National Park Service
U.S. Department of the Interior

Natural Resource Program Center

Thematic Accuracy Assessment Procedures

National Park Service Vegetation Inventory, Version 2.0

Natural Resource Report NPS/NRPC/NRR—2010/204

ON THE COVER
National Park Service staff record field data for a thematic accuracy assessment of a vegetation map at Joshua Tree National Park. National Park Service photograph by Chris Lea.

Thematic Accuracy Assessment Procedures

National Park Service Vegetation Inventory, Version 2.0

Natural Resource Report NPS/NRPC/NRR—2010/204

Chris Lea

National Park Service
Vegetation Inventory
P.O. Box 25287 (Academy Place)
Denver, CO 80225

Anthony C. Curtis

National Park Service
Southeast Coast Inventory and Monitoring Network
P.O. Box 806
Saint Marys, GA 31558

June 2010

U.S. Department of the Interior
National Park Service
Natural Resource Program Center
Fort Collins, Colorado

The National Park Service, Natural Resource Program Center publishes a range of reports that address natural resource topics of interest and applicability to a broad audience in the National Park Service and others in natural resource management, including scientists, conservation and environmental constituencies, and the public.

Natural Resource Reports are the designated medium for disseminating high priority, current natural resource information with managerial application. The targets a general, diverse audience, and may contain NPS policy considerations or address sensitive issues of management applicability.

All manuscripts in the series receive the appropriate level of peer review to ensure that the information is scientifically credible, technically accurate, appropriately written for the intended audience, and designed and published in a professional manner.

This report received formal peer review by subject-matter experts who were not directly involved in the collection, analysis, or reporting of the data, and whose background and expertise put them on par technically and scientifically with the authors of the information.

Views, statements, findings, conclusions, recommendations, and data in this report are those of the author(s) and do not necessarily reflect views and policies of the National Park Service, U.S. Department of the Interior. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the National Park Service.

This report is available from the National Park Service Vegetation Inventory website (http://science.nature.nps.gov/im/inventory/veg/index.cfm) and from the Natural Resource Publications Management website (http://www.nature.nps.gov/publications/NRPM).
Please cite this publication as:

Lea, C. and A. C. Curtis. 2010. Thematic accuracy assessment procedures: National Park Service Vegetation Inventory, version 2.0. Natural Resource Report NPS/2010/NRR—2010/204. National Park Service, Fort Collins, Colorado.

NPS 999/102741, June 2010

Contents

Page

Contents iii

Figures vii

Tables ix

Appendices and Exhibits xi

Executive Summary xiii

Acknowledgments xv

1.0 Introduction 1

1.1 Thematic Accuracy Assessment, NPS Vegetation Inventory, 1994-2009 1

1.2 Summary of Changes from 1994 Guidance 5

1.3 Key Requirements and Best Practices 6

1.4 Work Flow 9

2.0 Sampling Design 15

2.1 Sampling Design and Data Collection Objectives 15

2.2 Observation Area (Minimum Mapping Unit) Size 15

2.3 Inference Area Selection 18

2.3.1 Included and Excluded Map Classes 18

2.3.2 Areas with Training Sites 19

2.3.3 Adjustments for Access Costs 19

2.3.4 Adjustments to Achieve Sample Data Homogeneity in Observations 21

2.3.5 Adjustments for Independence Among Observations 22

2.4 Sampling Methods 24

2.5 Determining Sample Sizes 25

2.5.1 Determining Acceptable Levels of Error and Confidence 25

2.5.2 Number of Observations per Map Class *26*

2.6 Geographic Information Systems Design Considerations 28

3.0 Field Methods (Response Design) 29

3.1 Considerations that Affect the Selection of a Source of Higher Accuracy 29

3.2 Field Methodology 33

3.2.1 Field Observer Skill Level Considerations *33*

3.2.2 Safety and Efficiency *35*

3.2.3 Basic Navigation *36*

3.2.4 Field Navigation Supplements and Retaining Independence from Mapping *37*

3.2.5 Determination of Vegetation Type in the Field *37*

3.2.6 Moving or Reshaping Observation Area *38*

3.2.7 Recording Field Position (GPS methods) *40*

3.2.8 Reporting Field Data *41*

3.2.9 Missed Sites *41*

4.0 Data Analysis 43

4.1 Accuracy Assessment Data 44

4.2 The Sample Contingency Table 44

4.3 The Population Contingency Table 46

4.4 Definitions of Measures of Overall Accuracy 47

4.5 Users' and Producers' Accuracies Measures for Individual Map Classes 47

4.6 Computations of Accuracy Measures under Stratified Random Sampling 48

4.6.1 Computation of Measures of Users' Accuracies *48*

4.6.2 Computation of Measures of Producers' Accuracies *49*

4.6.3 Computation of Measures of Overall Accuracy *51*

4.7 Lumped Classes 52

4.8 Design and Formatting for Contingency Tables 53

5.0 Reporting Requirements 57

Literature Cited 59

Figures

Page

Figure 1. Work Flow of operational phases of an NPS Vegetation Inventory Project. 13

Figure 2. Relationship between Sample Size (N), Confidence Interval Width, and Confidence Level for two different values of the point estimate of accuracy ($\hat{p}$). 26

Figures 3A-D. Examples of Providing Field Maps Suitable as Navigation Aids for Thematic Accuracy Assessment, Thomas Stone National Historical Site (Maryland). 72

Figures 4A-D. Examples of Providing Field Maps Suitable as Navigation Aids for Thematic Accuracy Assessment, Joshua Tree National Park (California), USGS Conejo Wells 7.5 minute quadrangle..... 73

Figures 5A-B. Examples of Providing Field Maps Suitable as Navigation Aids for Thematic Accuracy Assessment, Joshua Tree National Park (California), USGS Yucca Valley 7.5 minute quadrangle..... 74

Tables

Page

Table 1. Comparison of field observation activities in the National Park Service Vegetation Inventory. 12

Table 2. Range of suitable minimum mapping unit (MMU) sizes for types of vegetation. 17

Table 3. Radius lengths for suitable observation area sizes. 17

Table 4. Standard sample size allocations for NPS Vegetation Inventory thematic accuracy assessment, based on map class area. 27

Table 5. Comparison of five possible methods for sources of higher accuracy for thematic accuracy assessment, using three major evaluation criteria. 32

Table 6. Notation used in Chapter 4. 43

Table 7. Sample contingency table for five sample and reference data vegetation classes.. 45

Table 8: Numerical example of a sample contingency table for five hypothetical sample data and reference data vegetation classes. 46

Table 9: Population contingency table for five hypothetical sample data and reference data vegetation classes.. 47

Table 10: Values of z_α for two-sided confidence intervals at selected confidence levels 49

Table 11. Calculation of users' accuracy point estimate and 90% confidence intervals for three forest types at Delaware Water Gap National Recreation Area aggregated (lumped) to a coarser thematic class. 52

Table 12: Numerical example of a sample contingency table for seven hypothetical sample data classes, represented by nine hypothetical reference data classes. 54

Table 13: Numerical example of a population contingency table for the five hypothetical sample data and reference data vegetation classes of Table 8. 55

Table 14. Accuracy assessment observation area size and results of access buffer application for individual Shenandoah National Park vegetation map classes. 69

Table 15. Sample contingency table for thematic accuracy assessment, Thomas Stone National Historic Site. 89

Table 16. Population contingency table for thematic accuracy assessment, Thomas Stone National Historic Site. 90

Appendices and Exhibits

Page

Appendix A. Minimum Mapping Unit (Point) Versus Polygon Sampling Designs for Thematic Accuracy Assessment 63

Appendix B. Sampling Design for Thematic Accuracy Assessment of Shenandoah National Park Vegetation Map 65

Appendix C. Examples of Providing Field Maps Suitable as Navigation Aids for Thematic Accuracy Assessment 71

Appendix D. How to Represent Accuracy Assessment Data when Map Corrections are Made 75

Appendix E. Glossary of Terms, Definitions, and Acronyms 79

Exhibit F. Examples of Contingency Tables 89

Exhibit H. Decision Tree for Relocating or Reshaping a Thematic Accuracy Assessment Reference Data Observation 91

Exhibit I: Accuracy Assessment Form 93

Executive Summary

This guidance is a revision and update of the 1994 guidance on thematic accuracy assessment for the NPS Vegetation Inventory (Environmental Systems Research Institute et al. 1994). The revision incorporates the sampling design and analysis principles described by the 1994 version as a sound starting framework and by augments them with practical experience gained by the NPS and its cooperators and contractors at more than 100 NPS parks over 15 years. NPS staff informally interviewed many individuals in and outside the NPS who have been involved in project production and oversight. We evaluated specifically thematic accuracy assessment practices in the field and in analysis at more than 30 NPS parks. As a result, this guidance is designed to be scientifically sound, but also practical to implement. While it specifically addresses NPS objectives, it may also be found useful as guidance for other organizations that assess the accuracy of vegetation maps.

Several formal and informal program reviews concluded that, while rigorous thematic accuracy assessments are a strength of the NPS Vegetation Inventory, the process has had some inefficiencies and/or objectives that were not primary NPS goals. Additionally, the 1994 guidance, which was written in the absence of operational experience within the National Park Service, acknowledged that operational testing and evaluation should be conducted and adjustments to the process be made, as needed. While the scientific (remote sensing, sampling, and statistical) literature provides guidance on sampling design and analysis, response design (field) methods that are specific to the vegetation science discipline are less often addressed. Many of the specific methods practiced within the NPS Vegetation Inventory that were reviewed in preparation of the 2010 guidance appear not to have been documented in the literature. A number of them evidently are *de novo* practices developed out of need by practitioners.

One program change that is beyond the scope of these guidelines, but is reflected in this revision, has been to better and more narrowly define the objectives of the thematic accuracy assessment for the NPS Vegetation Inventory and to eliminate or reassign activities better defined as production, rather than assessment, functions (see http://science.nature.nps.gov/im/inventory/veg/index.cfm).

The primary objectives of this guidance are to (1) provide the conceptual framework for thematic accuracy assessment within the NPS Vegetation Inventory, (2) describe minimal requirements for the process, and (3) give “best practices” guidance, including an acceptable range of procedural variations as alternative procedures. While the guidance outlines a basic procedure that is statistically rigorous and consistent with traditional methodologies, it is also recognized that individual site constraints may require reasonable variations in sampling design and data collection.

The guidance is organized to facilitate practical use in NPS Vegetation Inventory projects. Chapter 1.0 introduces the principles, terms, and work flow of thematic accuracy assessment to assist project managers in operational planning and budgeting. Each chapter from 2.0 to 5.0 describes requirements of the NPS Vegetation Inventory and suggested practices for one of the four major operational phases of thematic accuracy assessment: sampling design, field methods (response design), data analysis, and reporting. Titled sections and subsections within these chapters will assist project investigators with addressing specific technical and operational issues. More detailed “how to” information and examples from NPS projects on some issues are presented in appendices and exhibits at the end of the guidance.

Acknowledgments

Debbie Johnson (Aerial Information Systems [AIS]), John Menke (AIS), Ed Reyes (AIS), Arin Glass (AIS), Jim Von Loh (CoganTech), Dan Cogan (CoganTech), Jennifer Dieck (U.S. Geological Survey [USGS]), Kevin Hop (USGS), Sara Lubinski (USGS), Peggy Moore (USGS), Ann Cully (National Park Service [NPS]), Stephanie Perles (NPS), Mike Story (NPS), Karl Brown (NPS), and Tammy Cook (NPS) reviewed early drafts of this report and provided comments that greatly improved its quality. The knowledge of Tom Philippi (NPS) was especially helpful in the critical review and revision of computational methods. Steve Fancy (NPS) served as peer review manager for the report.

We thank the authors of the first version of this guidance, Mirjam Stadelmann (Environmental Systems Research Institute), Randy Vaughan (Environmental Systems Research Institute), and Michael Goodchild (University of California, Santa Barbara), for developing a solid foundation for thematic accuracy assessment upon which this document continues.

We recognize the many NPS managers, field staff, cooperators, and contractors who participated in the NPS Vegetation Inventory from 1994 to 2010 and whose efforts, struggles, experience sharing, and constructive criticism allowed us to improve this guidance.

Funding to support this work was provided by the National Park Service Inventory and Monitoring Program and Biological Resources Management Division.

1.0 Introduction

1.1 Thematic Accuracy Assessment, NPS Vegetation Inventory, 1994-2009

The objectives of the National Park Service (NPS) Vegetation Inventory (formerly known as the U.S. Geological Survey – National Park Service Vegetation Mapping Program) (http://science.nature.nps.gov/im/inventory/veg/index.cfm) are to classify vegetation as ecological community types in each of more than 250 NPS Inventory and Monitoring units ("parks") in the United States outside of Alaska and to map vegetation in each park using the park-specific vegetation classes. The primary data to be used for the vegetation classification are vegetation plot data that have been collected at or near the park. The primary data to be used for mapping is remote sensing imagery, supplemented by ancillary field data used for formal or informal modeling. The final products are descriptions of the classified vegetation, a digital spatial database (map) of the vegetation, and quality control data. An essential part of the quality control latter products is the data and report on a thematic accuracy assessment of each vegetation map class in the spatial database. Accuracy assessment is important because estimates of thematic errors in the data will allow data users to assess data suitability for a particular application.

Thematic accuracy assessment activities for the NPS Vegetation Inventory originally followed the guidance of Environmental System Research Institute et al. (1994), which specified a requirement of a minimum accuracy of 80% (for the point estimate of the sample mean) for the accuracy of every class mapped and assessed. Additionally, the ecological classification guidelines of The Nature Conservancy and Environmental System Research Institute (1994) specified that the vegetation classes would be mapped at the National Vegetation Classification (NVC) level of alliance, or, whenever possible, association (Federal Geographic Data Committee 1997, 2008, NatureServe 2009).

Three different reviews of the NPS Vegetation Inventory (National Park Service 1996, 1998, Moeller et al. 1998, U.S. Geological Survey 1999) examined accuracy assessment procedures and operations. In general, the reviews noted that thematic accuracy assessment was an important and rigorous component of the program, but that the costs of thematic accuracy assessment were of concern. In response to these reviews, the U. S. Geological Survey (1999) recommended that "the accuracy standard should be reevaluated by the technical leadership of the program. While the accuracy assessment standard and protocol is a critical part of the program, the specific standards criteria and all elements of the accuracy assessment protocol may not be necessary considering the planned applications of the data set. Thus, accuracy standards and the assessment protocol should be reevaluated with a goal of reducing cost and accelerating the completion of the program."

From 2003 to 2009, the National Park Service conducted informal "hands on" reviews of the procedures of Environmental Systems Research Institute et al. (1994). We examined thematic accuracy assessment procedures, operations, results, and products produced by cooperators and contractors at more than 30 program projects and evaluated how these followed the 1994 guidance. We interviewed a number of project practitioners. We provided direct oversight and guidance of thematic accuracy assessment operations at seven NPS units (Assateague Island National Seashore (Maryland-Virginia), Grant-Kohrs Ranch National Historic Site (Montana), Little Bighorn Battlefield National Monument (Montana), Joshua Tree National Park

(California), Shenandoah National Park (Virginia), Thomas Stone National Historic Site (Maryland), and Vicksburg National Military Park (Mississippi). Finally, as a part of a study to examine of vegetation classification issues throughout the United States (Lea 2008), we observed and assessed common problems that are encountered in field observation methodology while applying ecological field keys to the evaluation of more than 2,100 individual 0.5 hectare sites in 38 different vegetation stands at 20 different NPS park. These parks included Acadia National Park (Maine), Black Canyon of the Gunnison National Park (Colorado), Cape Cod National Seashore (Massachusetts), Congaree National Park (South Carolina), Cumberland Gap National Historical Park (Virginia-Kentucky-Tennessee), Delaware Water Gap National Recreation Area (New Jersey-Pennsylvania), Glacier National Park (Montana), Grand Teton National Park (Wyoming), Great Smoky Mountains National Park (North Carolina-Tennessee), Isle Royale National Park (Michigan), Joshua Tree National Park (California), New River Gorge National River (West Virginia), Ozark National Scenic Riverways (Missouri), Rocky Mountain National Park (Colorado), Sequoia National Park (California), Shenandoah National Park (Virginia), Voyageurs National Park (Minnesota), Walnut Canyon National Monument (Arizona), Yosemite National Park (California), and Zion National Park (Utah).

Two major issues with the ecological classification scheme that formed the basis of vegetation mapping in the NPS Vegetation Inventory became apparent. These issues certainly affected how accuracy assessments were being conducted, but were apparently unknown to and not addressed by previous reviews.

