stoms et al. 1994). Methods for estimating the width of the Perkal ε-band are discussed in Blakemore (1984).

- Width of boundary zones as defined by the mapper directly. An experienced photointerpreter will in most cases implicitly know the width of a particular transition zone and will probably choose the center of a transition zone between two classes as the estimated "best guess" of the position of the line. Interpreters could be instructed to collect this information as part of standard practice. Since the measure would be line feature based, this would allow transitions that occur at sharply demarcated boundaries (such as those that might occur as the result of abrupt changes in topography) to be distinguished from those that occur in ecotones. Each line would have its characteristic uncertainty, which would vary with community type as well as with physiographic and environmental factors. The main disadvantages of this approach are that the measure is solely based on interpretive judgment, and as such does not lend itself well to

quantitative analysis. Different interpreters will assign different uncertainties to a specific line, and short of repeating the interpretive process for a given area using different interpreters, there will be little ability to control for this effect. Also, depending on the method used for interpretation and data input, collection of this information may be cumbersome.

- Mixing ratios within vegetation polygons. This is a measure that would show the degree of heterogeneity within a particular polygon. Indirectly it permits the assignment of multiple classes to a polygon, in that the interpreter may classify an area as, say, 80% class A and 20% class B. The greater the heterogeneity in a particular polygon, the more classes form part of the mixing ratio, or the more balanced the mixing ratio. This type of measure would be polygon specific. The main difficulty in determining mixing ratios lies again in the interpretive nature of the measure. Field investigators would have to be carefully instructed on how to collect data for the estimation of mixing ratios and would probably have to extensively document their observations in order for them to be meaningful. Because of the interpretive judgment involved, mixing ratios should be reported no more precisely than to the nearest 10%.

- Size of unmapped inclusions within a polygon. An inclusion is a distinct vegetation community within a polygon that has not been mapped because it is smaller than the MMU. Estimating the size of such inclusions is a further measure of heterogeneity within polygons. As in the methods above, the main difficulty in employing the measure is the difficulty in defining a set of decision rules that permit field investigators to collect adequate data regarding inclusions in a consistent manner.

The ability to easily incorporate a variety of feature-based uncertainty measures in spatial databases is promising in terms of error management. Such information could be useful to producers and users of spatial data. Producers could benefit by gaining greater flexibility in representing error in data products. Users would benefit by being able to extract uncertainties for specific features rather than for entire classes. However, the success of any of these feature-based uncertainty measures ultimately depends on the ability to collect the data required to derive the measure reliably and cost effectively in the field. The degree to which this is feasible should be tested experimentally during the prototyping phase of the project. Such testing is recommended to determine on an experimental basis how well such uncertainty measures can be derived, and to what extent producers and users of the data can utilize the measures.

6.6 Quality Assurance

Accuracy assessment and quality assurance are related, in that a product subjected to well-defined quality assurance procedures is likely to be of higher accuracy. Generally, the implementation of a well-defined quality assurance program is one of the most valuable tools for reducing error in a data set. In order to be useful, quality assurance processes should consist of in-process reviews as well as reviews by independent personnel at critical junctures in the data production process. During process design, it is important to identify all critical steps in the production process, and the quality assurance procedures that will be used to verify the step, because the success of the following production step will probably depend on the successful completion of the step(s) preceding it. Implementing quality assurance steps at all critical junctures will minimize the need for expensive rework, if errors from early production steps are detected at a late phase of the project.

For the NPS/USGS Vegetation Mapping Project, critical quality assurance steps will probably include review and correction of the initial photointerpretation, reconciliation of field data with aerial photointerpretation, review and correction of manuscripts in preparation for scanning or digitizing, review of data input to verify quality of line work, review of advanced processing such as attribution of features, and final review of the data after incorporation into the database. As opposed to an accuracy assessment, which is based on sampling, quality assurance generally relies on exhaustive reviews. Reviews should be structured to incorporate feedback mechanisms, and they should encourage all personnel to assume joint responsibility for high quality in the product. Further discussion of quality assurance is found in Section 6 of the Field Sampling Methodology.

6.7 Effective Use of Local Expertise

Local experts may facilitate the vegetation mapping process, because they may know the location of specific classes based on experience. This is particularly useful for communities that cannot be delineated from aerial photographs, because these may be missed altogether in the mapping process. In addition to improving the accuracy of the product, local experts may also reduce mapping costs, because they are likely to have better information, not only on the vegetation characteristics of a certain area, but also on the best way to access the area.

6.8 Process Design

NPS/USGS vegetation mapping will be implemented as a series of park-specific projects. For each of these projects, the mapping process consists of a series of sequential steps, with the accuracy assessment forming the last step of the process. The objective of the accuracy assessment is to simply determine the accuracy of the data, not to correct errors. (Naturally, if error rates exceed requirements, rework may be necessary; however, if accuracy meets requirements, errors found in the sample used to verify accuracy will typically not be corrected.) In other words, accuracy assessment should be seen as independent of the mapping process itself. The advantage of this approach is that the assessment will be conducted at a time when the data will no longer change. (If, on the other hand, the accuracy assessment for thematic accuracy [for example] was conducted following the initial photointerpretation and field work, errors due to processing, such as incorrect coding for a particular vegetation class, would not be detected.) A disadvantage of conducting the accuracy assessment as the final step of the mapping process is that certain procedural errors may not be detected until all mapping is complete. However, this type of problem will be minimized by implementing effective quality assurance procedures.

7.0 The Accuracy Standard for the USGS/NPS Vegetation Mapping Program

7.1 Establishing Thematic and Positional Control

Since the compilation of the USGS/NPS vegetation data will depend on aerial photography as well as on field sampling, the recommended approach to accuracy assessment is to utilize actual ground checking as the source of higher accuracy. In order to successfully conduct field checking, investigators must be able to locate sampling sites for both positional and thematic accuracy assessment in the field. Given the high requirements for positional and thematic accuracy, the large scale of the product, and the small MMU, investigators must be able to locate sampling sites (i.e., coordinate locations) in the field with a considerable degree of accuracy. In general, sampling sites for thematic accuracy assessment do not need to be located with the same (high) level of precision as those for positional accuracy, since these sampling sites do not need to meet NMAS. However, to avoid thematic confusion due to positional errors, investigators should not rely completely on post-processing corrections to establish their true position in the field. Methods that can be used to locate well-defined sample sites in the field have been introduced briefly in Section 2.3.2. This section further discusses those methods that can be used to find coordinate locations that may not correspond to well-defined features.

Where no hard target point exists to identify the sample site on the ground (as is the case with sample sites for thematic accuracy assessment), locating a point in the field will be virtually impossible without the use of the Navigation Satellite Timing and Ranging (NAVSTAR) Global Positioning System (GPS). However, the process of determining coordinate locations by using GPS is not error free. The degree of error inherent in the position obtained from the receiver depends on various factors, including the number and type of receivers used and the method used to conduct the GPS survey. In general, the higher-precision GPS solutions are more costly. Before a solution is selected, the accuracy requirements of a given product should be carefully considered to determine whether the added expense of implementing one of the higher-precision GPS solutions is justified.

GPS surveying is a process by which highly accurate, three-dimensional point positions are determined from signals received from NAVSTAR satellites. A GPS is simply a range measurement device: distances are measured between the receiver point and the satellites, and the position is determined from the intersection of the range vectors. To overcome various signal

degraders and selective availability (i.e., the military's encryption of the signal), multiple GPS receivers are used in differential carrier phase GPS (DGPS) surveying mode to obtain the desired accuracy (U.S. Army Corps of Engineers 1991; and Federal Geodetic Control Committee 1988). There are presently three different modes for overcoming the inaccuracies of the encrypted GPS signal. The first and the oldest mode is conventional DGPS surveying utilizing a minimum of two GPS receivers. The second mode is real-time DGPS surveying utilizing a minimum of one on-site receiver and a real-time communication link with a distant GPS base station. The third mode is utilizing the Department of Defense's encrypted GPS receivers, which until recently were not available to other agencies of the government.

DGPS involves utilizing a base station GPS receiver placed on a known reference and adjusting the remote GPS receiver's data against the base station corrected to the known reference. Basically, a GPS receiver's antenna is placed directly above a known horizontal control survey marker, such as a National Geodetic Survey monument. The base station data are corrected against the known monument, creating a delta formula. This delta, or offset formula is then used to adjust the remote unit's data. Both units must be receiving data from the same satellites over the same period of time (U.S. Army Corps of Engineers 1991; and Frodge and Lanigan 1991).

A relatively new method is real-time DGPS. The corrected delta formula is converted to ASCII format and sent via a modem to a transmitter. This information is received by the remote GPS unit's radio receiver and modem, and is used to correct the remote receiver's data in real time. The major drawback of real-time DGPS is the inability to receive communication frequencies other than satellite communications in remote areas, which may cause skips in data transfer. If a skip in data transfer occurs, the measurements usually have to be repeated. The spatial accuracy of real-time DGPS is 2 to 3 meters, while DGPS is accurate to 1 meter or better (Bobbe 1992).

An alternative to DGPS is utilizing a Precise Positioning Service (PPS) GPS receiver instead of the Standard Positioning Service (SPS) receivers utilized commercially. PPS is an encrypted P-code service which is dynamic; it is the service utilized in the classified and controlled military GPS receivers. The advantage of these receivers, referred to as PLGRs (Precise Lightweight GPS Receivers) is that, depending on the receiver, they are accurate to 8 to 16 meters (32.8 to 52.5 feet) without requiring differential corrections. The disadvantages include lack of availability to the general public, insufficient accuracy for ground point control acquisition,

requirements for significant communication security as compared to commercially available GPS receivers, and inability to download the ephemeris data for post-processing to improve results (telephone conversation between Anthony Curtis, ESRI, and Kelly Bobbitt, Trimble Navigation, Aug. 2, 1994).

The use of GPS to assess the accuracy of data produced in the USGS/NPS Vegetation Mapping Program for both positional and classification accuracy will not only validate the product, but will simplify the accuracy assessment methodology. There are several ways to utilize DGPS to establish position during the accuracy assessment of the vegetation maps.