First, the accuracy requirement of a minimum of 80% for every class at the thematic (ecological) resolution of alliance or association (as these units are currently recognized by the NVC) proved to be difficult or impossible to achieve. With the exception of very small NPS units, even an overall (pooled, for all map classes) accuracy rate of 80% at these resolutions seldom was not consistently achieved at these resolutions and was not likely to be economically practical for the NPS Vegetation Inventory. The NVC was new at the time and the challenge of applying the most finely resolved levels to mapping was not well understood. This led to a gap between expectations and results. As an additional complicating factor, experience with and subsequent examination of NVC thematic units at the association and alliance level suggested that these terms often were being applied to different levels of thematic and ecological resolution in different regions of the United States (Lea 2008, Lea 2009). In the hindsight of this experience, it appears that some original assumptions about the NVC and some predictions about its mapping performance (The Nature Conservancy and Environmental System Research Institute 1994) were probably not reasonable, given the limitations in funding, time, and expertise available to the NPS Vegetation Inventory. While these issues often affected and were revealed in thematic accuracy assessment work within the NPS Vegetation Inventory, they cannot be addressed through accuracy assessment procedures.

Another issue that became clear was that mapping to the thematically coarser (higher) physiognomic levels of the NVC hierarchy as it existed in 1994 was of little help in resolving difficulties in mapping ecologically meaningful thematic units accurately. Higher levels of the original NVC hierarchy were populated by thematic classes that proved to be overly based on vegetation structure and to be often ecologically arbitrary (e.g., Formation and above in the sense of Federal Geographic Data Committee (1997) and Grossman et al. (1998), rather than in the sense of Federal Geographic Data Committee (2008)). For example, loblolly pine forest types of

the southeastern United States coastal plain and lodgepole pine forest types of the Rocky Mountains were grouped together at the relatively fine level of the [1997] NVC Formation, while floristically similar and spatially intergrading spruce-fir forests and spruce-fir-aspen forests in the Rocky Mountains were separated at the relatively coarse level of the [1997] NVC Subclass (NatureServe 2009). The structural and other criteria selected for defining these higher level units proved to be less helpful to mapping needs than anticipated. Difficulty in applying the upper levels of the 1997 NVC in both ecological classification and in mapping ultimately led to a considerable revision of the structure of the NVC (Federal Geographic Data Committee 2008) in order to introduce more ecologically meaningful middle levels. As with the issues with the finer levels of vegetation classification, these issues were outside the scope of NPS accuracy assessment operations, but certainly affected the realization of accuracy goals and drove accuracy assessment methodology.

The NPS, USGS, and other Vegetation Inventory partners were slow to recognize these issues and to address them in earlier project phases of ecological classification and map production. As a result, several well-intentioned *ad hoc* procedures of addressing them in the accuracy assessment phase became commonplace for projects. These proved to be difficult for the program to sustain.

Because many NVC associations and alliances were found to be difficult to map at the originally specified individual map class accuracy rates, particularly in the western United States, many thematically fine map classes were summarily lumped together to achieve the prescribed accuracy requirements, often to *ad hoc* project-specific and non-standard classes. This practice tended to defeat the advantages of mapping to a standard that were advocated by The Nature Conservancy and Environmental Systems Research Institute (1994). On the other hand, the thematically finer map classes that were originally attempted in mapping often represented (1) useful thematic (ecological) resolution and adequately accurate (if not 80%) spatial information to the user that the thematically coarser classes could not provide and (2) considerable investment in classification and mapping effort by the NPS. This situation often produced a "disconnect" within individual projects between the attribute information provided by the ecological classification function and the spatial information provided by the mapping function. The situation was further exacerbated by the fact that these two functions were usually provided by different organizations.

Thematic accuracy assessment campaigns often collected substantial amounts of data in the field, in order to allow *post hoc* review of field calls by the production team and to change field calls in order to increase accuracy more toward levels prescribed by the 1994 guidelines. This approach had several disadvantages. It proved to be inefficient to commit to a major field campaign (usually, 30 observations per map class) and to collect comprehensive data at each site only to discover basic errors in the ecological classification, the interpretive tools (field keys), and/or the mapping model. These issues might have been addressed more efficiently during the production activities of ecological classification and mapping. Another operational problem with this practice is that, although collection of sufficient field data to allow *post hoc* review by experts in the office or lab may indeed report higher map accuracy (i.e., the concurrence rate between map classes and field observations), it can be a methodologically cryptic process that impairs the clear understanding of the accuracy assessment results by a user. A major cost and efficiency consideration was that making an "office" determination of type requires substantial field

vegetation data collection effort, (rather than simple determination of type by a qualified user in the field using only the interpretive tools provided as products), and expertise that will not be available to users once the project funding has ceased. Finally, “correction” of field observations was occasionally guided to correspond with map class determinations. This practice violates a central premise that the accuracy assessment results should be as independent of the evaluated map data as possible.

Fuzzy sets theory (Gopal and Woodcock 1994) was often used to report higher (more acceptable) accuracy rates, taking into account that some mapping errors are more understandable than others. While this exercise may be helpful to map producers, a drawback of this approach for project evaluators is that there is no nationally consistent fuzzy sets standard. Each project developed its own criteria for the various levels of “correctness” of a field observation to map class match. This practice made it difficult to assess the quality of an individual project against comparable projects, a concern expressed by the 1994 guidelines (Environmental Systems Research Institute et al. 1994). A further drawback to this approach for users that was also expressed in the 1994 guidelines is that the fuzzy set criteria schemes reflected the perspective and values of producers of maps and ecological classifications and might not reflect the needs of potential users. Where the most thematically resolved map classes do not suffice for an application, the map user may use error rates in a contingency table to derive his/her own “fuzzy sets” or map class hierarchies in order to derive aggregated map classes that reflect user application needs; the criteria applied by the production team may be irrelevant to many of these needs. With the advent of more ecologically meaningful middle level units to the NVC hierarchy (Federal Geographic Data Committee 2008), it should become more possible to develop a more standardized scheme of fuzzy sets. Map class accuracies then might be reported at multiple levels in the NVC hierarchy that are both ecologically meaningful and also reasonably interpretable as a standard means of accuracy assessment at multiple levels of thematic resolution.

Three basic components of a thematic accuracy assessment are the sampling design, the response design, and estimation and analysis (Stehman and Czaplewski 1998). In evaluating these components, as provided for by of the 1994 guidance (Environmental Systems Research Institute et al. 1994) and as practiced within the NPS Vegetation Inventory in the subsequent 15 years, two general conclusions may be made:

(1) The 1994 guidance addressed sampling design and estimation and analysis well, and relied on well-documented and published methods. Published guidance in the remote sensing literature is generally available for these two components, which generally are not specific to the scientific discipline that is employed to create a mapping theme. Thus, this 2010 guidance follows mostly the same methodological approach as that in the 1994 guidance, in the components of sampling design and estimation and analysis. In addition, the 2010 guidance introduces some additional recognized published methods not addressed in 1994, especially those for analysis of stratified data and present more information on tactical methods (e.g. Geographic Information Systems) for sampling design gained from practical experience.

(2) In contrast to methods that apply to the other components, methods for response design (field assessment) are primarily tactical and specific to the methodologies of classification of the theme being assessed (in this case, vegetation science) rather than to general published principles of

remote sensing, sampling, and statistical analysis. In addition, the question of the level of methodological rigor that is required to produce an acceptable level of response design accuracy is heavily dependent upon practitioner objectives, which will invariably include cost/benefit considerations. Therefore, the published literature in remote sensing, sampling, and analysis seldom prescribes methods for the response design component and leaves such guidance to subject-matter experts in the discipline of the theme being assessed and to organization-specific objectives. With no program history, the 1994 guidance offered very little specific methodological guidance for the response design and, in fact, recommended a measure of operational testing, assessment, and revision, as needed (Environmental Systems Research Institute et al. 1994). As a result, individual NPS projects were largely left to address response design (field) methods in a somewhat *ad hoc* manner, and a range of common conventions for collecting and preparing field data for analysis emerged from these individual, but cumulative, project experiences. While each of these *ad hoc* methodologies has incorporated some degree of methodological rigor, they had not been assessed against the cost/benefit concerns raised by the program evaluations (National Park Service 1996, 1998, Moeller et al. 1998, U.S. Geological Survey 1999). Therefore, two goals of the 2010 guidance are (1) to summarize various response design (field method) tactics that have been used in the NPS program as a benefit to other organizations contemplating thematic accuracy assessment of vegetation maps and (2) to give guidance on the most cost-effective practices identified for the NPS Vegetation Inventory.

1.2 Summary of Changes from 1994 Guidance

In general, the guiding principles of the 1994 thematic accuracy assessment methods used are sound. However, they must be re-interpreted in the context of the practical experience that has been developed since 1994 and that was lacking at that time, not only within the NPS Vegetation Inventory, but within the community ecology and mapping disciplines in the United States. In response to the recommendations of past reviews, this revised guidance seeks to:

1. Address only thematic, rather than positional, accuracy. The 1994 version addressed both, but, since it was concerned mostly with thematic accuracy, this version is considered a second edition. Positional accuracy assessment guidance for the NPS Vegetation Inventory remains that prescribed by Environmental Systems Research Institute et al. (1994).

2. Reinforce that thematic accuracy assessment is primarily a user-oriented quality control objective.

3. Differentiate (and define) ground-truthing activities that are internal to the mapping process and whose purpose is to *improve* accuracy (e.g., calibration, verification, and validation) from those activities that are external to the mapping process and are concerned with *assessing* accuracy (e.g., validation and accuracy assessment). Allow reasonable opportunities for the former to occur in the ecological classification and the mapping processes, in order to eliminate incentive to enlist the accuracy assessment process toward meeting these needs. For more information on these processes see Table 1, Figure 1, Appendix E and http://science.nature.nps.gov/im/inventory/veg/index.cfm.

4. Define the most appropriate source of higher accuracy to be used for reference (field) observation data against which to compare the map data, taking into account the additional needs of cost and user relevance, along with accuracy.

5. Reinforce the differences between minimum mapping unit based and polygon based designs in accuracy assessments and that the former approach is required in the NPS Vegetation Inventory. See Appendix A.

6. Improve and make more efficient individual map class sample size allocation, given the increased understanding of the nature of vegetation distribution within National Park units.

7. Where appropriate, allow flexibility in minimum mapping unit sizes, given the increased understanding of vegetation spatial scales within parks and the scale of parks themselves.

8. Provide specific procedural guidance on sampling design.

9. Provide specific procedural guidance on response design (field methodology).

10. Provide updated procedural guidance on analysis (computational methods).

1.3 Key Requirements and Best Practices

Producers of ecological classifications and maps, project managers, and accuracy assessment practitioners for the NPS Vegetation Inventory should be familiar with the following key requirements (RQ) and best practices (BP) of the program.

(RQ) Since the spatial component (a Geographic Information System layer) of the vegetation database will be compiled as a series of park-specific projects, a separate thematic accuracy assessment will be conducted for each project. (BP) Single projects will usually be limited to individual parks, but, in some cases, multiple closely situated or contiguous parks will be mapped as a single project by the same team of producers.

(RQ) Thematic accuracy will be assessed using field observation by a qualified observer as the reference source of higher accuracy to which the map is compared.

(BP) The primary purpose of the thematic accuracy assessment is to inform users of the limitations of the individual vegetation map classes and of the relationship of the errors (confusion) between classes. While evaluation of overall map (project) accuracy against a minimum threshold is also often required in order to assess project performance, this need is more quickly addressed in the earlier and less extensive process of map validation (see Table 1, Figure 1, Appendix E, and http://science.nature.nps.gov/im/inventory/veg/index.cfm).

(RQ) As a quality control process that is intended to be independent from and external to the mapping process (e.g., Thomas et al. 2004), the thematic accuracy assessment is undertaken after work on the map and associated products (ecological classification and field keys) has ceased.

(RQ) As a process independent from mapping process, a thematic accuracy assessment must be carried out, even if every possible area the size of a minimum mapping unit was observed during

the mapping process. Errors of correspondence between map and user occur for reasons other than discrepancies between remotely sensed and ground perspectives of vegetation.

(BP) As a user-oriented process, the thematic accuracy assessment for the NPS Vegetation Inventory seeks to document error rates, but not error causes. While the results of the assessment may help producers understand mapping errors for improving future mapping efforts, they are not intended to further improve the map being assessed. Ground-truthing activities that are internal to the development or improvement of the map products include calibration and verification. See http://science.nature.nps.gov/im/inventory/vg/index.cfm and Appendix E of this guidance for an explanation of calibration and verification.

(RQ) The sampling design for thematic accuracy assessment will be stratified random sampling with the entire mapped park area the inference area (adjusted, as necessary for costs) and the map classes the strata (equivalent to simple random within each map class). The sample for each map class will be a set of observations selected by a common sampling scheme and centered on a site located by a set of x and y map coordinates. While the entire project (park) area should be treated as the inference area for the sampling whenever feasible, it is recognized that access, costs, and other logistical issues may prevent the entire park from being included in the inference area, particularly for the larger parks. In these cases, extending the results of the thematic accuracy assessment from the inference area to the rest of the project must be justified by assumptions, rather than by statistical inference.

(RQ) The accuracy assessment must be minimum mapping unit based. The area evaluated at each observation will be equivalent in size to the minimum mapping unit for the map class and must be large enough to be likely to contain thematic elements of the vegetation classes to be identified in the reference (ground) data. Polygon based designs are not acceptable (see Appendix A of this guidance). (BP) Multiple observations in a polygon are not only permissible; in many cases they will be required.

(RQ) In order to make reasonably precise statements about the accuracy of each map class, a sample size of 30 observations (sample units) will be allocated to more abundant map classes. Rarer map classes will be sampled with less frequency, with a minimum sample size of either five or as many independent observations as the class can accommodate. Therefore, it is recognized that the accuracy of rare classes will need to be stated with less precision than that of abundant classes.

(RQ) The default minimum mapping unit size (and, thus, the accuracy assessment observation area) is 0.5 hectares, but may be reduced or enlarged for map classes based on the spatial attributes of the vegetation stands represented by the classes and on cost and logistical issues.

(RQ) Individual observations must be independent from one another, as well as from the mapping process.

(BP) Observation sites will be located accurately enough in the field to make the sample data (map class) membership of each observation unambiguous, taking into account observation area size requirements and field positioning error. It is assumed that this will usually be accomplished through the use of sufficiently accurate Global Positioning Satellite (GPS) surveying methods.

(BP) For purposes of field efficiency, the default shape of the observation area around each observation site will be circular. However, field investigators will be given some discretion or guidance, as necessary, to minimally relocate and/or reshape the observation area in order to eliminate excessive heterogeneity (more than one clearly distinct vegetation type) within the observation area.

(RQ) To retain the principle of data independence (Congalton and Green 2009, p. 98) the sample data value may not be used as a factor (even as a partial factor) in determining the reference data (field call) value either before or after the fact.

(RQ) To the extent possible, the field observers will be unaware of ("blind to") the sample data value (the mapped class in which the observation occurs), when collecting or evaluating field (reference) data.

(RQ) The field observers will identify the vegetation to the ecological classification (rather than to the map class). Since vegetation classes are to be either equivalent to the map class or are to be nested uniquely within a single map class (in a many to one relationship between vegetation classes and map classes), identification of the vegetation class alone will determine whether the correct map class has been applied. This practice will avoid the need for a separate map class key and will provide the user with more information on map class composition than would a map class key. (BP) Observers will identify the vegetation type at each site as the best match out of any unit of the finest level of ecological classification to the vegetation in the observation area. The project field key will be the primary means of identification of vegetation types in the field. It may be supplemented by the ecological descriptions, where necessary.

(BP) The ecological classification and field keys should be thoroughly reviewed and quality controlled prior to undertaking the mapping and subsequent thematic accuracy assessment.

(RQ) The field observers must be minimally qualified (though not necessarily an expert), and trained, as necessary, to interpret the ecological field key (e.g., able to identify all plant taxa named in the field key for vegetation types developed for the unit and all taxa occurring in the area with which they may be confused, estimate cover accurately and precisely, interpret field key logic).

(BP) In order to maximize the degree of independence of accuracy assessment from map production, the best practice is that the mappers should not be involved in field observation data collection or evaluation for thematic accuracy assessment.

(BP) Interpretations of the field observations will not be modified after the fact, except to correct known egregious errors (e.g., a clear misidentification of a diagnostic plant taxon that is named in the field key).

(RQ) An individual observation will be considered accurately mapped if the value of the field call (reference data value) matches the value of or is included as a part of the map class in which the observation occurs (sample data value).

(RQ) Minimum field data that should be collected for each observation are: a unique identifier for the observation, park name (may be designated as a part of the unique identifier), reference

data value (the vegetation class observed ("field call"), the geographic position (typically measured with a GPS receiver) of the center point of the observation, estimated error of the geographic position, name of the field observer who evaluated the vegetation, date of the observation, and notes on vegetation type identification issues, if any.

(RQ) Thematic accuracy will be reported using contingency tables that report users' (100% minus commission error rate) and producers' accuracies (100% minus omission error rate) for each map class. These accuracies should be expressed as a percentage with a 90% confidence interval. The overall accuracy rate (accuracy rate pooled observations for all map classes) will also be reported.

(RQ) The thematic accuracy results (primarily the contingency table) and project methods will be reported in a thematic accuracy assessment report or report section for each project.