The best method is to use real-time DGPS, because real-time DGPS is of sufficiently high accuracy to establish position in the field without requiring post-processing. The field inspection team can navigate to the sample site position with a high degree of accuracy and can verify the classification of the sample site. Spatial accuracy can be assessed or measured in the same way, because of the high accuracy of the DGPS mode (1 to 5 meters, depending on the subscription service used, and less than 1 meter if post-processing is used).

By contrast, the PLGR system's spatial accuracy is only 8 to 16 meters, and post-processing is not possible, since the ephemeris data cannot be downloaded. Although this accuracy may be sufficient for thematic accuracy assessment, it is not sufficient to measure positional errors in 1:24,000-scale data expected to conform to NMAS.

However, because of difficulties with communication links or lack of subscription service transmitting sites, real-time DGPS may not always be available. Although satellite communication links are a possibility, the cost may not justify this option. If post-processed DGPS is utilized for site location, the locational data will be positionally accurate to 100 meters only, until post-processing has occurred. With this approach, the inspection team would have to navigate to a selected site utilizing conventional navigation and collecting GPS data. The classification accuracy would be assessed in the field, with the spatial data assessment taking place after the post-processing. As noted above, post-processed DGPS is the most accurate of the three methods (Evans 1992; and Adkins and Merry 1994).

Without the advantage of real-time DGPS and its high degree of accuracy, locating preselected sites for thematic accuracy assessment will be next to impossible, unless a hard feature like an improvement or a significant natural feature can be located in the source data and on the ground. To overcome these difficulties, the assessment team could select a general area for study instead of specific sample sites. The team could then utilize the digital data and uncorrected GPS to

study the area. The data could later be adjusted to the post-processed corrected DGPS data, and the final assessment of accuracy could be made. In other words, once the team located an accuracy assessment sample site as closely as possible, the team could record the raw GPS position and analyze the site in the usual manner. When the data were post-processed, the classification accuracy can be further evaluated by using the corrected position.

Positional accuracy does not depend on the availability of real-time DGPS. Instead, the hard features identified as control points are located in the field and their raw GPS coordinates are recorded. Once post-processing of these coordinates is complete, the post-processed (true) positions are compared to those in the digital database. The results of this analysis will be the positional accuracy of the end product.

In summary, the two most viable GPS solutions for establishing position in the field are conventional DGPS (utilizing post-processing to establish correct locations) and, where the appropriate communications links exist, real-time DGPS. Either one of these methods can satisfy the requirements for both thematic and positional accuracy assessments. In addition, both methods are commercially available. The use of the PLGR GPS technology would enhance the user's ability to locate preselected sites in the field for thematic accuracy assessment. However, this solution is not recommended, since it is not sufficiently accurate to establish positional control. In addition, military restrictions may further constrain the use of the system in the field.

7.2 Thematic Accuracy Requirements

For the USGS/NPS Vegetation Mapping Program, a per-class accuracy statement is required, and ideally each class will meet or exceed 80% accuracy. This accuracy must apply uniformly over the entire project area, and it must be applicable to the minimum level of classification detail represented in the vegetation data (Story 1994).

As discussed in Section 4.1.1, meeting or exceeding 80% accuracy on a per-class basis will be difficult, because the sample sizes used on a per-class basis will be too low to permit the derivation of a point accuracy estimate with a narrow confidence interval, even for abundant classes. This means that in order to pass a one-tailed hypothesis test, point accuracy estimates must be much higher than the required accuracy. Given the recommendation to decrease sample sizes to as low as 5, the rarest classes would never meet accuracy requirements, because they are expected to have confidence intervals with widths of 30% and more.

As an alternative, it is recommended that accuracy requirements be defined as being met if the appropriate hypothesis test indicates that estimated accuracy can be considered to be equal to required accuracy. This less stringent requirement is more reasonable, given that sample sizes on a per-class basis are unlikely to increase to much beyond 30 sample points, even for the most abundant classes.

Although per-class sample size is small, overall sample size should be sufficiently large in most cases to permit accuracy requirements to be more stringent. Therefore, it is recommended that the requirement for 80% or greater accuracy be maintained for the overall classification accuracy.

A 90% confidence level is recommended for both per-class and overall accuracy.

7.3 Positional Accuracy Requirements

The positional accuracy of the digital map product must be at least equal to National Map Accuracy Standards (NMAS) for 1:24,000-scale (7 1/2') maps. The NMAS was last revised in 1947. In order to comply with National Map Accuracy Standards (only the horizontal component will be considered for the USGS/NPS Vegetation Mapping Program), no more than 10% of the tested points can be in error by more than 0.02 inch (0.51 millimeter) when measured at the publication scale (Bureau of the Budget 1947). For a 1:24,000-scale product, this translates into an error of 40 feet (ca. 12 meters) in ground distance.

The National Cartographic Standards for Spatial Accuracy are currently being revised. The proposed new standards are based on work published by the ASPRS and are in their final review stage (as of January 1994)1[1]. The new accuracy standard is not necessarily in conflict with NMAS as published in 1947. It proposes to state accuracy as a standard error (RMSE) in the x- and y-coordinate directions rather than as a circular error with a 90% confidence level (as is indirectly implied by the NMAS of 1947). For a Class 1 product of scale 1:24,000, the RMSE is 6.0 meters (20 feet) at ground scale; for a Class 2 product, the maximum allowable RMSE is 12.0 meters (39_feet).

[11] ESRI is investigating if the final standard has been published.

Circular errors and the RMSE can be related provided that the following conditions are met (Merchant 1987):

1. The variances in the x- and y-coordinate directions are approximately equal,
2. Discrepancies are distributed normally around a zero mean, and
3. Sufficient test points (at least 20) are available to estimate the variances.

It is assumed that the final standard has already been published, or will be adopted in the very near future. The following discussion therefore refers to the revised standard, which will form the basis for positional accuracy statements associated with the USGS/NPS Vegetation Mapping Program.

Like the existing standard, the accuracy statements in the revised standard will apply to well-defined points. These are defined as features that can be clearly identified and located as discrete features in the map product as well as the source of higher accuracy used for comparison. Relatively few features fit the definition of well defined. Examples are sharply defined intersections (e.g., roads or railroads), corners of buildings, certain other discrete point features like benchmarks, or possibly targets that may have been placed for aerial surveys, provided that they are mapped in their precise locations (i.e., not just cartographically placed) with sharply defined symbols. Boundaries subject to interpretive judgment or certain natural boundaries subject to environmental fluctuations (e.g., river banks) generally do not qualify as well-defined points, even if they form intersections.

Since vegetation boundaries are nearly always interpretive, positional accuracy statements within the constraints of the National Cartographic Standards for Spatial Accuracy will not directly apply to the primary mapped features in the product, but will instead simply be well-defined features that are quite unrelated to vegetation boundaries. A positional accuracy statement in terms of the National Cartographic Standards for Spatial Accuracy should therefore be understood to indicate how well the map product is registered (geocoded) to its control, and not to indicate how well the vegetation polygon boundaries reflect their true positions on the ground.

Although positional accuracy statements in this context have little meaning in terms of vegetation boundary location, this information is important, particularly in locating vegetation communities. However, to assess the locational accuracy of vegetation community boundaries, alternative methods are required that address uncertainty in terms of more intangible quantities such as interpretive judgment and the inherently transitional nature of vegetation boundaries. Most of

these methods are experimental in nature and have not been tested operationally. They may, therefore, be time consuming and expensive to implement.

From an operational point of view, it is therefore recommended that positional accuracy testing be limited to the tests prescribed by the National Cartographic Standards for Spatial Accuracy. From an experimental point of view, it is recommended that some methods that provide accuracy measures for the position of vegetation boundaries be tested. These include the estimation of epsilon bands, the estimation of "ecologist/photointerpreter uncertainty" in the delineation of vegetation boundaries, multiple photointerpretations of the same area, the estimation of mixing ratios and inclusions, and the assignment of multiple classes to individual polygons (fuzzy membership). Because of their experimental nature, these methods have not been incorporated into standard procedures for accuracy assessment.

7.4 Dealing with Problem Areas and Classes

A problem area or class is defined as one in which accuracy cannot be adequately determined, or one in which accuracy requirements may for some reason not have been met. Such conditions may occur for several reasons:

1. It may be impossible to collect an adequate sample for the class, because the class is either rare or inaccessible. An inadequate sample may result in an unacceptably wide confidence interval for the class. If the accuracy requirement specifies the point estimate as needing to be less than the lower end of the confidence interval, the point estimate may need to be impossibly high in order to satisfy the requirement.

2. The class can be sampled adequately; however, it is easily confused with another class. The degree to which classes will be confused will depend to some extent on how well defined the classes are in the classification scheme and on whether the classes can be distinguished from one another, given the precision and accuracy of the field sampling techniques.

3. The number of control points in the area is insufficient to establish positional accuracy. Even if the number of points is sufficient, the points may not be distributed well through the map area.

4. The methods used to locate sampling sites in the field are not precise or accurate enough to adequately locate the site. This will pose a particular problem for classes that tend to occur in patches close to the minimum mapping unit in size.

5. Temporal variation has altered the composition of the vegetation in the area in terms of classification and vegetation boundary location. For example, a fire has burned part of several adjacent plant communities. The burnt area is now covered with new (fire-following) communities, and the boundaries of all communities adjacent to the burnt area have changed.

6. Climatic conditions during the sampling are different from those during the original survey, even though the same season is being sampled. As a result, some species typical of a given community have not yet become established, resulting in the assignment of the class to a different community. (This is similar to problem 2.)

The problems that may arise are too diverse to list here, particularly since some types of problems may occur infrequently, but some recommendations can be made to deal with the most common problems:

1. Relax accuracy requirements. This approach should be considered if, in the majority of cases, requirements cannot be met for reasons beyond the mapping contractor's control. For example, if assignment to a class cannot be done unambiguously, different interpreters may assign different classifications to the same class, and they may have equally valid justifications for the class assignment.

   Since any accuracy requirement is to some extent arbitrary, the difficulty with this approach is finding a sensible value that is not simply the lowest common denominator.

2. State actual accuracy rather than conformance to an existing standard. Rather than report whether the product meets a given accuracy requirement, simply state what the accuracy estimate is, together with its confidence interval. Like the approach above, this approach should be considered if accuracy requirements simply cannot be met for reasons beyond the mapping contractor's control.

   From a user's point of view, this approach is probably more helpful than a statement as to whether or not a map product meets a specific standard. However, from a producer's point of view, a requirement to simply state accuracy may provide no guidance as to the level of accuracy to be met and little incentive to achieve a certain level of accuracy.

3. In cases where small sample sizes prevent an accuracy estimate to be obtained with a

sufficient level of confidence, classes known to have similar error properties may be pooled to increase the sample size. This approach implies that a meaningful measure of a class's error properties can be developed.

4. Apply some sort of error weighting that takes into account confusion between inherently similar classes. In the extreme case, confusion between the most similar classes would not be counted as an error. How meaningful this approach is will depend on the classification method and the ability to define similarity between classes. One way to accomplish this would be to consider those classes as similar that are most often confused with one another.

5. Carefully document the instances in which accuracy is less than the minimum requirement (and the reasons for the violation). This approach should be considered if violations of requirements are restricted to a few classes or isolated areas. This approach is also recommended if discrepancies are due to temporal change or seasonal variability.

Prior experience with similar projects (e.g., Goodchild et al. 1994) suggests that meeting the proposed thematic and positional accuracy requirements will be difficult, even if sampling methods are purposely skewed in favor of selecting sampling areas that are classified correctly. Factors expected to impact the accuracy estimate include the interpretive nature of vegetation mapping, the transitional nature of many vegetation communities, and the limitations of the equipment likely to be used to establish sample locations (the latter two will also affect the estimation of positional accuracy). Even if accuracy standards are met, it will be difficult to interpret the meaning of the estimate, because differences in interpretation and actual misclassifications cannot be distinguished.

The extent to which the proposed accuracy requirements can be met can only be determined after the classification standard itself has been finalized and tested in the field. At this stage, it is recommended that the accuracy assessment procedure be field tested on diverse prototyping areas. The results of these tests can then be used to modify the assessment procedures or the proposed accuracy requirements for the USGS/NPS Vegetation Mapping Program itself.

8.0 Implementation

8.1 Overall Plan and Operational Scenarios

Accuracy assessment methodology depends to a considerable extent on the classification scheme used, the methods used for the original compilation of the data, and the accuracy requirements established for the data. As long as some of these issues remain unresolved, procedural recommendations must be based on assumptions that may or may not hold true. The current assumptions are as follows:

1. The MMU will be 0.5 hectare.

2. Initial vegetation boundaries will be delineated using aerial photographs; classification will then be refined using field sampling.

3. Vegetation mapping will be conducted separately for each park.

4. The accuracy requirements will be for 80% per-class accuracy for the vegetation classification and NMAS for positional information.

5. The mapping scale will be 1:24,000.

Operational scenarios have been defined on the basis of the frequency and abundance of each class. As such, the scenarios are class specific and not park specific. The sampling methods are essentially the same for all scenarios. However, the number of samples to be collected differs from one scenario to another.

From an operational point of view, it is intended to conduct a separate accuracy assessment for each park at the conclusion of the mapping process. As such, the accuracy assessment will be independent of the mapping process. The assessment process can be viewed as consisting of four basic steps:

1. Preliminary work and planning
2. Logistics planning for the field work
3. Field work
4. Analysis of field work and computation of results

The steps are illustrated in Figure 1.

Preliminary work and planning begins by analyzing the existing data to determine the classes to be sampled and their frequency and abundance. This analysis is used to determine the number of field samples to be selected per class. For each class, sample sites should be assigned randomly to polygons, with the constraint that all sample site locations be sufficiently far from polygon boundaries to minimize effects due to positional errors. Sample site locations should then be analyzed for possible accessibility problems and replaced with alternatives if necessary.

Once all sample sites have been determined, logistics planning for the field visit can begin. This includes a determination of the time and date of the field visit, the resolution of special situations (such as access problems due to administrative reasons), the order in which field sites should be visited, the personnel to visit the site, and the equipment to be used for the visit.