1.4 Work Flow

The thematic accuracy assessment for the NPS Vegetation Inventory follows completion of production work on the map, including validation to assess overall map accuracy, as needed (see Table 1, Figure 1, and Appendix E of this guidance and http://science.nature.nps.gov/im/inventory/veg/index.cfm). The first three phases of the NPS Vegetation Inventory thematic accuracy assessment are sampling design, response design (field methods), and analysis and estimation (Stehman and Czaplewski 1998). A reporting phase concludes the thematic accuracy assessment process. The next three chapters of this guidance address each of these phases specifically. Major activities of each phase are:

SAMPLING DESIGN (CHAPTER 2.0) STEPS

1. Determine map classes that are to be excluded from assessment, as necessary (see Subsection 2.3.1).

2. Measure area of all map classes to be included in assessment and determine sample sizes (see Subsection 2.5.2).

3. Determine individual map class inference areas (if less than their extent within the entire park) based on balance of access costs with representativeness, as needed (Subsection 2.3.3).

4. Allocate prescribed number of sites (from Step 2 above) to individual inference areas (from Step 3 above) and derive site coordinates.

5. Load coordinates into GPS receiver.

FIELD CAMPAIGN (RESPONSE DESIGN) (CHAPTER 3.0) STEPS

1. Contract, hire, or otherwise engage field observers (although it is part of the response design, this step is often conducted before the sampling design, out of practical necessity) (Subsection 3.2.1).

2. Meet with park managers to coordinate field campaign operational needs and restrictions (Subsection 3.2.2).

3. Train field observers (Subsection 3.2.1).

4. Field observers navigate to assigned sites (Subsections 3.2.3 and 3.2.4).

5. Field observers identify and record vegetation type at sites, primarily using field key, and record site positions (Subsections 3.2.5 through 3.2.9).

6. Review field data for egregious errors.

7. Compile field positional and attribute data into a Geographic Information Systems format and in PLOTS database.

ANALYSIS AND ESTIMATION (CHAPTER 4.0) STEPS

1. Label sites with both map class and field observation class (e.g. through a spatial join of field observation sites (points) to mapped classes (polygons) in a GIS).

2. Cross-tabulate sample data and reference data (typically using a pivot table) to produce raw contingency table (Section 4.3).

3. Calculate contingency table accuracy and confidence interval values (Sections 4.4 and 4.5)

4. Aggregate or adjust raw contingency table classes, as necessary (Section 4.6).

REPORTING (CHAPTER 5.0) STEPS

1. Write accuracy assessment report (or accuracy assessment sections of project report).

1.5 Validation and Accuracy Assessment for Small Parks

For smaller and less complex parks, the accuracy assessment may be employed as both a validation and an accuracy assessment step. Validation is a process of more limited sampling whose objective is to determine whether a final draft map meets an acceptably minimal threshold in order to proceed to the thematic accuracy assessment of individual map classes (see Appendix E of this guidance and http://science.nature.nps.gov/im/inventory/vg/index.cfm and for an explanation of validation). The results of validation address primarily the immediate needs of the map production team and project oversight team to evaluate the overall product before passing it on to a more committing thematic accuracy assessment campaign. Accuracy assessment is a process that seeks to determine the reliability of individual map classes and is primarily to inform the map user. In the course of evaluating individual map classes, the accuracy assessment will also provide a measure of overall map accuracy, although it has a high cost/benefit ratio for the latter purpose. Smaller parks are those with a calculated accuracy assessment workload of fewer than 150 observations, as determined by the sample size allocation procedures described in this guidance (Subsection 2.5.2).

In this small park scenario, if the accuracy assessment also indicates a project-wide accuracy that fulfills the requirements of the validation, no further field work is needed. If the accuracy assessment results indicate that the project-wide accuracy is inadequate, then the results are treated as a validation. The observations then may be used (“recycled”) as verification observations (Table 1, Figure 1, and Appendix E of this guidance and

http://science.nature.nps.gov/im/inventory/vg/index.cfm) in order to improve the map accuracy to a more acceptable level. The accuracy assessment is then repeated using new randomly selected sites. In this case, because the original accuracy assessment observations will have been used to amend map classes, they cannot serve to assess the amended map. Note that only those map classes that require changes pursuant to the first accuracy assessment attempt will need to be re-assessed in the second accuracy assessment attempt. The results from both accuracy assessment campaigns then may be combined to report the accuracy results for each class.

This approach is cost-effective because the difference in total cost between a validation and a per-class accuracy assessment at small parks is relatively small (e.g., up to 150 observations and in a setting where travel distance between sites is relatively trivial). Thus, the cost risk in embarking on a substantial accuracy assessment campaign before understanding project-wide accuracy problems is less of an issue than for larger park projects, and there is a benefit of dispensing with a separate validation step, if the map proves to be reasonably accurate. This approach still requires that producers conduct an adequate level of field calibration and verification during the classification and mapping phases.

For larger, more complex park projects, a separate validation step should be planned in order to avoid committing to an extensive accuracy assessment campaign on a map of possibly unacceptable overall accuracy. This step will likely be intensive enough only to assess the overall map accuracy and not intensive enough to identify individual problem classes.

Table 1. Comparison of field observation activities in the National Park Service Vegetation Inventory.

	ACTIVITY:				
	CLASSIFICATION PLOTS (RELEVES)	**MAP CALIBRATION**	**MAP VERIFICATION**	**MAP VALIDATION**	**ACCURACY ASSESSMENT**
INTERNAL OR EXTERNAL TO MAPPING PROCESS:	Internal	Internal	Internal	External	External
PHASE OF PROJECT:	Production (Ecological Classification)	Production (Mapping)	Production (Mapping)	Evaluation	Evaluation
SPECIFIC PURPOSE:	Ecological Classification (also map calibration)	Produces first draft of map	Checks and revises first draft of map	Checks final draft of map for acceptance	Checks final (accepted) product
NEED:	Required	As needed (recommended)	As needed (recommended)	Required, if acceptance of product is contingent on performance	Required
SAMPLING DESIGN:	Subjective	Subjective	Subjective (can be stratified random)	Stratified random	Stratified random
AMOUNT OF ATTRIBUTE DATA PER OBSERVATION:	Significant. Full floristic enumeration with cover of each species per stratum. Substantial environmental data.	Dominant / frequent / diagnostic species, as needed. Cover of these, as needed. Little environmental data.	Dominant / frequent / diagnostic species, as needed. Cover of these, as needed. Little environmental data.	Dominant / frequent / diagnostic species, as needed. Cover of these, as needed.	Relatively little. Minimally, a vegetation type classified at the site
NUMBER OF OBSERVATIONS:	5 plots per vegetation type	3-10 plots/ map class	3-10 plots/ map class, as needed	1-3 per class, (30-50 total)	5-30 points per map class

Further guidance on these processes may be found in Appendix E and at: http://science.nature.nps.gov/im/inventory/veg/index.cfm.

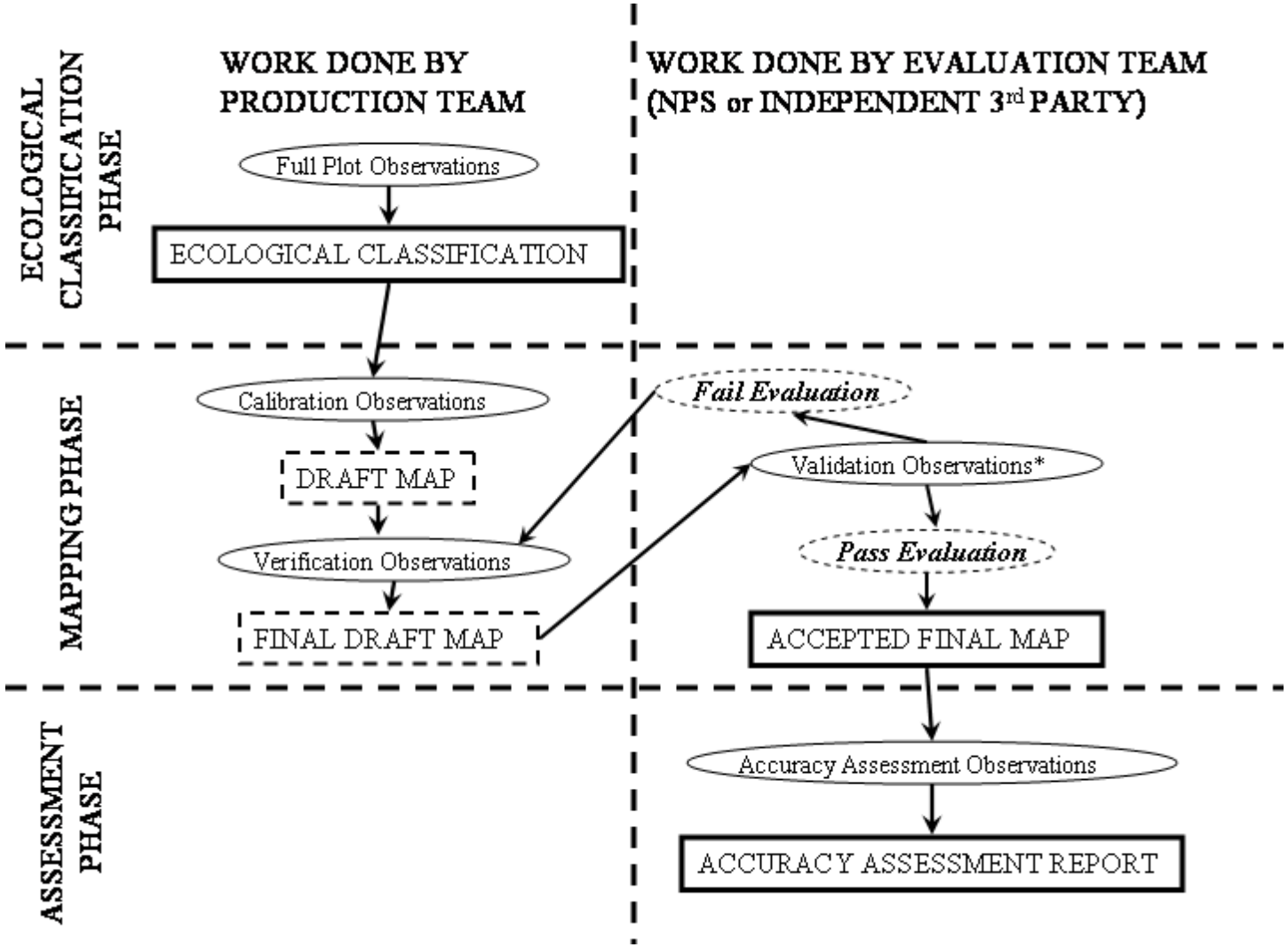

Figure 1. Work flow of operational phases of a NPS Vegetation Inventory Project. The diagram illustrates the relationship of the field activities for ecological classification, map calibration, map verification, map validation, and thematic accuracy assessment. Items in solid boxes with solid outlines represent products; items in boxes with dashed outlines represent intermediate products; Items in ellipses are field activities. Further guidance on these processes may be found in Appendix E and at: http://science.nature.nps.gov/im/inventory/vg/index.cfm.

2.0 Sampling Design

2.1 Sampling Design and Data Collection Objectives

The sampling design (Stehman and Czaplewski 1998) for accuracy assessment within the NPS Vegetation Inventory adheres to the scientific principles that govern sampling and statistical analysis, while also striving to be practical and cost effective. Specifically, the methodology should satisfy the following objectives (Stoms et al. 1994):

1. The methodology should be scientifically sound. In order to accomplish this need, 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 methodology should be economically feasible in view of both time and cost constraints.

3. The methodology 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 be based on the same theoretical foundation, so that the results of individual project assessments are comparable.

The objective of collecting observations for the thematic accuracy assessment is to draw inferences about the magnitude of discrepancies between the true attributes of a point or area and its representation on the map. Thus, the accuracy assessment objective contrasts with the objective of classification plots (releves), which is to provide raw data for the thorough description of vegetation types as efficiently as possible and is often accomplished by subjective sampling (Mueller-Dombois and Ellenberg 1974).

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.

Appendix C describes an example of a sampling design in steps for Shenandoah National Park (Young et al. 2009).

2.2 Observation Area (Minimum Mapping Unit) Size

While the issue of observation area size is of concern for field methods (Chapter 3), it also places some constraints on the sampling design and is treated here. As stipulated in the previous chapter, the thematic accuracy assessment observation area will be equivalent in size to the size of the minimum mapping unit designated for that map class.

The default (standard) minimum mapping unit (MMU) size for the NPS Vegetation Inventory is 0.5 hectare. While this size suffices for many vegetation types, mappers are often able to map vegetation that occurs in smaller stands because these stands contrast well with surrounding stands in imagery. An example of this is a small herbaceous wetland that is surrounded by an upland evergreen woodland. It is often advantageous to delineate these smaller stands when they are easily recognized and reasonably few in number because: (1) they represent vegetation types important to management and (2) designating them as separate map classes can increase the accuracy of surrounding or adjoining classes. This is because these small stands then will not occur as unmapped inclusions that the field observer may fail to recognize as different and that

may influence the field calls in the adjacent map classes. It is also important to consider that the minimum mapping unit size is more limited by the width of stands in the narrowest dimension, than it is by absolute size of the stands. For example, a linear stand of riparian vegetation that averages 5 meters wide and is 2000 meters long may not be fully discernible from a remotely sensed image even though it exceeds a 0.5 hectare in size.

Class-specific MMU size can be determined either *a priori* or *a posteriori*. Generally, mappers should attempt to map at a consistent MMU size for each class determined *a priori*, and such a stipulation may be a minimum project requirement. However, the mapping process itself will have provided insight as to the effective MMU size upon *a posteriori* examination of individual map units (polygons), as a practical matter. For example, if a mapper recognizes ten stands of a particular vegetation class and, if eight of these stands are measured on the map to be less than 0.5 hectare in size, then it is clear from the mapper's rendition that the vegetation type can be recognized consistently at scales smaller than 0.5 hectare. Since the primary purpose of the accuracy assessment is to inform users' about map class properties (rather than to assess project performance, as is the case with map validation), the MMU for this type might be decreased in size from original expectations for stand sizes. This would allow a larger number of mutually independent, non-overlapping assessment observations to be made in the smaller stands. The choice of mapping methods may also affect MMU size. Automated image analysis mapping methods tend to map vegetation in smaller stands, often classifying individual pixels or small groups of pixels as distinct stands, than do vector-based methods that rely on more direct human decisions about the individual stand size. Both types of methods must consider that the lower limits of MMU size are restricted by the minimum thematic unit size (equivalent to the "minimum area" for a community described by Mueller-Dombois and Ellenberg (1974). Areas mapped as stands of a vegetation type must be minimally large enough to have the species composition of the type adequately represented.

In the vector-based mapping methods often employed by projects of the NPS Vegetation Inventory, the relationship between the size of the MMU and mapping costs is not linear. That is, a doubling of MMU size, even project-wide, will result in a fairly modest decrease in mapping costs. Conversely, decreasing the size of the MMU for a limited number of map classes by a factor of two will result in a relatively modest increase in cost (provided features at the smaller MMU can still be mapped effectively). An order of magnitude change in the size of MMU usually is necessary to cause a substantial change in mapping effort and/or cost. Raster-based mapping methods tend to be more sensitive to MMU size, but, since these are often automated, cost is less of an issue than is an ecologically appropriate scaling of map units, compared to the minimum thematic unit. Thus, within the acceptable ranges of the NPS Vegetation Inventory and using 0.5 hectare as a default size, choice of MMU (and accuracy assessment observation size) of individual map classes should be guided primarily by appropriate spatial scales for recognizing stands of the types. For example, a suitable MMU size at which to evaluate widespread upland forest vegetation stands might be 0.5 hectare, but 0.25 hectare might be a more appropriate MMU size for rarer wetland or riparian vegetation types dominated by shrubs or herbaceous vegetation and that occur typically in small or linear stands.

Decreasing the MMU for large patch vegetation prevalent in many parks will not necessarily yield benefits in spatial resolution. In fact, it may actually decrease reported accuracy during the

assessment, especially if the MMU size approaches the size the minimum thematic unit for the vegetation type.

Table 2 depicts some possible ranges of minimum mapping unit size within the ranges allowed by the NPS Vegetation Inventory. Table 3 shows the radius distance for a circular observation area the size of each MMU and may be used in locating observation sites.

Table 2. Range of suitable minimum mapping unit (MMU) sizes for types of vegetation.

VEGETATION PHYSIOGNOMIC / ECOLOGICAL CATEGORY	MMU RANGE
Forests:	0.25 to 1.0 hectare
Humid to Semi-arid Woodlands:	0.5 to 1.0 hectare
Arid to Semi-arid Woodlands, Shrublands, and Wooded Shrublands:	0.5 to 2.0 hectare
Humid to Semi-arid Shrublands and Wooded Shrublands:	0.1 to 0.5 hectare
Herbaceous Vegetation and Non-vascular Vegetation:	0.1 to 0.5 hectare

Table 3. Radius lengths for suitable observation area sizes.