Figure 1 Accuracy Assessment Procedures Process Flow

a. Overall process flow

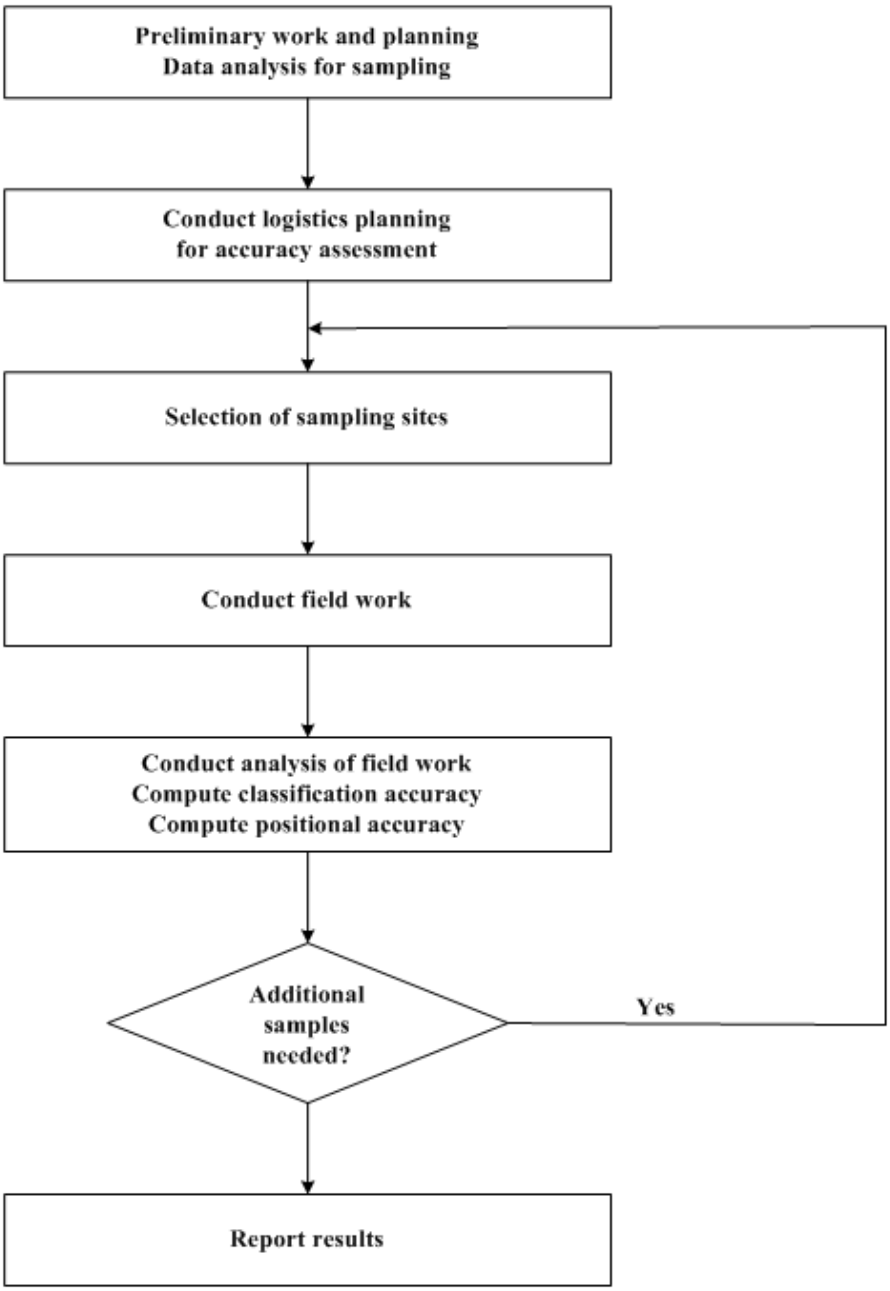


Figure 1 - Continued

b. Preliminary work and planning

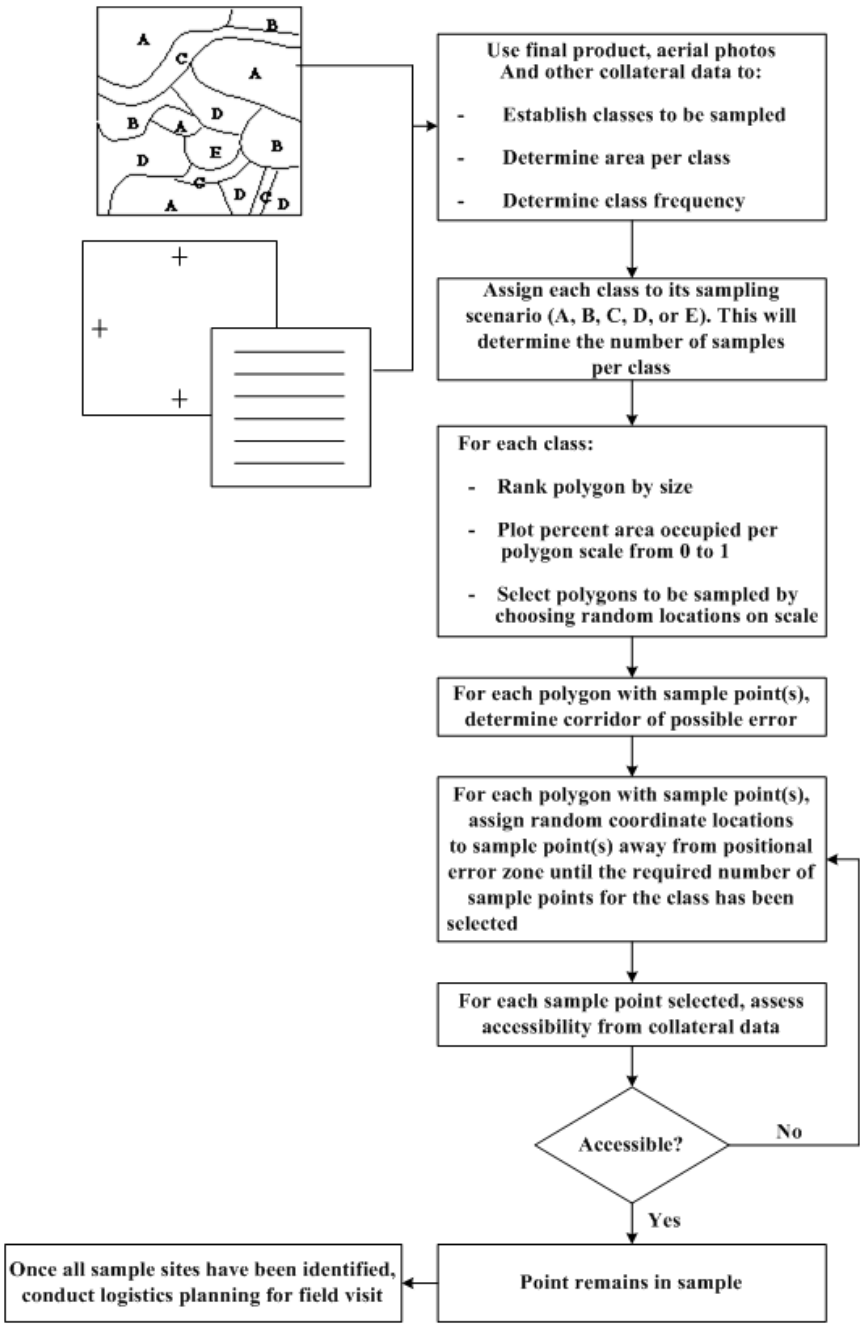


Figure 1 - Continued

c. Logistics planning for field work

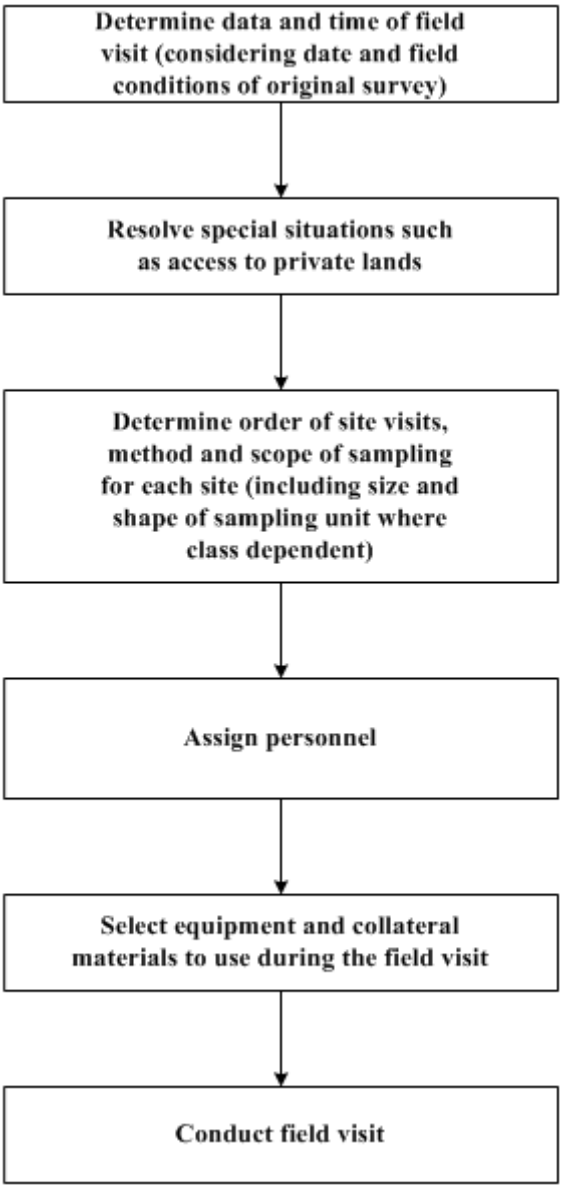


Figure 1 - Continued

d. Field visit

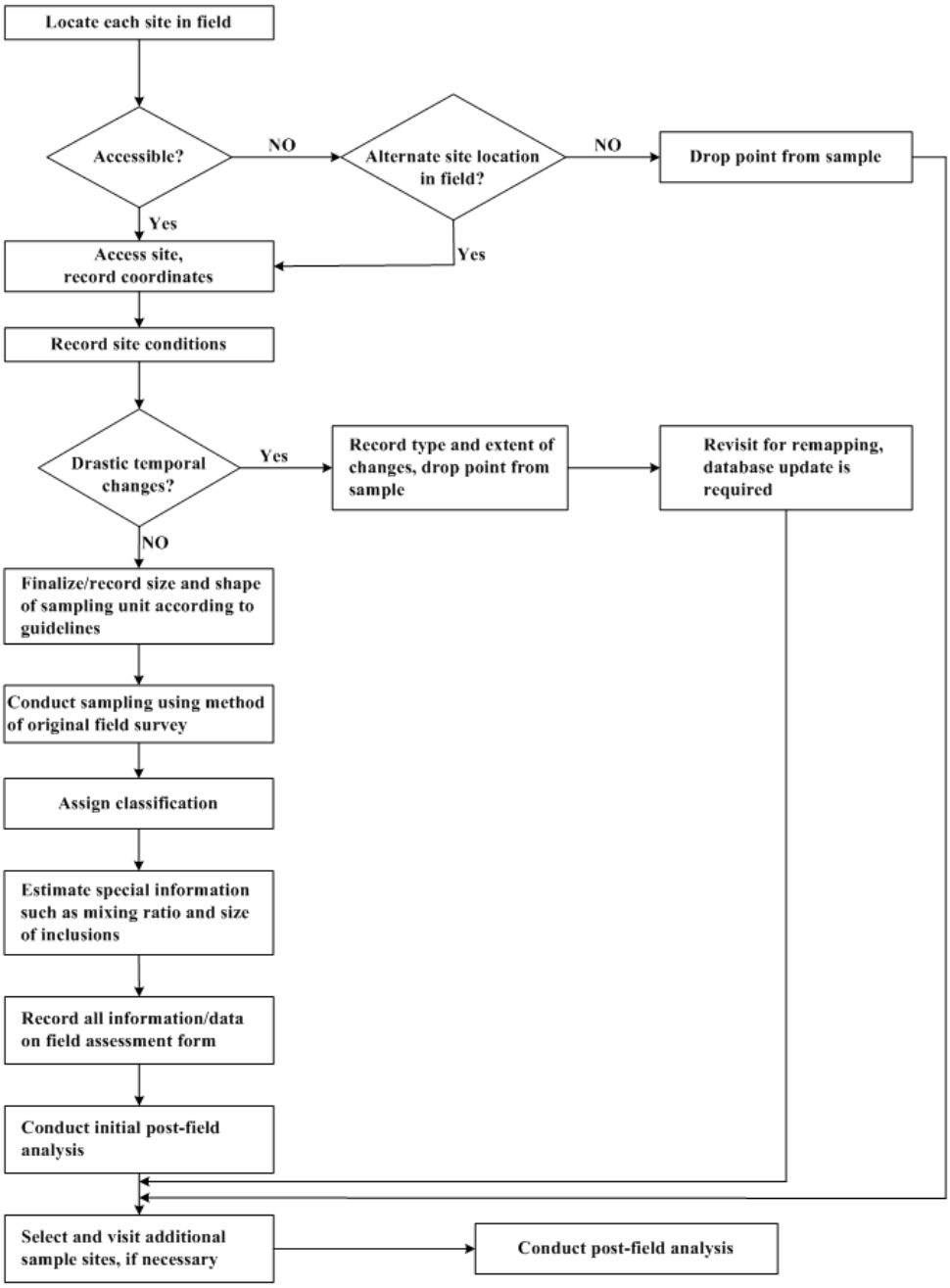


Figure 1 - Continued

e. Initial post-field analysis

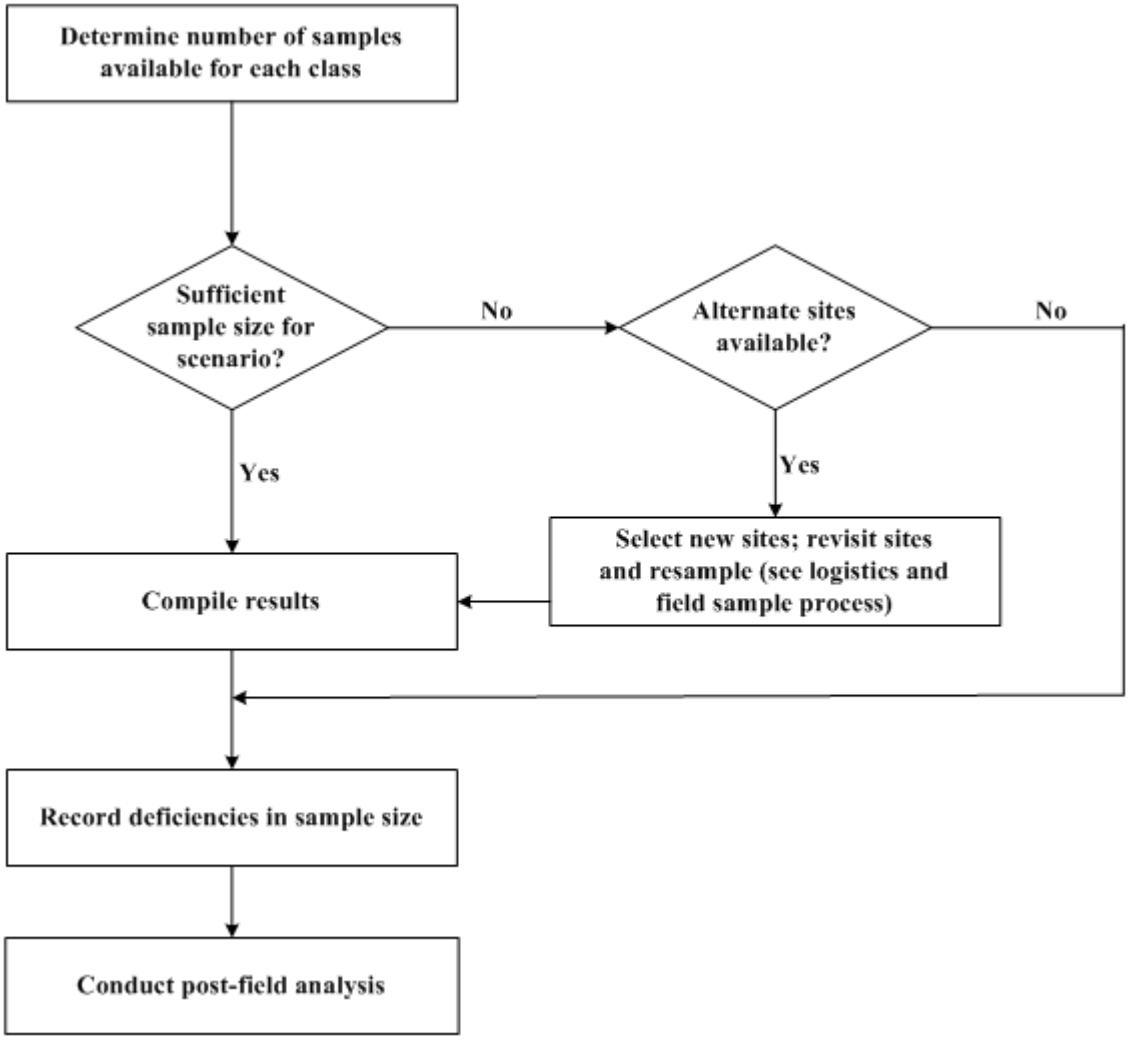
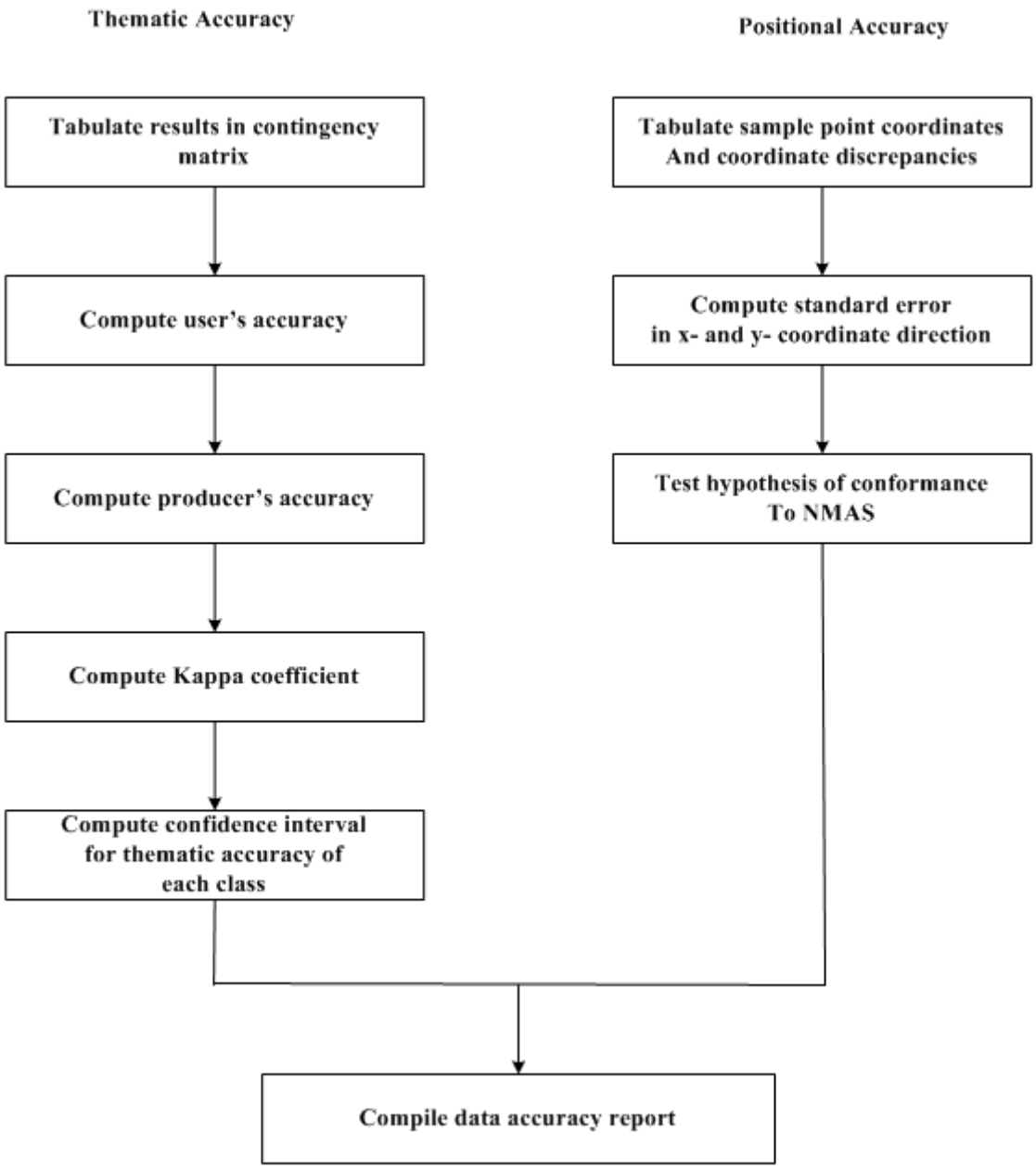


Figure 1 - Continued

f. Post-field analysis



The field visit itself is the most complex part of the assessment procedure. Investigators will need to locate each field site, determine its classification, and determine whether site conditions are appropriate for sampling. The last is necessary because a number of circumstances may prevent a sample site from being accessed or from being used, even if the site is accessible. If this occurs, sample sites may need to be replaced in the field or during initial post-field analysis. For accessible sites, an area equal to the MMU for that class will be observed in its entirety, and its classification and other pertinent information will be recorded on the field assessment form.

Once all sites have been visited, an initial post-field analysis will be conducted. The purpose of this analysis is to determine whether the data collected are sufficient to proceed with the final analysis or if a second field visit is necessary. A second visit may be necessary if, for example, the number of samples accessible in the field dropped below the number required for the class.

After completion of the initial post-field analysis (and any additional field work that may have been required), the data are analyzed for accuracy. The results of the analysis will be presented in the form of a report that will contain all pertinent information regarding thematic and positional accuracy.

8.2 Sampling and Data Collection

8.2.1 Number of Samples

It is recommended that the number of samples be varied according to the rarity of classes. Fewer samples should be allotted to rare classes.

Depending on the rarity of the class (in terms of both the area and the number of polygons), the number of sample sites selected can be decreased to as few as 5. Classes with fewer than 5 polygons should not be sampled, but observed in their entirety.

Wherever possible, a sufficiently large number of sample sites will be selected to permit the derivation of statistical measures to attain a satisfactory level of confidence. However, for some operational scenarios, especially those involving rare classes, achieving a satisfactory level of confidence may be impossible or prohibitively expensive.

The implications of the recommendation are that confidence intervals for thematic accuracy estimates will vary, depending on class. For classes with a sample size of 5, only a point estimate of the accuracy can be derived.

8.2.1.1 Total Number of Sample Sites