OBSERVATION AREA SIZE	OBSERVATION AREA RADIUS LENGTH[1]
0.1 hectare	18 meters
0.25 hectare	28 meters
0.5 hectare	40 meters
1.0 hectare	56 meters
2.0 hectare	80 meters

[1] - radius lengths are rounded to the nearest meter.

The use of the MMU sizes of 0.1 hectare and 2.0 hectares generally should be reserved for unusual situations. For example, 0.1 hectare has been used for very fragmented (small and/or linear) map classes that represent shrublands or herbaceous vegetation, but is generally inappropriately small for vegetation types that are dominated by trees. An example of this may be small patches of salt meadow hay (*Spartina patens*) wetlands within secondary or tertiary dune systems on a barrier island. A MMU size of 2.0 hectares might be used in upland vegetation of arid areas where diagnostic species, such as Joshua tree (*Yucca brevifolia*) in the Mojave Desert or saguaro (*Carnegiea gigantea*) in the Sonoran Desert, tend to be both well dispersed and to occur at relatively low absolute cover.

During the accuracy assessment campaign, it will be necessary to instruct field teams of the MMU size selected for each observation, without revealing the map class value of the observation (see Subsection 3.2.4 and Appendix C). This can be accomplished by embedding a code that designates the MMU size into the unique identifier for the GPS waypoint denoting the individual site (e.g., JOTR-AA-025-0110 and JOTR-AA-050-1154 denotes that a 0.25 hectare

area should be observed at site number 110, while a 0.50 hectare area should be observed at site number 1154 at Joshua Tree National Park).

2.3 Inference Area Selection

2.3.1 Included and Excluded Map Classes

All map classes representing vegetation that would be considered natural (including semi-natural) by the National Vegetation Classification, Version 2 (Federal Geographic Data Committee 2008) should be included in the accuracy assessment. Forest plantations (e.g., Norway Spruce Planted Forest, Red Pine Planted Forest) are considered semi-natural. These are somewhat prone to mapping confusion with fully natural forests and with one another and, thus, should be assessed.

Naturally occurring sparsely vegetated surfaces such as rock outcrops, talus, and intertidal flats usually support natural vegetation communities (often characterized by non-vascular vegetation) and should be treated as natural vegetation in the thematic accuracy assessment. It is permissible to omit these surfaces from the accuracy assessment if only one map class at the NVC level of Macrogroup (Federal Geographic Data Committee 2008) is used to represent them, as these classes will likely show little confusion with other classes. However, if classification distinctions between types are made at a level more finely resolved than the NVC Macrogroup level, then the map classes representing these types should be included in the inference area. An example would be a map class that represents mafic rock talus nonvascular vegetation and a map that represents silicaceous rock talus nonvascular vegetation. In the case of rock outcrop types that may occur on cliffs, some areas of these map classes may have to be eliminated from consideration for observations, as discussed above, because of the difficulty imposed by steep terrain.

Map classes that represent vegetation that would be considered cultural by the NVC (Federal Geographic Data Committee 2008) either may or may not be included in the inference area. When mapped to coarser thematic levels, such as NVC Cultural Group (Cultural Level 5) or coarser, map classes that represent cultural vegetation classes will be relatively easy to distinguish from other classes and the accuracy usually can be assumed to be very high, even in the absence of formal assessment. For this situation, it is acceptable to omit most cultural vegetation classes from the thematic accuracy assessment. However, if some vegetation types that are regarded as cultural are differentiated and depicted at thematically finer floristic levels, they should be included in an accuracy assessment. For example, it would not be necessary to assess a map class depicted as “Orchard;” but should map classes depicted as “apple orchard” and peach orchard” in a single project, they each should be assessed. Where map classes that represent vegetation that would be considered cultural (e.g., lawns) might reasonably be confused with map classes that represent vegetation that would be considered natural or semi-natural by the NVC (e.g., semi-natural old fields), these should be included in the inference area.

Truly unvegetated land cover surfaces (deep, open water, pavement and other human-made surfaces, extensive sand dunes, intertidal parts of sandy beaches, permanent snow or ice fields, etc.) and map complexes that represent a mixture of both unvegetated surfaces and cultural vegetation (e.g., developed areas) often are included as map classes in vegetation maps. These map classes may be omitted from the thematic accuracy assessment.

2.3.2 Areas with Training Sites

Accuracy assessment sites that happen to be co-located by chance with sites that have been used for mapping training or for classification plots are acceptable. Since the selection of sample sites for thematic accuracy assessment is independent of the selection of vegetation classification plot sites or mapping training (calibration or verification) sites, the probability that a training site will be selected as an accuracy assessment site is small, except for some very rare map classes. In these cases, the accuracy assessment sites visited may comprise a significant proportion of the area of the mapped class and may need to be visited to obtain a minimally adequate sample size. Even though the accuracy of these sites would be expected to be high, they should be included in the inference area, as other causes for thematic error (field key and observer error) will also affect accuracy. Additionally, the accuracy assessment seeks to evaluate all possible sites in the inference area, regardless of whether some may have been visited or not.

The assessment of sites that have been used previously for mapping training does not violate the principle of assessment independence, as long as the training site data (as well as the map class delineated at the site) are not revealed to the observer. The key consideration is that the application of the two different data sets be independent (Congalton and Green 2009), not whether the sites are independent.

2.3.3 Adjustments for Access Costs

As stipulated in Chapter 1, individual observation sites are to be selected using a simple random sampling design within each map class, with the entire project (park) area included as the inference area whenever feasible. This usually will be the case for smaller parks and for many medium-sized parks. Accessibility generally will be related to the size and shape of the park, the density of roads and trails, the amount of difficult terrain (steep slopes, dense vegetation) to traverse, and the frequency and size of travel barriers (e.g., cliffs, rivers, private lands). Where access costs are high due to distance and/or difficult terrain, the inference area may be reduced in size to more accessible parts of each map class. In large to very large parks, reducing the size of the inference area by applying a cost surface (access buffer) to limit the data collection to more accessible sites in order to make costs of data collection reasonable will be almost certainly necessary. The inherent assumption in this approach is that the less accessible areas that are not considered for sampling will have similar accuracy properties as the more accessible areas that are included in the inference area. The inference area included in the sampling design must be carefully described in the accuracy assessment report, so that the user understands to which park areas the accuracy assessment results apply by statistical inference and to which areas applying the results must be done by assumptions.

Not all map classes will be evenly distributed near or away from routes of easy access such as roads and trails. In mountainous western parks, roads may be denser in low elevation areas and high elevation (e.g., alpine) areas may be uniformly less accessible. In the Appalachian Mountains, trails and some roads often run along stream valleys or ridge crests, so that vegetation types that occur in these settings are more accessible, whereas vegetation occurring on steep middle slopes (e.g., boulder fields) may be less so. Some minimum degree of map class representativeness is desirable for each map class.

Except in situations of exceptionally difficult access, it is recommended that the most accessible 50th percentile of every abundant (50 hectares or more in area) map class should be included in

the inference population, and a minimum of at least the most accessible 30th percentile of abundant classes should be included in even the most difficult of access situations. For rarer map classes (less than 50 hectares), this percentile should be increased, on a class-by-class basis, in order to allow at least the appropriate map class sample size (see later in this chapter) to be considered, whenever feasible.

As a rule of thumb and in balance with the above guidance regarding individual map class representation, cost surfaces should be adjusted, as necessary, in order to allow each single field team to collect, at minimum, an average of 8 observations per field day throughout the accuracy assessment campaign.

Cost surfaces may involve effort and expense as well as time. Areas that would be either dangerous to access without expensive or time-consuming methods (e.g. technical climbing skills) may be eliminated from the inference area on the grounds of access cost or difficulty.

Where it is permitted and feasible, access by helicopter may be considered. Considerations for helicopter use include (1) if the number of difficult access sites is small and of critical concern or (2) if helicopter operations can reduce costs of field data production over other means. Helicopter time can be used very efficiently if multiple observation teams can be continuously delivered and retrieved at rally points, using either a cluster sampling or multistage sampling design (Section 2.4). One team might be making multiple observations along a route while another team is flown to a new route. Because the helicopter carrying capacity is often limited and because of the added safety factor afforded by the helicopter, one person observation teams should be considered. This approach makes best use of helicopter [hourly] expenses by keeping the pilot and helicopter continuously engaged in providing access, rather than in “down time.” However, such operations require fairly meticulous planning, timing, and execution.

Logistical problems other than distance or travel difficulty may increase access costs. Access to or through private or otherwise sensitive lands may increase costs (due to the need to gain permission) or even eliminate some sites from consideration.

In cases in which large or otherwise significant areas of the park must be omitted from the inference area of the thematic accuracy assessment for a NPS Vegetation Inventory project, park managers may find later occasion to do supplemental sampling in those areas for special studies or other purposes. For example, a portion of a park might be considered for federal Wilderness designation, with part of the justification for the designation is a concentration of a vegetation type that is habitat to a species of concern. If the proposed Wilderness Area is remote, costs may have prevented many stands of the type and all stands in the area of concern from being included in the inference area of the NPS Vegetation Inventory project. This might leave some doubt as to whether the map accurately represents the amount of the type in the management area of current concern. Because of the legal and public interest implications, a targeted accuracy assessment at this type only within this area only may be both feasible and in the interest of the proposal.

An alternative to omitting less accessible areas of the park from the inference area would be to employ access cost strata and to sample higher cost areas less densely than more accessible areas. This approach will require and additional layer of stratification within the map class strata and, thus, complicates the calculations of accuracy (see Subsection 4.4 and some references in

the Literature Cited section). It is likely that travel costs per observation could be quite high for those observations in less accessible strata in very large parks. On the other hand, it would allow the less accessible areas to be represented in the accuracy assessment.

2.3.4 Adjustments to Achieve Sample Data Homogeneity in Observations

In a digital vegetation map, vegetation data will be classified into nonoverlapping polygons with discrete boundaries that imply a fixed line along which there will be a distinct change from one vegetation class to another. On the ground, these delineations often appear more as transitional zones (ecotones) between individual vegetation types of varying width. Some accuracy assessment sampling designs have avoided potential sites of ambiguous (transitional between types) vegetation, often by locating field observation sites away from map class boundaries with other map classes. However, the purpose of employing an ecological and mapping classification for an inventory implies a clear class membership for any site within the study area, and the purpose of a thematic accuracy assessment is to evaluate the "goodness of fit" between the discrete (classified) model and the more continuous real condition. Thus, the term "ecotone" has no real meaning in the context of discrete classes because nearly any site will have more affinity for one class than for others, depending on the classification rules. The possibility of transitional or otherwise atypical vegetation should not be a factor in selecting sites for thematic accuracy assessment.

However, in order to define the sample data value of a site in order to determine the degree of match with the site's corresponding reference data value, it is usually necessary to identify individual observation sites as unambiguously belonging to a single map class (i.e., they have a single possible sample data value). In order to do this efficiently the site must be located sufficiently far from (or, in a Geographic Information System context, buffered from) the map class boundary to eliminate the possibility that the observed area was of mixed sample data values due to (1) confusion as to whether the observation area is wholly contained within the map class, (2) positional error due to site location (GPS) error and (3) allowable positional error in the map data. To calculate the buffer distance that would be required to eliminate this uncertainty, the square root of the sum squares of these error sources can be calculated using the following equation.

$$\text{Buffer Distance} = \sqrt{R^2 + F^2 + M^2}$$

In this equation, R is the radius distance of the observation area (from Table 3), F is the expected (e.g., 90th percentile field positioning (GPS) error distance, and M is the expected maximum positional error distance in the map. The value for the term M may be generalized to 12 meters for all NPS Vegetation Inventory projects that meet National Map Accuracy Standard (NMAS) requirements for positional accuracy of 1:24,000 scale products.

For example, if the observation area is a circular area the size of a 0.5 hectare minimum mapping unit (40 meters radius), the expected maximum field positioning error is 15 meters, and 12 meters is the allowable positional error in the map data for 90% of all map positions, the sample site should be positioned at least the root sum square of these three quantities from the polygon boundary (in this case, approximately 44 meters). Site positioning can be accomplished in a sampling scheme that is generated in a Geographic Information System (GIS) by creating a buffer of 44 meters on the interior of all polygons within the inference area, for this scenario.

The usual source of the largest positioning error, the radius of the observation area, might be reduced to allow more of the map classes to be included in the inference area if it can be assumed that including a small amount of a second vegetation type in the inference area (or if the logistics of the field observer taking the time to adjust a small number of site positions) will not affect the classification of the site. For example, if having up to 10% of the observation area for a map class occurring in another vegetation type is acceptable for correctly classifying the site, then the buffer from map class boundaries might be reduced to about 35 meters, in the case above, since this amount of buffering should keep from 90% to 100% of the area of every observation area within the target map class.

If adjoining polygons have the same sample data value (but are retained as separate polygons because they have different values in other attributes), the boundary between them should be dissolved prior to buffering polygon interiors. This is for purposes of assigning thematic accuracy assessment only. For example, if a Douglas-fir Forest polygon of height class 4 adjoins a Douglas-fir Forest polygon of height class 3, it is permissible for an accuracy assessment site to fall on the boundary between them and to include parts of both polygons, since the attribute of interest (the vegetation type) in the sample data will be homogeneous throughout the observation area, even though another attribute (height class) will not be.

The buffering step will largely eliminate ambiguity as to map class membership in individual observations, since it ensures that a single map class will completely contain the observation area. Assuming that map class (polygon) boundaries coincide well with real vegetation boundaries, this step may bias the sample somewhat, since class membership away from the transition zone may be more pronounced, thereby increasing the chance of agreement between the class assignment in the database and the class assignment made during the accuracy assessment. However, unlike intentional buffering to avoid transition zones, the practice is not arbitrary, since it uses a decision rule based on real sources of positioning error in the test process, rather than on sources of thematic error in the data. Also, there is no assurance that vegetation type boundaries as viewed on the ground will match map class boundaries.

It is also desirable or necessary to eliminate ambiguity in the reference data values for the observations. This ambiguity can occur where an unmapped vegetation class boundary is encountered within the accuracy assessment observation area, and it may require field adjustment in the original site positions. This need is addressed in the next chapter.

2.3.5 Adjustments for Independence Among Observations

The preceding subsections describe how the inference area may be reduced in extent in order to eliminate (1) non-vegetated map classes, (2) areas of high access cost, and (3) boundary areas where map class membership of observations might be uncertain. After these steps have been completed, the positioning of individual observation sites must be considered, in order to maintain independence between observations.

Observation sites will be assigned within an individual map class by allocating them using simple random sampling with the map class (amended as above, as needed) serving as the inference area. Site selection is usually accomplished in a GIS using a program that selects a random x (easting or longitude) value and a random y (northing or latitude) value along a grid of sufficiently fine resolution relative to the map classes (e.g., National Park Service 2008). In such

a sampling scheme, the individual sites are represented as dimensionless points and would coincide only in unlikely cases in which the same x and y values are selected more than once. However, due to the need to observe a sufficiently large area (the size of the MMU for that map class) around the point in order to determine the reference (field) class value, the selection of a sample size of points that is relatively large compared to the map class area is likely to produce some site locations that are close enough to one another so that the observation areas associated with these points overlap in space, particularly so in map classes that have limited total area.

For the purposes of thematic accuracy assessment, the inference population can be regarded as a potentially infinite set of individual observations in time, as well in space. Errors in correspondence between map and observer include factors other than mapping error, such as field key error, reasonable user interpretation error, and vegetation that was not described in the ecological classification step. In that context, observations of overlapping observation areas, even those that overlap completely, could be made independent by simply having different and independent field teams observe and identify them. However, because of the costs associated with access and the difficulty in recruiting qualified field observers, this approach would be largely impractical for the NPS Vegetation Inventory. Field teams most often will need to gain access to and observe sites sequentially by their proximity to one another to save access time.

If significant portions of one or more observation areas overlap spatially and the same field observer is visiting them in short sequence, it would be unreasonable to expect the observer to be able to make the field calls independently from one another, since the observer's decision on the initial call naturally would tend to influence the decision on a subsequent call that includes much of the same observation area and vegetation. In order to maintain independence of observations, the observation areas of individual observations should be independent from one another (non-overlapping), although, as is the case in Subsection 2.3.4, it may be acceptable for independence of field calls by the same observer if adjacent observation areas overlap by a very small area (e.g., less than 10%). It is not only permissible, but is often necessary, for polygons to be populated with multiple sample sites.

As a general rule, individual observation sites should be located so that their areas overlap minimally or not at all (sampling without replacement). In a site allocation methodology, this might involve generating the requisite sample size for a map class, then to use an unbiased (random selection) means of eliminating individual sites whose observation areas overlap those of other sites. The eliminated sites are then replaced by sites generated from a subsequent round of site allocation, with priority for inclusion in the sample being given to sites generated in earlier rounds. If the map class area is very limited relative to the number of sites and associated observation areas that must be accommodated, it may be more efficient to generate a larger number of sites than is necessary with each round of site allocation and eliminate those that encroach upon sites produced in earlier rounds of allocation. The rules developed for priority of allocation will affect the location and, in some cases, the maximum number of sites that can be accommodated in a map class. This is because subsequent selections of site locations often will be limited in their accommodation within the map class without overlap with other sites because of specific placements of sites generated in an earlier allocation round.

2.4 Sampling Methods

There are a number of methods for drawing samples from a population. The most common are random, stratified random, multistage, cluster, and systematic sampling (Cochran 1977, Thompson 2002), and subjectively representative sampling (Mueller-Dombois and Ellenberg 1974). Other approaches exist, but they are essentially variations on the five methods listed here. Because of the statistical test requirements of the accuracy assessment process, systematic sampling and subjectively representative sampling are inappropriate for purposes of NPS Vegetation Inventory thematic accuracy assessments. Sampling methods employed are essentially some combination of the other three approaches.

Individual map classes are the inference areas of interest for the NPS Vegetation Inventory that require an accuracy statement. The standard design employed by NPS is stratified random sampling (Cochran 1977, Thompson 2002) with all map classes (pooled) the inference area and the individual map classes the strata, with a sufficient number of observations (sample units) within each class (e.g., 30 sample units) in order to make a reasonably precise statement about the accuracy of that individual class. The number of observations that is sufficient to make a precise statement about the accuracy of each of the [usually] many map classes that comprise the map requires much more effort than would be needed to make a similarly precise statement about the entire map. A smaller number (e.g., 30-50) of observations may be used to produce a precise statement for the entire map for validation, a process that evaluates the map accuracy for acceptance or rejection (as in a contract performance assessment) prior to the per class accuracy assessment. Nevertheless, once the per-class accuracy assessment is made, one may also use the pooled sample from the individual classes to make a more accurate and precise statement (than the validation) about overall map accuracy. However, having employed stratification by map class across the entire map requires that the accuracy assessment sample design be viewed and analyzed as a stratified random design, rather than as a simple random sampling problem. Computational methods and issues are discussed in Section 4.5.

Although little used to date by the NPS Vegetation Inventory, multistage (e.g., two-stage) sampling (Cochran 1977, Thompson 2002), and cluster sampling (Cochran 1977, Stehman 1995, Stehman and Czaplewski 1998, Thompson 2002; Stehman 2009) might be required in very large parks to reduce access costs even more than would simply eliminating distant parts of map classes through a cost surface (access buffer). For multistage sampling, the park might be divided into sections of relatively uniform map class diversity. The primary sampling units (e.g., a section or block of sections) could be randomly selected until the entire range of class variability is selected. A second stage of stratified random sampling of the vegetation classes would then take place within the primary units. For cluster sampling, a range of secondary sampling units (areas in which different vegetation types are concentrated near one another) are grouped together into primary sampling units. These primary units might then be sampled randomly, and all secondary units within the selected primary units assessed.

Compared to single stage sampling methods, multistage and cluster sampling approaches are more sensitive to effects of spatial autocorrelation of errors and would seem most appropriate where the variability in vegetation is more or less uniformly distributed. Additionally, because some classes, particularly rarer ones, will be highly restricted within a park and because large parks will have, if anything, less uniformity in vegetation distribution than will smaller parks, it is almost a certainty that this design would have to be supplemented by designs that are sensitive

to the distribution of sporadic or concentrated phenomena (Cochran 1977) and that can capture some vegetation types that are underrepresented in the primary strata.

2.5 Determining Sample Sizes

2.5.1 Determining Acceptable Levels of Error and Confidence

Accuracy of a vegetation map may be evaluated from the perspective of whether the accuracy of an entire map or the accuracies of its individual classes meet some specified threshold of minimal accuracy, as in evaluating project performance. In such a case, expressing the results of the assessment as a hypothesis test would be appropriate. Since the NPS Vegetation Inventory accuracy assessment process is now more narrowly defined as a user information tool after the project, rather than a project performance tool, this aspect of evaluating accuracy is now covered by the validation step.

By the start of the thematic accuracy assessment campaign for a park, the map will have been determined to have met a minimum accuracy threshold. The objective of the NPS Vegetation Inventory thematic accuracy assessment is to inform the map user of the accuracy (limitations) of individual map classes with a reasonably reliable estimate of the true accuracy of each map class. The reliability of the accuracy estimate is reflected in its precision.

Precision is a measure of dispersion of the probability distribution associated with a measurement. For the precision of an accuracy estimate (e.g., a mean from a sample) to be understood, a confidence interval is required. A confidence interval is an interval within which we have a specific level of confidence that the true value of an estimate lies; the narrower the confidence interval associated with an estimate, the more precise is the estimate. The width of a confidence interval is affected by (1) the sample size used to derive the point estimate, (2) the confidence level that the user selects, and, (3) for binomial distributions, the value obtained for the point estimate derived from the sample (see Figure 2). The first two factors may be controlled by the investigator conducting the sampling; the third factor is a property of the inference population being sampled and is often unknown or poorly known by the investigator during the planning of the sampling. Larger sample sizes will result in a narrower confidence interval, as will lower confidence levels. Smaller sample sizes and higher confidence levels will widen the confidence interval. For binomial distributions, point estimates of the mean that approach 0.5 (50%) will widen the confidence interval, whereas, with binomial (percentage) data, estimates that approach either 0.0 (0%) or 1.0 (100%) will narrow the interval. Figure 2 shows the relationship between the sample size and confidence interval width for confidence levels of 99%, 95%, and 90%, for two somewhat extreme values of $\hat{p}$: 0.5 (50%) and 0.9 (90%).

Of the parameters that can be controlled by the study design, the acceptance of a lower confidence level can narrow the confidence interval width, as does increasing sample size (particularly if the starting ample size is small). 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%. The NPS Vegetation Inventory uses a confidence level of 90%. Therefore, the width of the confidence interval will vary with changing sample size. Since thematic accuracy sample size for a map class varies with class abundance for the NPS Vegetation Inventory (Section 2.5.2), confidence intervals for map classes that occupy fewer than 50 hectares will be wider than those for more abundant classes.

Using an estimation model of the normal approximation to the binomial distribution, a sample size that ensures a confidence level of 90% with an acceptable two-sided confidence interval width of 20% for an accuracy estimate (p) of 0.5 would require nearly 30 (27) observations per map class (Figure 2,; see also Goodchild et al. 1994). For NPS Vegetation Inventory purposes, this allocation has proven to be a reasonable balance between precision and cost, provided field campaigns can collect individual observations expeditiously (large and/or complex parks may require more than 1,000 observations).

Computations of confidence intervals from sample data are discussed in Chapter 4.

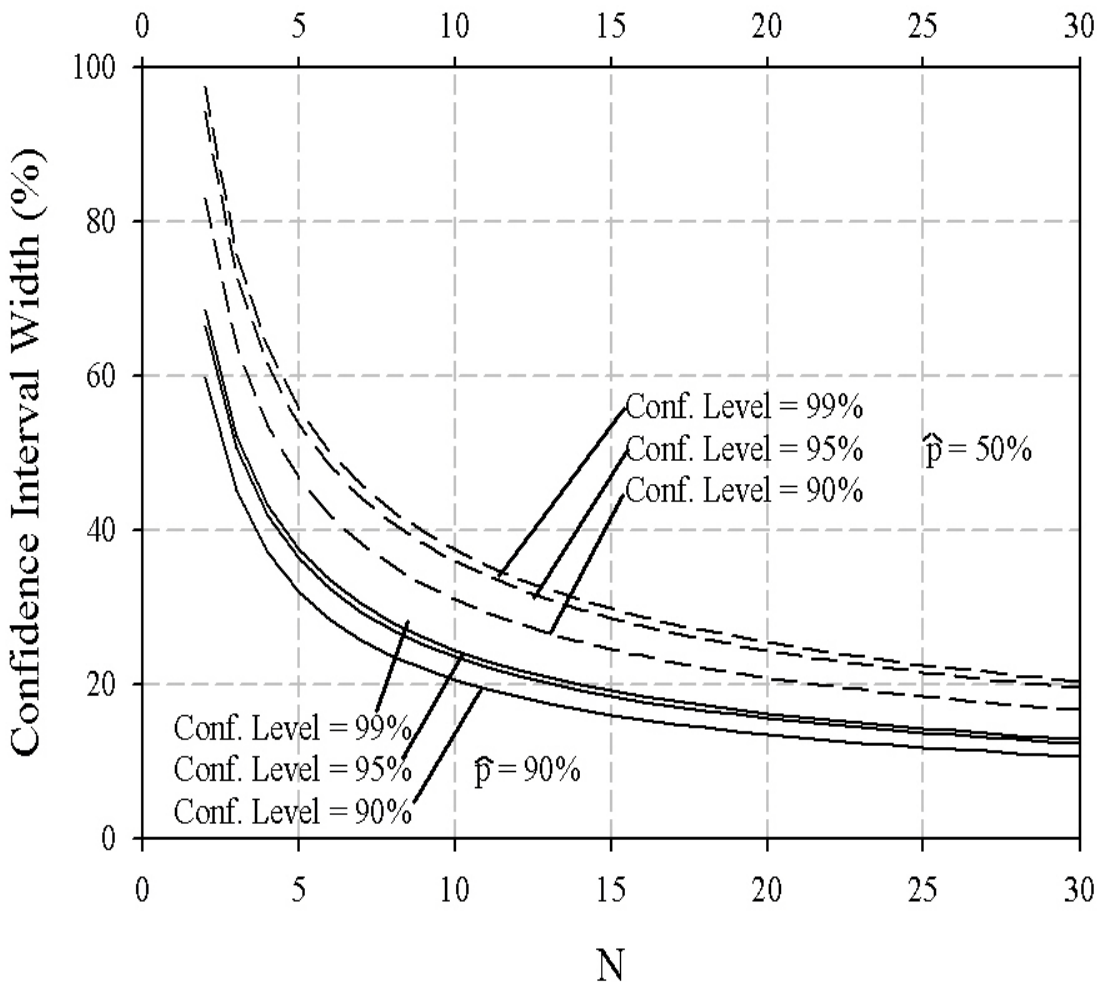

Figure 2. Relationship between sample size (N), confidence interval width, and confidence level for two different values of the point estimate of accuracy ($\hat{p}$).

2.5.2 Number of Observations per Map Class

The 1994 accuracy assessment guidelines (Environmental Systems Research Institute et al. 1994) listed two approaches that may be taken for weighting observations by map class. They are:

1. Sample size can be made proportional to the abundance and frequency of the map class. With this approach, the rarer classes would receive few to no observations. This is unacceptable, since all map classes are potentially individually important to park management, and, thus, it is desirable to have a reasonably precise estimate of accuracy for each class. Conversely, the sample size that might be generated for the most abundant classes would be an inefficient use of field time, since the rate of confidence interval width reduction per observation decreases with increasing sample size, representing a decreasing benefit per unit cost (see Figure 2 in Chapter 4). Furthermore, travel time required to reach all sites in a sample of randomly allocated sites of given size is greater for abundant (spatially extensive) map classes than it is for rarer map classes, which generally have a more limited distribution (i.e., sampling is more efficient in rarer map classes and often significantly so).

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

Because it is important for the user to understand the error properties of each map class, regardless of its abundance, an ideal scenario might place an equal number of observations in each map class. This number should be large enough so that the point estimate of accuracy for the class is reasonably reliable, but small enough to avoid excessive time and cost. Because an area representing a minimally adequate stand size needed to evaluate a vegetation type must be observed at each site and because of the logistical site location constraints discussed earlier in this chapter, some rarer map classes (those that cover less total area) may not accommodate this ideal number of observations.

The 1994 guidelines recommended the second approach, with a sample size of either 30 (for abundant and fragmented classes), 20 (for abundant, but less fragmented, or less abundant, but more fragmented classes), or 5 or fewer (for rare or very rare classes) (Environmental Systems Research Institute et al. 1994). In the practice of allocating sampling sites in NPS parks using these three alternative samples sizes, it was found that many to most map classes in smaller parks (e.g., Cogan 2007, Patterson 2008) and some less abundant map classes in larger parks are too sparsely populated to make meaningfully precise estimates of accuracy. To correct that deficiency and, to take into account the site locating constraints and minimum mapping unit sizes, a blend of the two approaches described above is now employed to assign numbers of observations to individual map classes according to one of three scenarios (see also Table 4):

Table 4. Standard sample size allocations for NPS Vegetation Inventory thematic accuracy assessment, based on map class area.

MAP CLASS TOTAL AREA*	NUMBER OF OBSERVATIONS PER MAP CLASS
> 50 hectares	30*
8.33 to 50 hectares	0.6 per hectare*
< 8.33 hectares	5*

* - as measured before buffering for cost surface (access buffer) or for map class boundary buffer.

** - or as many spatially independent (non-overlapping) observation sites as map class area, MMU size and other considerations will allow.

Scenario A: The class is abundant. It covers more than 50 hectares in total area. The map class receives the maximum sample size of 30.

Scenario B: The class is relatively abundant. It covers at least 8.33 hectares, but no more than 50 hectares in total area. The map class receives a sample size of 0.6 observations per hectare of the map class (= one observation for every 1.67 hectares of map class area). (This ratio allocates observations at an equal density rate to 30 observations per 50 hectares).

Scenario C: The class is relatively rare. It covers less than 8.33 hectares in total area. The map class receives 5 observations (the recommended minimum sample size).

When necessary, the minimum sample size should be further reduced to the maximum number of spatially independent (non-overlapping) observation sites that can be accommodated within the map class. This situation will occur most often in Scenario C, but may occur occasionally in Scenario B and, rarely, in Scenario A. Situations that will require further reduction are: (1) the map class is extremely rare (e.g., a map class that covers a total of 2 hectares can accommodate a maximum of 4 non-overlapping observation areas of minimum mapping unit size of 0.5 hectare and will probably accommodate fewer due to the need for observations to be spatially independent from one another), (2) the map class is extremely fragmented (in small polygons) or linear (so that the number of available sites around which an observation area of the minimum mapping unit size that is entirely within the target map class will be limited).

Map class areas specified above are computed prior to any reduction of map class inference area to accommodate access (cost surface) and/or observation positioning concerns (see Subsection 2.3.3 through Subsection 2.3.5).

Additional observations for or stratification by the degree of fragmentation of map classes (more polygons) (Environmental Systems Research Institute et al. 1994) is no longer employed. Such a design would be generally of interest to producers, but they represent an additional level of stratification, complicating accuracy computations. This practice also spreads field observations more than does a simple random sampling design, increasing the access costs. Additional observations for or stratification by the degree of fragmentation of map classes (more polygons) (Environmental Systems Research Institute et al. 1994) is no longer employed. Such a design would be generally of interest to producers, but they represent an additional level of stratification, complicating accuracy computations. This practice also spreads field observations more than does a simple random sampling design, increasing the access costs.

2.6 Geographic Information Systems Design Considerations

Appendix C provides an example of the specific GIS procedures that might be employed to achieve the requirements and best practices described in this chapter. Especially with larger parks, it is advisable to perform one to several “dry run” rounds of generating the field campaign sample, in order to gain an understanding of how the project-specific criteria will affect both representation and cost efficiency and other logistical concerns and to modify the criteria, as needed. The output of a sampling design for a thematic accuracy assessment is a set of site [x and y] coordinates that can be loaded into a GPS receiver for navigation by field observers. After the “dry run” rounds confirm the adequacy of the design criteria, the round which actually allocates sites for the field campaign should be declared *a priori*.

3.0 Field Methods (Response Design)

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

Many of the techniques for measuring uncertainty in mapped classes were developed for remote sensing in order to provide users and producers with ways to assess the thematic accuracy of remotely sensed land classifications. In this kind of accuracy assessment, it is common practice to select a sample of observation sites and to compare the class that was assigned to each site by the mapping effort being evaluated with that obtained using a source of higher accuracy than that of the map. The source is usually a form of ground truthing that is obtained by human observation and interpretation in the field.

A successful thematic accuracy assessment of a map requires that data be collected at reference sites using a response design (Stehman and Czaplewski 1998). The response design is an assessment methodology that produces reference data that can be assumed to be of significantly higher accuracy than the map itself, although no assessment methodology can produce infallibly accurate results. The methodology used in the response design should be independent from the mapping results that are being assessed. The choice of an appropriate methodology of assessment depends on three main factors: (1) accuracy of the method, (2) cost of the method, and (3) the relevance of the method to the user. Sources of higher accuracy other than the most accurate means possible may be used to increase sampling efficiency if the loss of confidence in the identification is acceptable (Environmental Systems Research Institute et al. 1994).

During the course of the NPS Vegetation Inventory from 1994 to 2009, five methodologies have either been considered and/or used. They are:

1. Collection of field data using the same methodology as the vegetation classification (e.g., establish a classification plot (or releve) at the site and then return to the office for analysis and assignment of the site to a vegetation type).

2. Field observation and class assignment in the field by an expert in the vegetation classification for the area. The expert is usually the ecologist that developed the vegetation classification or similarly experienced person. Minimal other data are collected.