Sampling will be stratified by class. The total number of sample sites will therefore vary with the number of classes present in the park and the number of samples to be selected for each class. If the recommendation regarding sample sizes in rare classes is adopted, sample sizes may vary from 5 to 30, depending on the rarity of the class (see Section 8.2.1.2).

8.2.1.2 Number of Sites per Class

The following sample sizes are recommended for the operational scenarios described in Section 8.1:

Scenario	Description	Polygons in class	Area occupied by class	Recommended number of samples in class
A	Abundant. Many polygons that cover a large area.	≥ 30	≥ 50 ha	30
B	Relatively abundant. Class has few polygons that cover a large area.	< 30	≥ 50 ha	20
C	Relatively rare. Class has many polygons, but covers a small area. Many polygons are close to the MMU.	> 30	< 50 ha	20
D	Rare. Class has few polygons, which may be widely distributed. Most or all polygons are close to the MMU.	≥ 5, ≤ 30	< 50 ha	5
E	Very rare. Class has too few polygons to permit sampling. Polygons are close to the MMU.	< 5	< 50 ha	Visit all and confirm

8.2.2 Sample Site Selection

It is recommended that all sample sites be selected randomly, stratified by class. It will be permissible for there to be multiple sample points per polygon. However, the number of sample sites per polygon will be weighted by polygon area.

One possible weighting scheme is as follows:

1. For a given class, compute the percent area for each polygon.

2. Plot the percentages on a scale from 1 to 100 (i.e., polygons with a larger area will occupy a larger section of the scale bar).

3. Select random numbers between 1 and 100. The random number (sample) selected will be allocated to the polygon whose position on the scale includes the random number. Since larger polygons occupy a larger section of the scale, they have a higher chance of including samples.

4. Once the number of samples per polygon is determined, sample coordinates within the polygon will be randomly allocated.

5. Sample coordinates will be located away from boundaries to be sufficiently far from the boundary to minimize the effects of positional error in the data. At a minimum, the sample sites will be positioned such that the MMU to be observed will be wholly contained within the polygon.

For each sample site selected, a sample area equivalent to one MMU will be observed. Currently, the MMU is set at 0.5 hectare. However, it is recommended to define MMUs as class specific. The coordinates selected as the sample site location will be the center of the area to be observed. Ideally, the area to be observed should be circular. (If an MMU of 0.5 hectare is used, this will correspond to a circle with a 40-meter radius). However, in cases where a circular MMU will not fit into polygon boundaries (as may be the case for riparian zones), or where the layout of a circular area will be impractical, investigators will be allowed to vary the shape of the MMU.

8.2.2.1 Stratification

Samples will be stratified by class and weighted by polygon area. In the case of positional accuracy assessment, an attempt will be made to stratify samples by area (i.e., allocate an equal number of sample points to each quadrant). If this is not possible, a simple random sampling scheme will be employed.