3. Field observation and class assignment in the field by a qualified and trained, but non-expert observer (as in Method 4); the observer also collects supplemental data on species composition and cover and other parameters that would aid in identification of the type. The field calls are assessed for matches to the map data. Supplemental data for the mismatches are evaluated in the office by an expert (as in Method 2) and the field call is modified, as necessary. The inherent assumption is that the field call is accurate, unless superseded by the office call. In order to maintain the principle of reference data independence from the sample data, and to avoid biases in the reference data determinations, experts in the office must make their determinations from the supplemental data without only out knowing the sample data value of the site that is being evaluated and without knowing the field observers' assessment call. Both of these principles are sometimes violated.

4. Field observation and class assignment by qualified and trained non-expert field observer. The observer uses the classification materials (usually a field key to vegetation types) to make a

relatively objective determination of type. “Qualified and trained” requires that the observer meet the requirements of Subsection 3.2.1, but has not developed the classification. Where appropriate, minimal amounts of supplemental data (e.g., a list of dominant species) may be collected. The purpose of these data is restricted to the need to review observer qualifications periodically and/or to detect egregious errors that would justify elimination of observations from analysis. However, supplemental data that would enable a *post hoc* review of observer determinations (e.g., cover of dominant species in all strata) are not collected.

5. Aerial videography.

Method 1 was suggested by Environmental Systems Research Institute et al. (1994) as an ideal method for a higher source of accuracy. It is undoubtedly the highest possible source of accuracy that could be used for thematic accuracy assessment reference data and, thus, has many of the advantages recognized for Method 2.

This method has never been employed by the NPS Vegetation Inventory because it is clearly cost-prohibitive. Normally, an average of three to ten classification plots (releves) per type have been collected to construct the vegetation classification, and the data needs for this method are time consuming enough that only two to four such plots normally can be collected per field day. Assuming the ideal situation in which most or all map classes represent a unique vegetation type, the accuracy assessment would require 30 plots per classified type, in many or most cases, or about three to ten times the number of plots used to build the classification. Furthermore, the requirements of random sampling for statistical representativeness means that travel to a individual accuracy assessment site would require, on balance, much more time than travel to an individual plot that has been used to build the classification (because classification plots are usually subjectively and more conveniently located). It is also clear that this methodology fails to take into account that many uses of the map will rely on the interpretive products that many users will employ to define vegetation types (descriptions, field keys) as the best source of field determination of type by the user, when field checking is required.

Method 2 will also produce fairly high accuracy (i.e., trueness to the classification) for the reference sites because an expert in the vegetation of an area is usually the most accurate source of understanding of the vegetation class concepts (Milliken et al. 1998), apart from to the intensive data collection and analysis methods of Method 1. Method 2 is best suited for users who require the truest representation of the map to the classification, especially when funding limits the ability to achieve this during classification and map production. It requires that there be no question as to the accuracy and representativeness of the classification and as to the ability of the field observer to interpret the classification.

Drawbacks to this method are that, while it is an accurate and cost effective, when available, it is generally impractical for NPS Vegetation Inventory thematic accuracy assessment campaigns. The number of experts in vegetation classification for a particular park area is usually very limited and, where they can be found, they are seldom available long enough to spend the required time in the field. Often, the only available pool of experts is the individual or the group who developed the classification for the unit. A consideration of this consequence is that this methodology then would fail to test the interpretative products (e.g., field keys, vegetation

descriptions) as important links from classification concepts to understanding by a more general user audience.

Method 3 has been used often by NPS Vegetation Inventory projects, as something of a compromise between the data intensive needs of Method 1, the expertise required by Method 2, and the availability of staff that are qualified for Method 4. It is more feasible than Method 2 because it can employ non-experts for the field campaign, and can produce reasonably accurate results, but has several drawbacks. It requires that enough field data beyond an assessment of type be made (often, cover of all dominant species in each vegetation stratum), so that the expert in the office can make as many informed assessment in the office as possible. This may include recording cover of multiple species in each stratum over an area as large as 0.25 to 0.5 hectare (rather than the 0.01 to 0.1 hectare typical of classification plots), taking multiple photographs, and other intensive data collection methods. It usually requires a substantial time investment at each site (e.g., 20-40 minutes). When combined with travel time in parks that are large and/or have access challenges, this method incurs relatively high costs. Additionally, there is considerable inefficiency in this method, in that many or most calls in the field will match the map class, and the data are not used. Since two different methods used to determine the reference data values for individual observations, and the method used at individual observations is usually neither explicitly available to the user nor completely independent from the mapping, the response design methodology can be somewhat cryptic and difficult to understand. Related to this concern is the ambiguity as to what method defines the reference data value (for placement in the contingency table), for observations for which values obtained for both reference data evaluation methods show a mismatch with the sample data value.

Method 4 has also been used by NPS Vegetation Inventory projects. It generally is not as high a source of accuracy as Methods 1, 2, or 3. However, the gap in accuracy differences between Method 3 and Method 4 can be minimized by adequate coordination of mapping and ecological classification criteria during the mapping phase and by adequate recruitment and training of observation teams in methodology prior to accuracy assessment, as described in this chapter. Method 4 has the advantage of not requiring experts in the field. It has further cost advantages in that site data collection (5 to 15 minutes per site) is considerably less than for Method 3. It requires far less data review time in the office. Additionally, understanding of the methodology is largely straightforward. For a non-expert field user who needs to verify map accuracy in the field prior to implementing an application, it has the advantage of employing the level of resources and expertise most likely to be available.

Because it controls less for the factors affecting the interpretation of the classification by experts through the type descriptions and the field key and is somewhat less accurate to the truest possible interpretation of the vegetation classification, Method 4 has been referred to as a “correspondence assessment.” This can be an advantage to a user, if the user is a non-expert who must interpret the classification in the field and is likely to be limited to applying these products to the assessment of vegetation, as is the case with a number of applications.

Method 4 has the disadvantage in that in will report lower accuracies than Methods 1, 2, or 3 and controls less for sources of error other than the mapping process (if there is a desire is to learn more about those causes). To be reasonably comparable to Methods 1, 2, and 3, it requires an investment in recruitment, training, and/or supervision of field observers.

Although it has not been used by the NPS Vegetation Inventory, Method 5 (aerial videography) was mentioned by Environmental Systems Research Institute et al. (1994) as a potentially cost effective source of reference data of higher accuracy than the map. While low altitude aerial videography undoubtedly allows more accurate identification of vegetation composition than does the higher altitude aerial photography that usually is employed for the mapping, it would likely be substantially less accurate than ground-based observation because of the general inability to view vegetation features under the uppermost stratum and to identify many plant species other than large woody species with confidence. Because of the "bird's eye" perspective and its inability to detect many ground level features of vegetation, it would have decidedly decreased relevance to the field user, compared to the other four methods. It likely would result in cost savings over ground-based methods in at least the largest parks. It may not be applicable in parks or park areas that have low level overflight restrictions. Because of the program's lack of experience with aerial videography, it remains a largely hypothetical method for the NPS Vegetation Inventory, and is noted here mostly as an method that has been regarded as an acceptable option on the full spectrum along the cost/accuracy gradient of various possible sources of higher accuracy reference data for the purpose of thematic accuracy assessment.

Table 5 compares the five possible methods of higher accuracy sources, using ranked scores of three evaluation criteria. Weighting the three criteria of accuracy, cost, and user relevance relatively evenly, Methods 2 and 4 have the most advantages. Method 1 scores more favorably than Method 5 and nearly as favorably as Method 3, but since cost considerations make it impractical for the NPS Vegetation Inventory, it is disqualified for consideration for program use. Method 2 is the best alternative because it employs the highest reasonable source of accuracy and it requires minimal data collection. However, it is infrequently available to projects. All criteria considered, Method 4 is normally the best readily available approach for the NPS Vegetation Inventory and is described in this chapter. The accuracy difference between Method 4 and the other methods that have been experienced in projects can be narrowed with better planning and coordination between field ecologists and mapping teams (e.g., more attention to mapping models for the ecological classification, more timely production and early testing of field keys (Faber-Langendoen et al. 2007)) and by careful, hiring, training, and early oversight of field observation teams during the assessment.

Table 5. Comparison of five possible methods for sources of higher accuracy for thematic accuracy assessment, using three major evaluation criteria.

METHOD	CRITERION:			VALUE[3]:
	ACCURACY	COST	USER RELEVANCE	
1. Classification Plots	Highest (1)	Impractical (5)	Fair (4)	10.0
2. Expert in the Field	Higher (2)	Low (2)	High (1.5)[2]	5.5
3. User in Field, Expert Office Review	High (3)	Fair (3)	Fair (3)	9.0
4. User in the Field	Good (4)	Lowest (1)	High (1.5)[2]	6.5
5. Aerial Videography	Fair? (5)	Unknown[1] (4)	Low (5)	14.0

Note: numbers in parentheses denote rank order (of five methods) of favorability for the criterion.

[1] - Probably Good for large parks; Low for small parks. [2] - May depend on intended use. [3] – Rank sum of criteria; lower scores have higher value.

Method 3 (Method 4 with a *post hoc* review of limited floristic data collection) is a less desirable methodology for the NPS Vegetation Inventory, but may be employed if the following conditions are met:

-Field data collected for *post hoc* review are limited to a list of dominant and/or frequent species in each stratum (e.g., not species cover estimates, photographs, etc.). This measure is intended to save field time and also direct limited field time toward carefully assessing criteria needed to follow the field key. It will limit the ability to "overrule" reference data values made in the field to the more significant errors. "Reasonable but wrong" calls will often stand. Exhibit I is an example of a field data assessment form that meets these criteria.

-Neither the identity of both the map class (sample data value) nor the reference call that has been made by the field observers can be revealed to the reviewing expert prior to review. Revealing the map class identity is a clear violation of independence in a hypothesis test. Revealing the observers' reference data values will likely bias the reviewer's judgment. The reviewer knows that he/she is reviewing only reference data values that are mismatches with the sample data values and is likely to avoid the same reference data when there is ambiguity. The reviewer must be given an unbiased opportunity to agree with, as well as disagree with the field observers' calls. If the reviewer lacks sufficient information to make a judgment about the best reference data call, then no alternative value need be offered, and the field observers' reference data values will be regarded as correct.

-Data used in mapping (e.g., properties of a vegetation image and other remotely sensed data) may not be used to influence either reference data value.

If these procedures are followed, then only field data for sites with mismatches between the sample data value and reference data value need be reviewed and the expert reviewer's assessment from the [limited] field data may be regarded as the most correct reference data value, when there is disagreement between the field observer and the reviewer.

3.2 Field Methodology

3.2.1 Field Observer Skill Level Considerations

While the NPS Vegetation Inventory accepts Method 4 as a best practice, a minimal amount of preparation and oversight is necessary to assure that observer skill and that the accuracy gap between use of Method 4 and Methods 1, 2, and 3 is minimized, within a reasonable amount of project control.

Field observers should be hired and/or vetted and trained, as needed to represent an adequate skill level to:

Identify all species named in the field key, as well as likely look-alikes. For example, if northern red oak (*Quercus rubra*) is named in the key, then observers should be able to recognize this species whenever it is encountered, to avoid errors of omission of northern red oak in the key. If the similar-appearing scarlet oak (*Quercus coccinea*) is also present in the study area, observers should be able to recognize it and differentiate it from northern red oak, regardless of whether or not scarlet oak is named in the field key (in order to avoid errors of commission in situations in the key that call for recognizing northern red oak). Familiarity with most of the flora is

necessary, although this knowledge level may be less than is required to collect classification plot data.

Estimate cover with reasonable accuracy and precision (repeatability). All field observers should be able to make cover estimates of sufficient precision and accuracy to accurately diagnose key dichotomies that require evaluation of a numerical cover threshold. A good criterion for training and evaluating observers in this skill is that observers should be able to estimate continuous cover of any species to within 20% (greater than or less than) of the true continuous cover value of the species in a given observation area of MMU size. As a practical training exercise in testing and refining observer skills, the true cover value may be estimated by the instructor or may be estimated as the mean value of several independent trainee estimates.

Accurately estimate the extent of the observation area (as defined by the MMU size for the map class). During training and/or early in the campaign, even experienced field observers should practice estimating observation area extents by measure the prescribed radius length for all map class MMU sizes (Table 3) outward from the plot center (site position) with a measuring tape, for a circular observation area of MMU size. For training purposes, it is often preferable to do this in four directions from the plot center, with each measured radius about 90 degrees from the two nearest measurements. Once observers have gained familiarity with these radius distances, the extent of the observation area may be estimated visually or paced. However, where the presence of a diagnostic species or condition occurs sufficiently near the visually estimated boundary to create uncertainty as to whether a field key threshold has been exceeded or not, the observation area the observer should either pace or measure the radius distance with a tape in this direction in order to resolve this question.

Consistently identify situations that represent an overly heterogeneous field observation area. (See Subsection 3.2.6). Caution should also be employed that observers do not attempt to "over think" the situation (e.g., to attempt to delineate a boundary between two associations within a common alliance) or to delineate distributions of diagnostic individual species that do not coincide with ecological boundaries.

Efficiently and safely navigate and travel off-trail in the terrain to be expected. Except in parks with good road and trail access and moderate distance and terrain challenges, field work in thematic accuracy assessment requires that observers have a higher level of physical fitness, backcountry travel skills, equipment skills, and judgment than might be satisfactory for classification plot data collection, for which observation locations are not as constrained by statistical requirements and may be located nearer to access corridors.

A helpful training exercise that reduces among-observer variability is to have observers key out the same observation area and arrive at the most likely vegetation type independently from one another (e.g., write their answer down before declaring it). Their answers may be compared to one another and/or to the trainer's answer. Causes for discrepancies should be discussed. Consistency between trainees but a discrepancy between the trainees and the trainer may suggest problems with the key that were overlooked during field key creation or verification (or with the trainer!). Discrepancies between an individual trainee and the majority of answers may suggest issues with trainee performance.

If field observer skills in species identification are still in question following a training or evaluation period, it is acceptable to ask observers to record dominant, frequent, and/or diagnostic plant species (by presence) in each stratum. The purpose of this is to allow the accuracy assessment campaign manager to catch egregious identification errors that would cause serious field keying errors (e.g., recording species that do not occur in the study area), rather than to take the time to provide enough data to allow re-assignment in the office of field calls (e.g., cover estimates of these species). The practice of recording species presence may be dropped once the manager has confidence in the observer(s) ability to identify species.

In a lengthy and committing field campaign, the field campaign manager should review and assess the data and or observers frequently (e.g., weekly) in the early stages, including performing an early accuracy analysis (e.g., calculating the correspondence rate of the field calls with the corresponding map classes and comparing this with the accuracy rate derived from the map validation for large discrepancies). Unless the observers initially are working in areas that have significantly lower accuracy than the map as a whole, a lower accuracy rate than that of the validation may suggest performance problems with the field observers that can be addressed early, if needed.

3.2.2 Safety and Efficiency

Execution of the Method 4 field methods generally requires only one observer at a site. However, compared to plot data collection, accuracy assessment field work usually requires much more cross-country travel away from established travel corridors such as trails and roads and usually requires travel to more distant sites, exposing observers to more risk of mishap. Additionally, the location of individual target vegetation types may coincide with potentially difficult or hazardous terrain, such as steep or unstable slopes. Thus, for safety, it is often desirable to send observers to the field in at least teams of two, particularly in remote areas and/or in areas where the possibility of crime may be a concern.

When teams of two persons are necessary, it is more efficient to split the more skilled members (e.g., botanists) into different teams and pair each botanically skilled person with a less skilled team member (e.g., a student intern), since the second person's primary purpose is more for safety than to make technical observations. The backcountry travel abilities of the less botanically skilled member should be relatively well matched to that of the primary observer; otherwise, the pairing could create inefficiency or an adverse safety situation. This approach has the dual advantage of (1) efficiently allocating the expertise required to collect the data and (2) affording less skilled employees the opportunity to learn more.

Particularly when teams of a single person are used, a daily travel plan should be submitted to a responsible person. The plan should include the GPS positions of the sites to be visited, clear instructions on check-in at the end of the day, and actions to be taken the event the check-in does not occur. The plan should establish methods of routine and emergency contact procedures (e.g., cellular phones, radios or satellite phones, emergency personal locator beacons, where applicable).

In some cases, teams have traveled in groups of two persons or more and split into teams of a single person for short arrays of sites and then checked in at a "rally point" at a predetermined time. This approach can afford a reasonable balance of project efficiency and safety. It especially

can work well if visibility is sufficient to allow individuals to remain largely in sight of one another (e.g., in arid areas of sparser vegetation).

When “out and back” (from a road or trail access) arrays of sites are contemplated in a single trip (e.g., in a single field day), it is generally most efficient to visit the farthest site in the array first (bypassing closer sites) on the trip out and progressively visiting the less distant sites on the return trip. If the array cannot be finished in a single trip, the return trip then will not require another long approach from the access point. This approach can also enhance safety, as observers will less likely find themselves far from an access point late in the day.

Particularly in steep terrain, some potentially hazardous areas might be eliminated *a priori* from the inference area through GIS analysis. For example, the inference area for rock outcrop types might exclude those stands that occur on sites that exceed a safe slope angle (e.g., cliffs).

The importance of exercising sound safety judgment should be stressed to field observers. Not all potentially hazardous field situations can be recognized in the sample design. No observation is worth a significant risk of an accident or becoming lost, and observers should understand that their own judgment about what is safe for them always should overrule the need to access individual sites.

Individual parks may have unique access and safety challenges. The park units and/or contractors or cooperators conducting the work may have specific requirements and/or recommendations for safety in field operations. Other than the general considerations common to all thematic accuracy assessment field campaigns that are discussed here, planning for access and safety challenges is best done in the context of each individual project. Early communications between field staff and local managers about expectations is important to a successful field campaign.

3.2.3 Basic Navigation

Observers usually will navigate to the observation site using a GPS. GPS positional accuracy requirements should follow sampling design and analysis considerations, GPS model capabilities and methods, and field limitations (e.g., degradation of GPS accuracy under certain situations, such as heavy vegetation cover and proximity to a GPS signal blocking feature such s a cliff).

When nearing the site, the observer will simultaneously be watching the GPS screen and walking. Special care should be taken in placing one’s footing. It is also useful to plan one’s approach route to the site when far enough away to anticipate travel difficulties (thick brush, drop-offs, etc.). The quickest and safest path to the site may not always follow the shortest distance to it. If the site itself proves to be unsafe or impossible to approach, it may be possible to record the nearest position and the direction and distance to the site (offset) and still observe the prescribed observation area around the site adequately enough to make a field call (this often occurs if the type is relatively distinct or otherwise easily keyed). If the original observation area cannot be adequately observed from this alternate position, the alternate position may be recorded and the alternate observation area around this position may be observed. If, after GIS comparison of the alternate position with the map class, the alternate observation area is found to be positioned entirely within a single map class (even if not the originally targeted class), the observation may be used in analysis with the map class recorded by the *post hoc* comparison as the sample data value (even if this is not the originally targeted class). If the alternate position

indicates that the observation area has included more than one map class, the project *a priori* rules for determining the map class (sample design value) in these cases should be employed.

Upon arrival at the plot center (site coordinates), the observer should immediately mark it (e.g., with a stake or flagging) to make relocation easier when walking around the observation area. It is recommended that equipment not needed for the observation (e.g., packs) be placed here to avoid losing track of it. If the observer becomes disoriented, they may always re-navigate back to the plot center and gear from the site coordinates.

3.2.4 Field Navigation Supplements and Retaining Independence from Mapping

As an aid to field navigation, especially for locating sites within small and/or linear polygons, field observers have been provided with copies of the map (either in hard copy or digital form on a portable field computer with display), with the map class boundaries delineated but unlabeled. Improvements in GPS real-time accuracy since 1994 and acceptable methods of buffering (see Subsection 2.3.4) have greatly reduced the need for this. In addition, this practice introduces the concern about biasing some field calls by revealing how the mapper treated the values of some nearby sites (as being the same or different from one another). If a field observer sees that multiple sites are enclosed in the same mapper's polygon, they may be influenced (biased) to making some of the field calls to being also consistent within the polygon, whenever there is some ambiguity about the field call. Conversely, the field observer may be influenced to make different field calls for sites in adjoining polygons with a map class boundary between them, even though such a situation may not be the best field representation of the vegetation. Of course, it cannot be assumed that the mapper's delineation of boundaries is accurate and that the polygons represented actually are homogeneous, and a premise of the NPS Vegetation Inventory is that sites are attributed in the field independent of the mapping process. This potential problem is especially acute in smaller parks, where many or most polygons may be populated with accuracy assessment sites and many polygons will also require multiple sites.

In most cases, providing the field teams with an image (and other data that might help navigation to and location of sites, such as elevation contours and stream locations), but not the mapper's rendition of stand boundaries (polygon boundaries), will address positioning concerns that cannot be alleviated by sampling design. In rarer cases, individual polygons represent stands that are at or near the MMU size, but must be used for locating sites because of overall class rarity. If these polygons are to receive only one site, the boundaries of these individual polygons (only) may be delineated on the field map as an aid to accurate site placement within the target map class. In cases in which multiple sites within a linear polygon are to be visited, showing incomplete boundaries (in the vicinity of the sites only) is acceptable.

Appendix D provides examples of guidance for these situations.

3.2.5 Determination of Vegetation Type in the Field

Determining the reference data value of an observation site can be done in several ways. Using the field key that is developed for the project from local observation of vegetation as a primary arbiter to make determinations as to vegetation type occurring at a site as observed in the field has several advantages. Unlike a reading of a number of all possible vegetation type descriptions or expert judgment of an individual, use of a field key more or less forces all users to apply the same relatively objective criteria in making field determinations. This helps to assure that the

process would have a significant degree of repeatability among users (making the accuracy rate recorded a more meaningful expectation for future map users). Also, by directing the user to focus on the most differential criteria between vegetation types, it is a relatively efficient means of guiding the user relatively quickly toward a correct (or at least, good) diagnosis and away from less plausible answers.

As ecologists have cautioned, it is true that "the key is not the classification," in the sense that the field key applies relatively few technical and/or diagnostic criteria out of a series of many classification criteria. Even the best crafted field key cannot account for all vegetation variation that might be encountered. As a means for a non-expert to apply rapidly the best criteria for a diagnosis of an individual vegetation stand, any key will trade some diagnostic accuracy for the sake of efficiency and repeatability. Observation teams should train with the field key prior to applying them to the accuracy assessment in order to apply key criteria relatively consistently within the project. The vegetation descriptions be consulted, when necessary, in order to resolve ambiguous situations (decision points) in the key.

A field key that is useful for thematic accuracy assessment should make it possible for a user to key to any vegetation type identified during the project and not only those types that are represented by map classes in a 1:1 relationship. Vegetation types that were not mapped as individual classes because the mapper believed them to always occur in stands smaller than the defined minimum mapping unit (MMU) size may be encountered in stands that are larger than the MMU size as errors of omission. Furthermore, where several types are always mapped together as a part of a map complex or at a higher hierarchical level of vegetation classification, identification of accuracy assessment sites in the field to the finest classified level, (rather than only to the finest mapped level) will provide the map user with useful information about the relative contributions of each of the finer vegetation classes to the coarser map class (Czaplewski 2003).

Some map classes that represent cultural vegetation (in the sense of Federal Geographic Data Committee 2008) or non-vegetated land cover that might not be included in the sample, might be encountered as errors of omission within natural and semi-natural vegetation map classes. Therefore, it is a good practice to allow them to be identified in the field key, even if these are not represented by map classes being assessed. They often are differentiated from natural and semi-natural vegetation at the start of the field key.

3.2.6 Moving or Reshaping Observation Area

The practical necessity of defining the accuracy assessment inference area so that the sample data value for each single observation in the sample is unambiguously homogeneous (i.e., that each observation site has only a single map class membership) is discussed in Chapter 2. As a practical matter, it is equally necessary for the field call value of each observation to be reasonably homogeneous (i.e., that the field observation area does not cover more than one distinct vegetation type). Field keys are designed to work in relatively homogeneous stands. Note that homogeneous means internally homogeneous (at the stand or individual observation level); it does not mean "typical" of the vegetation type.

Despite the employment of measures to ensure sample data homogeneity within the observation area (see Subsection 2.3.4), excessive thematic heterogeneity in the reference data within the

observation area can occur. This is because unmapped areas of vegetation types that are different from that represented by the mapper in a single polygon can occur. These may include inclusions (stands of types different from the majority type in the polygon and that are smaller than the MMU size). These are rightfully ignored by the mapping but may still affect the field key performance. They may also include errors of omission (stands of types different from the majority type but larger than the MMU size and, thus, should have been mapped as separate stands).

For example, a field observer might navigate to a site (represented by a point with specific coordinates) and evaluate a 0.5 hectare circular observation area around the site. The site may be dominated by (e.g., 75% of the site is occupied by) dry upland meadow vegetation, with no wetland plant species present, but there may be a distinct boundary within the observation area between the dry meadow vegetation and a wetland vegetation type that covers 25% of the observation area and has nearly 100% cover of wetland plant species over that 25%. A field key criterion may specify that a wetland type must be selected, if wetland species are present in significant quantity, which would be the case if they cover 25% of an observation area at 100% cover. Thus, the key answer for the site would be a wetland type, even though most of the observation area and possibly the site location itself is a non-wetland type.

Where there is a distinct boundary between types in the observation area, the field observer should be given some latitude to moving the site to accomplish a reasonable amount of homogeneity in the observation area. Two considerations are in order.

First, because the observer should be unaware of the locations of the map class boundaries, the site location (observation plot center) should be moved the minimal distance in order to achieve this homogeneity. Because vegetation boundaries as defined by individual polygons that represent map classes and vegetation boundaries as seen in the field do not always correspond to one another, there is some (small) risk of unknowingly relocating the observation area partially into a different map class, which would create ambiguity as to map class value of the site, even as it eliminates field observation ambiguity for the site. The relocation should be done despite this concern because, in most cases, this will not affect map class value. For cases in which this action does cause map class value to be ambiguous, the accuracy assessment analysis should have a set of *a priori* rules to resolve the situation (e.g., one might record the map class represented by the observation area center as the most probable class, one might record the map class occupying most observation area as the most probable class, one might discard the site from analysis, etc.). If the site relocation places the site unambiguously into a different map class than that to which the site was originally assigned, the site is simply assessed as an additional observation within the different map class, rather than within the original map class to which it was assigned. The importance of recording the observation point center in the field, rather than relying on the original navigation position, becomes clear here.

Secondly, in order to balance analytical accuracy with efficiency and consistency, the field observers should be given some guidelines on how distinct a boundary between types must be in order to move the site (minimally) to a more homogeneous location. As a general rule, boundaries between two different vegetation types that are distinct only at the NVC association or NVC alliance level (Federal Geographic Data Committee 2008) usually will appear as being quite gradual on the ground and locating them will be imprecise. In these cases, if the observer

recognizes a transition at all, it is still best for them to key the vegetation in place, which is the most desirable course, from the perspective of the need to reduce ambiguity in sample data value. Boundaries between vegetation types that are distinct at the NVC Macrogroup level (Federal Geographic Data Committee 2008) will often be more distinct, and uniting two very different stands in a single observation area (e.g., mixing a wetland type stand with an upland type stand) may produce spurious results in keying or otherwise determining a single (most probable) vegetation type. In such a case, minimal relocation of the site would be warranted.

For situations in which two or more very different vegetation types are perceived within the prescribed observation area around the original site location, *a priori* rules are needed as to which vegetation type should be evaluated. Generally, the best rule of thumb is for the vegetation type occupying the point representing the original (navigated to) site. If this stand does not occupy an area as large as or larger than the MMU size for the map class, then the vegetation type representing the next nearest stand to the site may be selected and the site moved (minimally) so that the observation area is completely within a single vegetation stand, and so forth. Splitting the observation area is problematic and generally is not recommended because (1) each part of the observation area that is attributed to an individual vegetation type will represent an area smaller than the prescribed MMU size (and which is less likely to sufficient attributes to key the vegetation type) and (2) it is left ambiguous as to which vegetation type (field call) is actually represented at the site (point). Splitting the observation area into multiple vegetation stands and recording the percentages of the area occupied by each type might be warranted in cases in which map classes that are evaluated at a particular MMU size are usually occupied by a series of recognizable vegetation types that typically occur in patches smaller than the MMU size (e.g., a fine-scale vegetation mosaic). This situation most often involves wetland vegetation, such as occurs on riparian bars and pond shores. Splitting the observation area for these types can provide an estimate of relative abundance of small patch vegetation types within a larger generalized map class. However, this situation is germane to a relatively small amount of vegetation and should be used sparingly, if at all, for thematic accuracy assessment.

An example of a decision tree (key) that may be followed in cases of excessive heterogeneity found in the field observation area as viewed from the original site (waypoint) is provided in Exhibit H. The tree may be amended or decisions may be further specified in order to fit project decision rules). It is intended to serve as a training or learning tool for field observers and is, therefore, laboriously explicit. Once observers have practiced the decision process enough under vegetation conditions at a project area, the decisions usually can be made quite rapidly.

3.2.7 Recording Field Position (GPS methods)

Because of the occasional need to relocate a thematic accuracy assessment site and the possibility that this relocation may place the observation in a different sample data class, it is imperative that field observers record the site position by a GPS receiver in the field, in order to record the site that was actually observed with the highest level of positional certainty. A GPS receiver that can average multiple positions (e.g., 30-100) is recommended. Coordinates (northing/easting or latitude/longitude) should be recorded to an appropriate level of precision required to confidently place the observation within map class in which it actually was recorded, rather than in the map class in which it was intended, in cases in which the two may differ from each other (e.g., as in Subsection 3.2.3). Additionally, an estimate of the positioning error (usually, as estimated and as displayed by the GPS receiver) should be recorded, in the event that

this error is large enough to create ambiguity about the map class membership of the observation. There are a number of ways to deal with this type of uncertainty; it is probably best to try to minimize it (see Subsection 2.3.3). Projects should have an *a priori* plan as to how much map class membership uncertainty is acceptable and what to do in cases of sample data value ambiguity.

3.2.8 Reporting Field Data

Field observers will report their findings on a field assessment form. At a minimum, the field assessment form should contain the following information:

1. Name of the observer(s).
2. Name of the park.
3. Date of the field visit.
4. Name or identifier of the sample site location.
5. GPS coordinates of the sample site in the field and the method by which the coordinates were determined (e.g., real-time Differential GPS, Wide Area Augmentation System (WAAS)-enabled GPS, multiple uncorrected GPS positions averaged, GPS with offset, single uncorrected GPS position, estimated position from 7.5 minute U.S. Geological Survey quad sheet, etc.).
6. Estimated maximum (e.g., 90% to 99% confidence level) error of the field GPS positioning (usually given by the GPS unit).
7. Shape and size of the observation area used at the site, if different from the standard shape or the prescribed size for the site.
8. Methods of site relocation, when necessary.
9. Type of vegetation at the site (usually from the field key).
10. Description of special conditions that may have been encountered (e.g., field key did not work, unusual vegetation conditions that may affect the identification of type, secondary field call (if significant ambiguity about primary field call), obviously undescribed vegetation type, etc.).

Minimally, data for items 1-9 should be entered into the PLOTS database, if an automated field data entry form is not used.

3.2.9 Missed Sites

It is possible that some thematic accuracy assessment sites will not be visited due to unforeseeable circumstances (e.g., safety concerns, lack of time, reassignment of an observation to an alternate map class due to problems positioning it within the original class (as in Subsection 3.2.3), lost data, sites overlooked). Previous projects have guarded against a shortfall in the prescribed map class sample sizes by “overstocking” the sample with redundant or replacement

sites to assure that the prescribed sample size (see Subsection 2.5.2) is achieved. Good planning should minimize sites lost for these reasons; nevertheless, some map classes with less than a full complement of their prescribed observations may be anticipated. Rather than add replacement sites, it is suggested that the prescribed sample size be attempted and, if a sample size of at least 90% of the prescribed sample size for the map class (e.g., 27 sites out of a sample size of 30 under Scenario A) is actually visited and recorded in the field, that this lesser sample size be deemed adequate for the class. If the field campaign fails to achieve 90% of the prescribed sample size, then reasonable additional efforts should be made to revisit sites or replace them.

4.0 Data Analysis

This chapter provides the definitions of the measures of accuracy used in thematic accuracy assessment for NPS vegetation mapping. It provides the logic of estimating those measures from field-collected accuracy assessment data, and of estimating the uncertainty of those estimates as approximate confidence intervals. The computational steps are provided for use with the most common sampling design for NPS vegetation mapping accuracy assessment: explicit choice of sample sizes per map class (Subsection 2.5.2, Table 4), and simple random sampling of observations within each mapped class. Computational spreadsheets are provided (see http://science.nature.nps.gov/im/inventory/vg/index.cfm) are provided in order to assist in these computations for this common case. Note that if two-stage or cluster sampling is used (as noted in Section 2.4), the logic of estimating parameters from the accuracy assessment data is the same, but different computations are required, and the user should consult the relevant literature, e.g., Stehman (1997, 2009) for cluster sampling.

Table 6. Notation used in Chapter 4.

SYMBOL	DESCRIPTION
i	Row subscript or map class (sample data class) (denoted in lower case)
J	Column subscript or field vegetation class (reference data class) (denoted in capitals)
	Population (Complete) Contingency Table
p_{ij}	Proportion (fraction) of the total area of inference with field vegetation class J and map class i
P_{i+}	Proportion (fraction) of the total area of inference in map class i
P_{+J}	Proportion (fraction) of the total area of inference in field vegetation class J
$\hat{P}_{+J}$	Estimated proportion (fraction) of the total area of inference with field vegetation class J
	Sample (AA observations) Contingency Table
N	Total number of observations (accuracy assessment sites) in sample contingency table
n_{iJ}	Number (raw count) of observations (sample units) classified as map class i and as field vegetation class J.
n_{i+}	Total number (raw count) of observations (sample units) in row i (map class i)
N_{+J}	Total number (raw count) of observations (sample units) in column J (field vegetation class J)
	Areas (in hectares)
$A_{total}=A_{++}$	Total area in inference area (all map classes combined) (known from GIS).
A_{i+}	Total area (in hectares) in map class i (known from GIS).
$\hat{A}_{+J}$	Estimated total area (in hectares) of map class J had all map objects been classified using the reference data (field) methodology.