8.2.2.2 Initial Positioning of Sample Sites

In terms of sample site selection, it is recommended that samples be positioned randomly within polygons, regardless of any ecotonal variation that may exist in the area. The only restriction on the positioning of the sample site is that the site must be located sufficiently far from the boundary to eliminate confusion due to positional error either in the data itself or in the ability of field personnel to find the sample coordinates in the field. For example, if the positional accuracy of the data is estimated to be 12 meters at a 90% confidence level and field personnel can position themselves within 15 meters of the true location of the point, the sample site should be positioned at least the root sum square of these two quantities from the polygon boundary (in this instance approximately 19 meters).

In terms of in-field sample site location, it is recommended that differential GPS be used to locate sample sites in the field, where possible. This solution (whether real time or dependent on post-processing) is the only one that can satisfy requirements for both thematic and positional accuracy. In case real-time DGPS is not available, investigators will only be able to navigate to preselected coordinates to within 100 meters. In this situation, investigators will be given some freedom in selecting a coordinate within a 100-meter radius of the predetermined one, as long as it remains in the appropriate community type.

8.2.2.3 Sampling between and along Gradients

The objective of the accuracy assessment is to identify a given community, not the location of a community boundary. The sample plot, which is equivalent to one MMU in size, should therefore be wholly contained within the community type.

8.2.3 Sample Collection

Samples should be analyzed utilizing the same field techniques as were employed when the identity of the community was originally established. It is assumed that community type will be established by using a combination of aerial photointerpretation and field surveys in the form of sampling. It is therefore recommended to utilize sampling to
verify community identity during the accuracy assessment. The field survey should be carried out by personnel familiar with this technique, but not by the same personnel who conducted the original field survey. If community identification is possible through simple observation, this will be an acceptable alternative to reconstructing the original sampling method. Once the data for the sample site have been selected, the class assignment should be made according to the same classification rules used in the original classification.

8.2.4 Reporting Results

Field personnel will report their findings on a field assessment form. The format of this form should be similar to the format used for the original classification. This will be an added check that the assessment methods will be as similar as possible to the methods used for the original survey. In addition to the standard items to record, the field assessment form should be used to record special conditions, such as drastic temporal changes, variation in the shape or size of the MMU, or relocation of a sample site to an alternative position. At the same time, reasons for deviating from the standard assessment procedure should be documented.

8.2.5 Field Assessment Form

At a minimum, the field assessment form should contain the following information:

- Name of the investigator(s).
- Name of the park.
- Date and time of the field visit.
- Name or identifier of the sample site location.

- GPS coordinates of the sample site in the field and the method by which the coordinates were determined (i.e., real-time DGPS, raw DGPS coordinates). If raw DGPS coordinates are used, the form should contain a description of the decision rules the investigator uses when navigating to the site, especially if the site coordinates are different from those originally selected.

- Sample site conditions in terms of environmental gradients, such as soil type, temperature, and overall development stage of the vegetation.

- Shape and size of the MMU used in the assessment. A description of the decision rules used to alter the MMU, if it does not conform to the standard shape and size.

- Site classification and a description of the procedures and methods used for the classification.

- All raw data collected to determine the classification (such as species counts).

- Reasons for dropping sampling sites and decision rules used to locate alternative sites.

- Description of special conditions that may have been encountered (e.g., drastic temporal changes, abnormal developmental stage of vegetation).

8.3 Statistical Methods

8.3.1 Positional Accuracy

To conform to the requirements of the proposed National Cartographic Data Standards for Spatial Accuracy, positional accuracy should be computed as the standard error of test point discrepancies in the x- and y-coordinate directions. From the standard error, a CMAS at a 90% confidence level will be derived. A 2 hypothesis test will be utilized to determine if the estimated accuracy meets or exceeds the established standard. The computational methods to be used to derive these quantities are presented in Sections 5.3, 5.4, and 5.5.

8.3.2 Attribute Accuracy

Attribute accuracy will be presented in the form of a contingency matrix, from which the following information will be extracted:

- Overall accuracy with 90% confidence interval
- Per-class users' and producers' accuracy with 90% confidence interval

Overall accuracy should be computed and adjusted with a Kappa coefficient. A t-hypothesis test will be utilized to determine if the estimated accuracy meets or exceeds the required standard of 80% at a 90% confidence level.

Both producers' and users' accuracy will be computed for all classes sampled. All accuracy estimates will be computed with a 90% confidence interval. A t-hypothesis test will be utilized to determine if the estimated accuracy of each class can be considered to be equal to the required class accuracy of 80% at a 90% confidence level.

The computational methods for deriving these quantities are described in Sections 5.2, 5.4, and 5.5.

8.3.3 Reporting Results

Results will be reported in the form of an accuracy report that describes both the thematic and positional error properties of the data. The report will include a map showing the position of the sample sites to give users an indication as to whether the accuracy assessment is applicable to the entire study area. Also included will be a brief outline of the assessment procedure or appropriate references to permit readers to review the actual assessment methodology.

For positional accuracy, the report will include the test point coordinates and their discrepancies from their true values. Also included will be the results of all computations, including the standard error in the x- and y-coordinate location, the CMAS at a 90% confidence level, and the results of the hypothesis test that determines whether accuracy requirements have been met. The

report will note any special conditions, such as points that were dropped from the sample as outliers.

For thematic accuracy, the report will consist of a contingency matrix showing all classifications. Samples that were dropped will be noted together with the reasons for their removal from the sample. The report will present both overall and per-class classification accuracy, the associated confidence intervals, and the results of the hypothesis tests that determine whether accuracy requirements have been met.

8.3.4 Labeling of Products

In addition to making a statement as to whether the product conforms to a specific requirement, it is recommended that the actual thematic and positional accuracy of each class and the associated confidence intervals be stored as part of the database. This information will be especially useful in situations in which a given class does not meet the accuracy requirement. In those situations, knowing the actual accuracy estimate will permit the users of the data to make informed decisions as to the usefulness of data that do not completely conform to established requirements. To conform to existing standards, such as SDTS, it is also recommended that other data quality information be incorporated in the database, such as reports on lineage, completeness, and logical consistency, which for the most part address expected levels of processing error.

Other measures of uncertainty, such as those associated with the position of polygon boundaries in vegetation transition zone, may also eventually become part of the database. Examples include the Perkal e-band, estimated "investigator" or "ecologist" uncertainties for individual lines, or estimated mixing ratios within polygons. Most of these address the uncertainty of individual features (either lines or polygons). However, since these measures are more experimental in nature, their inclusion in the database is not currently part of the standard recommended procedure. However, it is recommended that the feasibility of deriving these types of measures be tested during the prototyping phase of the project.

How accuracy-related information will be incorporated into the database will ultimately depend on the database design. Possible options include related database tables that link, for example, class-specific thematic accuracy values to individual polygons. Other possibilities include the use of a data quality coverage or a combination of data quality coverage and related tables.

8.3.5 Corrective Measures

The type of corrective measures to apply, if the vegetation data for a specific park fail to meet thematic or positional accuracy requirements, will have to be determined on a case-by-case basis. A failure to meet accuracy specifications may result from improper procedures during the compilation of the map but may also be due to limitations inherent in the vegetation mapping process itself. If excessive error rates result from improper procedures, the mapping process may have to be repeated, starting at the point in the process where the procedural error occurred. This of course is only possible if the procedural error can be isolated. If excessive error rates occur for reasons outside the mapping contractor's control, such as ambiguities in the classification scheme, the classification scheme itself may need to be modified. If such modification is not possible, the accuracy requirements for the given area or class may have to be relaxed.

During prototyping, mapping procedures, classification methods, and accuracy assessment procedures will be exercised. This will help investigators detect possible weaknesses in the classification and mapping process, as well as determine the extent to which accuracy requirements can be met. These tests will be used to refine the mapping and classification process as well as the accuracy assessment procedures and accuracy requirements. It is hoped that these refinements will permit a mapping and classification process that has a high probability of meeting the accuracy requirements to be devised.

8.4 Alternative Methods

The recommended approach for accuracy assessment described in previous sections has four basic precepts:

1. The entire park will be included in the sampling strategy.

2. Observations of vegetation classes will be ground based.

3. Ground sampling techniques will be identical to those used in the initial classification.

4. The number of samples per class will vary with the abundance and size of classified polygons.

This approach is intended to set a rigorous standard while still taking into account cost considerations. Nevertheless, the approach is still very ambitious and assumes a very high level of resource availability. For this reason, it is appropriate to outline some alternatives or variations on the basic approach which take into account the inevitability of logistical problems and limited resources. It is also recognized that there will be instances where the area in question is so small as not to warrant a complex sampling procedure. For some areas, it may be feasible to literally check everything.

Alternative approaches that can be used to reduce costs or overcome accessibility problems fall into two broad categories. First, strategies for sampling less than the entire mapped area of the park need to be considered (subsampling strategies). Second, observational strategies may be varied. In other words, sources of higher accuracy other than ground-based sampling may be used to increase sampling efficiency if the loss of confidence in the identification is manageable.

8.4.1 Subsampling Strategies

Two-stage or subsampling of a large park could proceed in one of two ways. First, the primary sampling units (e.g., a section or block of sections) could be randomly selected until the entire range of class variability or a statistically meaningful number of samples (e.g., 30) are selected. Stratified random sampling of the vegetation classes would then take place within the primary units. This technique would seem most appropriate where the variability in vegetation is more or less uniformly distributed.

A more efficient technique would be to consciously select subsamples that encompass the greatest amount of potential diversity in vegetation. This could proceed by an analysis of biophysical diversity in much the same way that reconnaissance sampling is undertaken (see Section 3.1.4 in Field Sampling Methodology). The subsamples could then be laid out as blocks that capture clusters of diversity or as transects that coincide with known environmental gradients. The latter would serve as a means of detecting subtle patterns of error.

The number of these primary samples would largely be determined by the clustering of diversity in the park as well as available resources. However, enough samples would also have to be selected to take into account important variables that influence the reliability of mapping such as photo quality or an individual interpreter's ability. That is to say each interpreter's work or major fluctuation in photo coverage (e.g., color balance, date) would need a representative sample.

A probability of introducing significant bias exists with this technique, but the actual process of photointerpretation favors the technique as a reliable indicator. Consistency

and accuracy in photointerpretation depends on the identification of a clear and unambiguous relationship between the ground conditions and the photo signatures. If this relationship is not apparent, errors will be propagated across the mapping area. Even limited sampling will identify this problem trend. Other errors or inconsistencies caused by poor technique are more likely to be clustered in the work of a given interpreter. The sampling design can account for this type of error.

8.4.2 Observational Strategies

The high cost of acquiring numerous ground-based samples in very large or inaccessible parks suggests that other means be employed to validate the accuracy of mapped vegetation classes. In circumstances where access problems are limited to a few critical areas, validation crews might be transported by helicopter to sites for observations from the air while hovering. Variations on this approach could involve the use of 35-mm or 70-mm cameras for documentation. Landing the helicopter for reconnaissance could also be used, if feasible. In any case, positions would be fixed using GPS.

This technique offers the advantage of getting very close to the vegetation in addition to rapid transport. However, it is probably not feasible for extensive use. Helicopters, particularly those that might carry large crews, are very expensive to rent or lease. Additionally, the noise levels may not be acceptable. For safety reasons, helicopters will also face restrictions on the working altitude in dry dusty environments.

Airborne Videography

A more promising alternative would be the use of fixed-wing airborne videography. Video data has been used with varying success in a number of natural resource assessment and monitoring projects. (Graham 1993; Marsh et al. 1991: Myhre et al. 1990; Stutte and Stutte 1990). Airborne videography provides several advantages: It is a relatively inexpensive and rapid method for acquiring large-scale imagery. The video tapes are available for viewing and analysis immediately after or even during the flight. Videography provides a less biased sample by freeing the observer from access restrictions; and it provides an extended opportunity for analyses and decision making by experts not able to visit the field sample sites. Noise levels from fixed-wing aircraft are also much less than those from helicopters.

The zoom capabilities of video cameras also offer the advantage of different scale perspectives. Wide-angle views help with understanding the overall landscape while zoomed images may be suitable for identifying individual species. A video camera with 16-mm lens is capable of picture element resolutions of 3.5 feet at 3,000 foot altitude above-ground level (AGL), and 5.8 feet at 5,000 feet AGL (Evans 1992). Portable monitors in the aircraft would offer an additional opportunity to compare the direct observations of the vegetative cover with the video images.

Field testing of this technique should focus on the ability to acquire precise geographic locational information for the images and the limitations of interpreting vegetation classes from videography. Errors are introduced into the georeferencing of video data by the fact that the GPS data received by the aircraft antenna are assumed to be for the video frame center. Excessive tip and tilt of the aircraft during filming can introduce significant error especially for higher altitudes (Bobbe et al. 1993). The relatively slow time for GPS updating when compared to the capture of video frames, also limits positional accuracy. However, some researchers have made innovative use of the time codes that are a standard for each frame of a television picture to extrapolate positions of frames between GPS updates (Graham 1993). The positional accuracy limits of this approach and others should be field tested.

While natural color video is the most likely choice, it is recommended that other portions of the electromagnetic spectrum, such as near-infrared, be tested to determine if they enhance the system's ability to discriminate vegetation types and improve interpretation.

Interpretation keys for use with videography may either be developed using video prints that have been field verified or keys may be based on qualitative descriptions of the vegetation. Both approaches should be tested. An important part of error management for either method would be to restrict the use of any but the most confident sample observations. Those that are in any doubt should be replaced with a ground-based observation or not used.

Aerial surveys with video will be the most efficient when the flight lines are straight with few turns. This will need to be factored into the sampling design so that video transects actually become the subsamples from which the stratified random samples are taken.

Existing Field Plot Data

As stated in Section 3.1, existing field plot data may also be useful for accuracy assessments if they are accurately documented and reflect conditions at the time the imagery was obtained. These data would also be extremely valuable for initial field reconnaissance, which is probably their most beneficial use.

8.4.3 Strategies for Using Alternatives

The decision on when and where to use the alternative procedures will be based on the nature of the park lands and available resources. The major factors to be considered include the size of the park, the accessibility of park lands, and the inherent difficulty of classifying the vegetation. The following discussion addresses the conditions under which the various alternatives would be appropriate.

Size

The parks range in size from less than 1 hectare to over 1 million hectares. The median size is approximately 2,500 hectares; the mean size is approximately 50,000 hectares. This does not include the areas of interest surrounding the parks which might, in some cases, double the area to be mapped. Approximately 30% are less than 500 hectares in size. Another 55% of the parks range from 500 hectares to the approximate mean size of 50,000 hectares, while the remaining 15% increase steadily in size to the largest, which is Yellowstone, at 1,043,330 hectares.

For those parks under the median size of 2,500 hectares, it may be feasible to actually visit every classified unit if the accessibility is good and diversity is relatively low. For parks in the range of 2,500 hectares to the mean size of 50,000 hectares, the basic recommended approach of ground observation sampling for the entire area may be feasible. Parks larger than the mean will most likely be considered for subsampling as well as using airborne videography because of the large number of points that would have to be sampled if the entire area was covered. At this point, these guidelines are intuitive and will have to be considered within the context of a given park unit.

Accessibility

Accessibility will generally be related to the size of the park and the lack of roads and trails. However, access or transportation problems could also be a problem in medium or small-size parks because of extensive wetland areas or sensitive biological preserves. Access may also be limited by surrounding private l ands such as those along some scenic waterways.

Access by helicopter should be considered if the number of hard-to-reach sites is small and of critical concern. If the access problem is widespread, then videography should be considered. If the inaccessible areas are relatively few and are amply represented by similar areas that can be accessed, the more reasonable approach may be to simply extrapolate from the known conditions and not attempt to sample the difficult areas.

Vegetation Complexities

Several factors can contribute to the inherent difficulty of classifying the vegetation of a park. These include subtle environmental gradients, great species diversity, histories of repeated disturbance, or lack of obvious dominants in the community. Extremely complex vegetation may limit the use of airborne videography if identification becomes the issue. However, subsampling as a strategy for large parks will still be applicable even with very complex classification issues if a representative sample of all the variation can be selected with great confidence. Even for moderate size parks, subsampling may be preferable if the complexity of the vegetation results in an extraordinary number of classes.

8.5 Local Options

The degree to which the basic recommendation for sampling will have to be modified for local conditions will have to be evaluated on a case-by-case basis. The use of subsampling, variable scenarios for sampling sizes, and sources of higher accuracy other than ground observations implies that the thematic accuracy of the same class in different parks may be stated with a different confidence interval. Nevertheless, this should be manageable given a complete understanding of the methods and technologies. To reach the required understanding, operational testing of the procedures will be necessary.

9.0 Operational Testing

9.1 Testing the Proposed Method

It is recommended that the proposed methods be tested during the prototyping phase of the project. During this phase, areas in several parks will be classified using the proposed classification scheme. After the map data for the prototype areas are compiled, a prototype accuracy assessment can be conducted using the accuracy assessment procedures described above. The aim of the testing will be twofold:

1. To determine whether the classification and accuracy assessment methods are reasonable, given the nature of the vegetation data being collected. The evaluation will include determining whether the sampling scenarios are reasonable and whether the accuracy requirements can be met with the given classification scheme and sampling methods.

2. To determine whether these methods can be implemented given cost and time constraints.

The following information will be of interest in making these determinations:

- The number of classes typically found in a park and how the number varies with park size. This number will depend on the classification scheme. Information about the number of classes to be expected will give an indication as to the total number of samples that may have to be collected for a park if the current assumptions about frequency and abundance are correct.

- The percentage of area occupied by each class in the prototype areas and the frequency of the polygons that belong to a specific class. This will help determine whether the current definitions of frequency and abundance are reasonable.

- The ease with which the sampling sites can be located in the field. Also, the percentage of preselected sites that have to be dropped in the field because of such factors as inaccessibility or temporal change. Difficulty in locating and accessing field sites and the loss of a large percentage of sampling sites will affect cost, because field visits will need to be repeated.

- The ability to delineate a sampling area equivalent to the MMU once the sample site has

been located. Related to this, the ease with which data can be collected within the MMU and the amount of time required to collect data for a given site: modifications of the shape or area of the sampling unit may have to be made.

- The extent of the modifications investigators need to make to the sampling procedure in the field to deal with special situations. The need to make extensive modifications is a good indication that the overall process should be modified.

- The positional and thematic accuracies obtained for the prototype areas and the main contributing sources of error. Mapping or sampling procedures may need to be improved, or accuracy requirements may need to be modified.

9.2 Refining the Process

The results of the operational testing may be used to refine or change the accuracy assessment procedures. The refinement of current procedures is likely to include the following:

- Modifying the definitions of class abundance in terms of polygon frequency and the area occupied by the class. The modification of the definitions may include options for park-specific definitions.

- Modifying the sample sizes. Depending on the accuracy assessment budget and the total number of classes encountered in a park, it may be possible to raise the minimum number of samples, especially for classes deemed important from an ecological point of view. Of course, if the number of classes is too large to allocate a sufficient number of samples to each class, it may also be necessary to decrease the number of samples.

- Modifying the sampling unit, the sample collection methods, or the methods by which sample sites are selected and positioned in the field. These types of modifications will probably be driven by cost constraints.

- Modifying the accuracy requirements. This will be justified if the main reason that classification accuracy requirements are not met is the classification scheme itself, as opposed to data collection error.

- Modifying sampling procedures to include two-stage sampling and the use of airborne videography as the source of higher accuracy.

10.0 Glossary of Terms and Definitions

Accuracy	Measure of the absence of error. Also a measure of how close an estimate is to its true value.
Binomial Distribution	Probability distribution used to describe discrete random variables.
χ^2 Test	A hypothesis test used for the comparison of variances. χ^2 is an index of dispersion.
CMAS	Circular map accuracy standard.
Confidence Interval	An interval within which an investigator will have a specified level of confidence (typically 90%, 95%, or 99%), the true value of an estimate lies.
Confidence Level	The degree of confidence an investigator has in an estimate. Typically expressed as a percentage (e.g., 90%, 95%, or 99%).
Continuous Random Variable	A random variable with an infinite set of outcomes.
Contingency Table	Table constructed for classifying count data. The entries in the cells show the number of observations falling into a particular category. For accuracy assessment, the table is used to determine the degree of misclassification that has occurred between classes. Also referred to as error matrix, confusion matrix, or misclassification matrix.
DGPS	Differential GPS.
Discrete Random Variable	A random variable with a finite set of outcomes.
Error	In the case of vegetation mapping, the difference between the vegetation class specified at a point <x,y> in the database and the class observed in the field. Errors may be thematic or positional in nature.

Estimate	Value computed from sample data that approximates a population parameter (also point estimate).
GPS	Global Positioning System.
Hypothesis Test	A type of statistical inference where an investigator makes a decision concerning the hypothesized (expected) value of a random variable (Kleinbaum and Kupper 1978). The results of the test typically permit an investigator to conclude whether the estimated value is the same or different than the hypothesized value.
Kappa Index	Index that corrects for chance agreement in a contingency matrix.
Inclusion	Distinct vegetation community whose extent is less than the minimum mapping unit.
Mixing Ratio	Community composition within a particular polygon in situations where class assignment is ambiguous due to the presence of transition zones or inclusions. For example, the mixing ratio of a particular polygon may be 80% class A and 20% class B.
NAVSTAR	Navigation Satellite Timing and Ranging.
NMAS	National Map Accuracy Standards.
NPS	National Park Service.
Normal Distribution	Bell-shaped probability distribution.
Parameter	A numerical population descriptor.
Point Estimate	(See estimate.)
Population	The elements to be studied or described in a given experiment. In the case of accuracy assessment, the population represents the totality of all errors.

Positional Error	Discrepancy between the coordinate location of a point in the database and the coordinate of the same point in a source of higher accuracy. Positional error generally applies to well-defined points only and is expressed as a root mean square error (RMSE).
Precision	The degree of conformity among a set of observations. As such, precision is a measure of dispersion of the probability distribution associated with a measurement and expresses the degree of repeatability of a measurement.
Producers' Accuracy	Extracted from the contingency matrix. Probability that a reference sample has been classified correctly. An incorrectly classified reference sample is also referred to as an error of omission.
Random Variable	A variable whose values are numerical events that cannot be predicted with certainty. Random variables may be continuous or discrete.
RMSE	Root mean square error. A measure of positional accuracy that encompasses both the effects of bias and random error.
t-test	Hypothesis test used for the comparison of means.
Thematic Error	Discrepancy between the vegetation class at a particular point in the database and the class observed at the same coordinate location in the field. Generally expressed as a proportion or percentage.
Users' Accuracy	Extracted from the contingency matrix. Probability that a sample from the classified data actually represents that category on the ground. A sample with a different category on the ground than in the classified data is also referred to as an error of commission.
USGS	United States Geological Survey.

11.0 Literature Cited

Adkins, Kirk F. and C. J. Merry. 1994. Accuracy Assessment of Elevation Data Sets Using the Global Positioning System *American Society for Photogrammetry and Remote Sensing,* 60 (2).

American Society of Photogrammetry and Remote Sensing. 1989. ASPRS Interim Accuracy Standards for Large Scale Maps. *Photogrammetric Engineering and Remote Sensing,* 55 (7): 1038–1040.

Aronoff, S. 1985. The Minimum Accuracy Value as an Index of Classification Accuracy. *Photogrammetric Engineering and Remote Sensing,* 51 (1): 99–111.

Blakemore, M. 1984. Generalisation and Error in Spatial Data Bases. *Cartographica,* 21, (2 and 3): 131–139.

Bobbe, T. 1992. Real-Time Differential GPS for Aerial Surveying and Remote Sensing. *GPS World,* 3 (7).

Bobbe, Tom, Dennis Reed, and Jim Schramek. 1993. Georeferenced Airborne Video Imagery. *Journal of Forestry*, August: 34–37.

Colwell, Robert N., ed. American Society of Photogrammetry. 1983. *Manual of Remote Sensing*, Vol. 2, 1601–1605.

Chrisman, N. R. 1984. The Role of Quality Information in the Long-Term Functioning of a Geographic Information System. *Cartographica,* 21 (2 and 3): 79–87.

Congalton, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37: 35–46.

Congalton, R. G. and K. Green. 1993. A Practical Look at the Sources of Confusion in Error Matrix Generation. *Photogrammetric Engineering and Remote Sensing,* 39 (5): 641–644.

Congalton, R. G. and R. A. Mead. 1983. A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing,* 49 (1): 69–74.

Congalton, R. G., R. G. Oderwald, and R. A. Mead. 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering and Remote Sensing*, 49 (12): 1671–1678.

Evans, David L. 1992. Using GPS to Evaluate Aerial Video Missions. *GPS World*, 3 (7).

Evans, David L. 1992. Using GPS to Evaluate Aerial Video Missions. *GPS World*, July/August: 24–29.

Federal Geodetic Control Committee. 1988. *Geometric Geodetic Accuracy Standards and Specifications for Using GPS Relative Positioning Techniques.* Reprinted with corrections: 1 August 1989.

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

Ginevan, M. E. 1979. Testing Land-Use Map Accuracy: Another Look. *Photogrammetric Engineering and Remote Sensing*, 45 (10): 1371–1377.

Goodchild, M. F., Biging, G. S., Congalton, R. G., Langley, P. G., Chrisman N. R., and F. W. Davis. 1994. *Final Report of the Accuracy Assessment Task Force.* California Assembly Bill AB1580, Santa Barbara: University of California, National Center for Geographic Information and Analysis (NCGIA).

Goodchild, M. F., G. Q. Sun, and S. Yang. 1992. Development and test of an error model for categorical data. *International Journal of Geographical Information Systems*, 6 (2): 87–104.

Gopal, S. and C. Woodcock. 1994. Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets. *Photogrammetric Engineering and Remote Sensing*, 60 (2): 181–188.

Graham, Lee A. 1993. Airborne Video for Near-Real-Time Vegetation Mapping. *Journal of Forestry*, August: 28–32.

Kleinbaum, D. G. and L. L. Kupper. 1978. *Applied Regression Analysis and Other Multivariable Methods*. Boston: Duxbury Press.

Lillesand, T. M. and R. W. Kiefer. 1979. *Remote Sensing and Image Interpretation*. New York: John Wiley and Sons.

Marsh, S. E., J. L. Walsh, C. T. Lee, and L. A. Graham. 1991. Multispectral Analysis of Hazardous Waste Sites Through the Use of Bi-spectral Video Remote Sensing System and Standard CIR Photography. *Photogrammetric Engineering and Remote Sensing*. 57: 1221–1226.

Merchant, D. C. 1985. Accuracy Specification for Large Scale Maps, *Photogrammetric Engineering and Remote Sensing*, 51 (2): pp. 195–199.

Merchant, D. C. 1987. Spatial Accuracy Specification for Large Scale Maps. *Photogrammetric Engineering and Remote Sensing*, 53 (7): 958–961.

Mikhail, E. M. and F. Ackerman. 1976. *Observations and Least Squares*. New York: University Press of America.

Mueller-Dombois, D. and H. Ellenberg. 1974. *Aims and Methods of Vegetation Ecology*. New York: John Wiley and Sons.

Myhre, R. J., L. A. Graham, and C. W. Sumpter. 1990. Airborne Video Technology: Protecting Natural Resources with Remote Sensing. *Proceedings, 3rd USDA Forest Service Remote Sensing Applications Conference*. American Society of Photogrammetry and Remote Sensing. Tucson, Arizona, pp. 293–298.

National Park Service. 1993. *NPS Geographic Information Systems Sourcebook.* Chapter 3 of Spatial Data Characteristics.

Rosenfield, G. H., and K. Fitzpatrick-Lins. 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 52 (2): pp. 223–227.

Skidmore, A. K. and B. J. Turner. 1992. Map Accuracy Assessment Using Line Intersect Sampling. *Photogrammetric Engineering and Remote Sensing*, 58 (10): 1453–1457.

Snedecor, G. W. and W. G. Cochran. 1976. *Statistical Methods.* Ames, Iowa: The Iowa State University Press.

Stoms, D., Davis F., Cogan, C. and K. Cassidy. 1994. Assessing Vegetation/Land Cover Map Accuracy for Gap Analysis, *Draft Handbook*, University of California at Santa Barbara, Biogeography Laboratory.

Story, M., and R. G. Congalton. 1986. Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, 52 (3): 397–399.

Stutte, C. A. and G. W. Stutte. 1990. Quantification of Field Crop Stress Areas Using Videography. *Proceedings, 123 Biennial Workshop on Color Aerial Photography and Videography*. American Society for Photogrammetry and Remote Sensing., Sparks, Nevada, pp. 88–93.

Taylor, P. J. 1977. *Quantitative Methods in Geography—An Introduction to Spatial Analysis.* Boston: Houghten Mifflin Company.

U.S. Army Corps of Engineers. 1991. *NAVSTAR Global Positioning System Surveying.*

U.S. Army Topological Engineering Center. 1991. *GPS For The Hydrographic Surveyor: Meters Today, Centimeters Tomorrow.*

United States National Map Accuracy Standards. 1947. U.S. Bureau of the Budget.

Van Genderen, J. L. and B. F. Lock. 1977. Testing Land-Use Map Accuracy. *Photogrammetric Engineering and Remote Sensing*, 43 (9): 1135–1137.

12.0 List of Authors and Contributors

Mirjam Stadelmann*
Anthony Curtis*
Randy Vaughan*
Marian Bailey
Charles Convis
Environmental Systems Research Institute, Inc. (ESRI)
380 New York Street
Redlands, CA 92373

Michael Goodchild*
Frank Davis
National Center for Geographic Information Analysis
3510 Phelps Hall
University of California at Santa Barbara
Santa Barbara, CA 93106

Xiaojun Li
Kathy Goodin
Dennis Grossman
The Nature Conservancy
1815 North Lynn Street
Arlington, VA 22207

***Contributing authors**