4.1 Accuracy Assessment Data

The fundamental data from an accuracy assessment are a list of accuracy assessment observations, with a geographic position (northing/easting or latitude/longitude), a sample data (map class) value representing a vegetation type, and a reference data (field call) value representing a vegetation type, for each observation. The reference data values for vegetation types are considered to be accurate. In a GIS context, the individual field observations are often represented as a point layer having a field in an attribute table that identifies the reference data value. The map classes are usually represented as a polygon layer, with a field identifying the map class. Assuming that both data layers have sufficient positional accuracy, the positions of the field observation site layer may be plotted onto the map class layer, and a spatial join may be performed to create a table of the accuracy assessment observations attributed with fields for both their field vegetation type (from the point layer) and their mapped vegetation type (from the polygon layer).

Ideally, where the map classes represent the classified vegetation types in a 1:1 relationship, the list of vegetation types would be the same as the list of map class types. For these map classes, a match between the values of both the vegetation type field and the map class field constitute an accurately mapped site. For some vegetation types, the mapper may represent multiple vegetation types in a single map class, as in map complexes in which multiple vegetation types normally occur together or in similar settings and are clearly distinguishable on the ground, but are not distinguishable from a remote sensing perspective. For map classes which the mapper has asserted to represent multiple vegetation types that occur in a one to many hierarchical relationship (as in a map complex), any field observation that has been asserted (*a priori*) to have membership in the map class is counted as a correctly mapped site. Conversely, in the much less common situation in which multiple map classes represent a single vegetation type (e.g., two map phases of the type), a field call of that vegetation type is counted as a correctly mapped for either map class. This is equivalent to lumping either field types to a single map (complex) type, or lumping multiple map classes to a single field type, in order to produce a 1:1 relationship between map and field classes for the purposes of the accuracy assessment.

4.2 The Sample Contingency Table

An initial summary of the accuracy assessment data is a contingency table of counts of observations, with sample data values (vegetation map classes) as rows and reference data values (vegetation types as identified on the ground) as columns, as illustrated in Table 7. Table 8 (Section 4.8) and Table 15 (Exhibit F) provide numerical examples of a sample contingency table. In notation of data terms, lower case subscripts refer to the sample data value or row; upper case subscripts refer to the reference data value or column (Table 6). The value in the cell n_{iJ} is the number of accuracy assessment observations mapped in class (row) i that were found to be of class (column) J in the field. The values in the shaded cells along the diagonal represent counts for correctly classified observations, where the reference data (column) vegetation type matches the mapped vegetation type (row) value.

This contingency table is often referred to as a confusion matrix, misclassification matrix, or error matrix (Congalton and Mead 1983, Congalton, Oderwald, and Mead 1983, Story and Congalton 1986, Rosenfield and Fitzpatrick-Lins 1986, Congalton 1991, Czaplewski 2003, Brewer et al. 2005, Congalton and Green 2009). However, those names are also applied to the

population contingency table, which has the same layout but reflects the entire mapped area, not the accuracy assessment sample.

Table 7. Sample contingency table for five sample and reference data vegetation classes. n_{iJ} is the proportion or fraction of the total area that is actually vegetation type *J* and mapped as class *i*.

	Reference Data (Observed Vegetation Types on Ground)					
Sample Data (Map Classes)	**A**	**B**	**C**	**D**	**E**	**Row Total (n_{i+})**
a	n_{aA}	n_{aB}	n_{aC}	n_{aD}	n_{aE}	n_{a+}
b	n_{bA}	n_{bB}	n_{bC}	n_{bD}	n_{bE}	n_{b+}
c	n_{cA}	n_{cB}	n_{cC}	n_{cD}	n_{cE}	n_{c+}
d	n_{dA}	n_{dB}	n_{dC}	n_{dD}	n_{dE}	n_{d+}
e	n_{eA}	n_{eB}	n_{eC}	n_{eD}	n_{eE}	n_{e+}
Column Total (n_{+j})	n_{+A}	n_{+B}	n_{+C}	n_{+D}	n_{+E}	$n_{++}=n_{total}$

The row totals in the sample contingency table are the numbers of accuracy assessment observations collected for each mapped class. They reflect the deliberate choices of sample sizes for each map class stratum in the accuracy assessment sampling design (see Subsection 2.5.2 and Table 4). Note that these totals differ proportionally from the row totals in the population contingency table, which are the fractions of the target area mapped as each class. These fractions of the target area mapped as each vegetation type are known exactly from the GIS summarization of the map. If A_i is the inference area of map class *i* (calculated by a GIS) and *A* is the total area of the total inference area of all map classes (also calculated by a GIS), the proportion of the target area mapped as class *i* is A_i/A.

In the numerical example of Table 8, the sample sizes selected follow the recommendations in section 2.5.2, possibly minus sites that could not be sampled or were sampled in alternative classes because of positioning error. Thus, the row totals for each of the map classes *a*, *b*, *c*, and *d* are approximately 30 observations. Map class *e* has many fewer observations (11), possibly because it is a rare class (covering a limited amount of area) with fewer opportunities for placing spatially independent observations.

Unlike the row totals, the column totals in the sample contingency table have no clear interpretation. The column totals are the number of accuracy assessment observations found to belong to each vegetation class. If the accuracy assessment observations were selected as an equal probability simple random sample, these column totals, divided by the total number of accuracy assessment observations, would estimate the fraction of the target population or inference area in each vegetation class. However, the numbers of accuracy assessment observations for each map class were chosen in the sampling design and are not proportional to the prevalence of each map class in the target area (population) as they are in the population contingency table. Doubling the number of accuracy assessment observations in any map class (e.g., selecting 60 instead of 30 accuracy assessment observations in map class a) would by itself dramatically change the relative values of the column totals.

Table 8: Numerical example of a sample contingency table for five hypothetical sample data and reference data vegetation classes.

	Reference Data (Observed Vegetation Types on Ground)					
Sample Data (Observations Selected Within the Map Class)	**A**	**B**	**C**	**D**	**E**	**Row Total (n_{i+})**
a	20	7	1	0	2	30
b	7	15	8	0	0	30
c	2	0	27	0	1	30
d	1	4	7	8	8	28
e	1	0	1	1	8	11
Column Total (n_{+J})	31	26	44	9	19	**129**

Because the measures of accuracy are defined from the population contingency table, the statistical analysis of an accuracy assessment uses the sample contingency table, the data from the accuracy assessment samples, to estimate the population contingency table, including the proportion of area in each field vegetation class, and then computes estimates of the various measures of accuracy or error rates.

The previous NPS Vegetation Inventory guidance (Environmental Systems Research Institute et al. 1994) suggested calculating users', producers', and overall accuracies from the sample contingency table of observation counts. This practice did not account for the unequal probability sampling imposed by the stratification by map classes. While appropriate for users' accuracies, it will give inaccurate results for producers' accuracies and overall accuracy.

4.3 The Population Contingency Table

Measures of accuracy are defined on the population contingency table as shown in Table 9 (Stehman and Czaplewski 1998) and in numerical examples in Table 13 (Section 4.8) and Table 16 (Exhibit F). As in the sample contingency table, rows are defined by the sample data values (lower case subscripts), and the columns are defined by the reference data values (upper case subscripts in this table). However, the values in each cell are the proportion of the target area in the corresponding true and mapped vegetation classes, rather than the raw count of observations. The row sums p_{i+} are the proportions of the total area mapped as type i. The column sums p_{+J} are the proportions of the total area that are truly class J, which is not known, but can be estimated from the reference data values.

The population contingency table should be reported along with the sample contingency table, as a summary of between-class error relationships and class accuracy statistics.

Table 9: Population contingency table for five hypothetical sample data and reference data vegetation classes. p_{iJ} is the proportion or fraction of the total area that is actually vegetation type *J* and mapped as class *i*.

	Reference Data (Observed Vegetation Types on Ground) (*J*)					
Mapped Classes of AA observations (*i*)	**A**	**B**	**C**	**D**	**E**	**Row Total (p_{i+})**
A	p_{aA}	p_{aB}	p_{aC}	p_{aD}	p_{aE}	p_{a+}
B	p_{bA}	p_{bB}	p_{bC}	p_{bD}	p_{bE}	p_{b+}
C	p_{cA}	p_{cB}	p_{cC}	p_{cD}	p_{cE}	p_{c+}
D	p_{dA}	p_{dB}	p_{dC}	p_{dD}	p_{dE}	p_{d+}
E	p_{eA}	p_{eB}	p_{eC}	p_{eD}	p_{eE}	p_{e+}
Column Total (p_{+J})	p_{+A}	p_{+B}	p_{+C}	p_{+D}	p_{+E}	$p_{++}=1$

4.4 Definitions of Measures of Overall Accuracy

The overall accuracy is defined as the proportion of the total area (all individual map classes pooled) that is correctly mapped, or equivalently, the probability that a random point in the target area is classified correctly by the map. This overall accuracy is simply the sum of the shaded $p_{i=J}$ terms along the diagonal divided by the total area to produce a proportion or percentage:

$$Overall\ Accuracy = \left(\sum^{i=veg\ classes} p_{iI} \right) \Big/ p_{++}$$

Note that while the maximum overall accuracy when all areas of the target population are mapped to the correct vegetation class is 1 (100%), under a completely random assignment of locations to map classes, overall accuracy would be greater than 0, as due to chance some locations would be assigned to the correct map class. Under completely random mapping, the amount of chance agreement is the sum of the product of the corresponding row and column sums:

$$Chance\ Agreement = \sum_{i=J}^{veg\ classes} p_{i+}p_{+J}$$

The kappa coefficient is another measure of overall accuracy that is scaled to reach a maximum of 1 (100%) under perfect accuracy, but to have a value of 0 under random mapping, by subtracting off the expected chance agreement from both the numerator and the denominator:

$$kappa = \frac{Overall\ Accuracy - \sum_{i=J}^{veg\ classes} p_{i+}p_{+i}}{1 - \sum_{i=J}^{veg\ classes} p_{i+}p_{+i}}$$

4.5 Users' and Producers' Accuracies Measures for Individual Map Classes

Per-class accuracies also can be estimated from the data, and often are more informative to a user if they are differentiated into producers' and users' accuracies (Story and Congalton 1986). There

are two forms of class-specific accuracies, one based on row classes and one based on column classes. **Users' accuracy** is conditional on mapped classes, and thus rows. It is defined as the probability that a location mapped as class *i* is in fact class *I*: p_{iI}/p_{i+}. Users' accuracy 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: if the user goes to a location mapped as class *i*, what is the probability it is in fact vegetation class *I*? **Producers' accuracy** is conditional on the true vegetation class in the field, and thus is based on columns. The producer accuracy for class *J* is the probability that a location of vegetation class *J* in the field is mapped as class *j*: p_{jJ}/p_{+J}. Producers' accuracy is so named because it may inform producers of remotely sensed and mapped data how readily a class may be detected by mapping whenever it occurs on the grounds.

Similarly, there are two forms of class-specific error rates. **Errors of commission** are conditional on the map classes: the probability that a location mapped as class *i* is in fact not class *I* when observed on the ground. The probability or rate of errors of commission is $(1 - p_{iI})/p_{i+}$, or $p_{i\neq J}/p_{i+}$. Conversely, **errors of omission** are conditional on the field classes: the probability that a location of class *J* in the field is not mapped as class *j* is $(1 - p_{jJ})/p_{+J}$, or $p_{J\neq i}/p_{i+}$. Note that the users' accuracy p_{iI} plus the sum of the errors of commission for that map class $\sum_{J\neq i}^{J} p_{iJ}$ sum to 100% of that map class p_{i+}.

Despite the implications of their names, both users' and producers' accuracies can be important to map users, from different management perspectives. Users' accuracies are often of importance to a user who is contemplating a visit to a particular mapped site for an application and who may be concerned about the consequences of not finding the vegetation type that is mapped (e.g., the costs of a fruitless trip). Producers' accuracies are often of importance to managers who wish to understand the true area of a type that may be of management or conservation interest, since the producers' accuracies will estimate the true areas of these types that may have been missed in mapping. These users will often not be visiting a site. Users of a map and accuracy assessment data should note that neither type of accuracy can predict with site specificity where errors have occurred; they can only estimate the probability of an error type at a given site.

4.6 Computations of Accuracy Measures under Stratified Random Sampling

This section explains the computation of the estimates of measures of accuracy for the common case of sampling that is stratified by map class. Czaplewski (2003) (page 133) offers slightly different forms of these equations. Additionally, the NPS Vegetation Inventory web site (http://science.nature.nps.gov/im/inventory/vg/index.cfm) provides programmed spreadsheets that perform these computations after the data for the sample contingency table, and the areas mapped into each class are entered.

4.6.1 Computation of Measures of Users' Accuracies

As long as simple random sampling within each map class was used to select accuracy assessment observations, the users' accuracy for a map class can be estimated directly as the fraction of the accuracy assessment observations in that map class that were found to have the correct vegetation class in the field:

$$est(User's\ Accuracy\ class\ i) = \hat{U}_i = n_{iI}/n_{i+}$$

The standard deviation about that estimate is based on the binomial distribution:

$$var(\hat{U}_i) = U_i(1 - U_i)/n_{i+}$$

Approximate 90% confidence intervals are based on the z distribution:

$$90\%\ c.\ i. = U_i \pm\ 1.645\sqrt{U_i\,(1-U_i)/n_{i+}}\quad + 1/2n_{i+}$$

The term $1/2n_{i+}$ is a correction for continuity (Snedecor and Cochran 1967). This correction accounts for the fact that 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.

Such confidence intervals would be expected to contain the true value (from the population contingency table) of the users' accuracy 90% of the time. 1.645 is the critical z value for 90% confidence intervals. If other widths of confidence intervals are desired, the appropriate values for z can be substituted from Table 10.

Table 10: Values of z_α for two-sided confidence intervals at selected confidence levels (from Zar 1996)

Two-sided Confidence Level	α	z_α
99%	0.01	2.326
95%	0.05	1.960
90%	0.10	1.645
85%	0.15	1.440
80%	0.20	1.282

The simplicity of the computations for users' accuracy is the result of the recommended sampling design being optimized for estimating users' accuracy. Simple random sampling (independent of mapped class) would have made the computations equally simple for producers' accuracy, overall accuracy, and Kappa. However, as noted in Subsection 2.5.2, rare vegetation classes would have few if any accuracy assessment observations, and thus have very poor estimates of users' accuracy. Stratification of the accuracy assessment observations by mapped class increases the ability to estimate users' accuracy for less common vegetation classes, at the cost of more complicated computations and poorer estimates for the other measures of accuracy.

4.6.2 Computation of Measures of Producers' Accuracies

Producers' accuracy for a map class could be computed as weighted averages of the counts within each column, with weightings based on the fractions of the target area mapped as each class and the number of observations sampled in that map class: $weight_i = (A_i/A)/n_{i+}$. However, all remaining computations are built on these weightings, so it is simpler to estimate the population contingency table directly from the sample contingency table, and use those estimates in the remaining computations.

The fraction of the target area with map class *i* and field class *J* is estimated as the product of the fraction of the target area mapped as class *i* times the fraction of map class *i* found to be field class *J* in the sample contingency table (n_{iJ}/n_{i+}). Note that the proportion of the target area mapped as class *i*, are the row totals A_i/A, and are known exactly from the GIS summarization of the map. Therefore,

$$\widehat{p_{iJ}} = (n_{iJ}/n_{i+})(A_i/A)$$

These estimates are a generalization of the estimates of users' accuracy from the special case of $i=J$ to all possible combinations of *i* and *J*. (In fact, for more complex sampling designs where the users' accuracy cannot be estimated directly, users' accuracies can be estimated from the $\widehat{p_{iJ}}$ and $\widehat{p_{i+}}$ estimates). The variances of these estimates are similar in form to those of the users' accuracies, with each term in the numerator multiplied by the constant term (A_i/A):

$$var(\widehat{p_{iJ}}) = \frac{\widehat{p_{iJ}}(\widehat{p_{i+}} - \widehat{p_{iJ}})}{n_{i+}} = \frac{(n_{iJ}/n_{i+})(1 - n_{iJ}/n_{i+})}{n_{i+}}(A_i/A)$$

The column totals $\widehat{p_{+J}}$ are estimates of the proportion of the target area in each field vegetation class *J*, and are equivalent to the stratified estimator across mapped vegetation types:

$$est(Proportion\ of\ target\ area\ in\ vegetation\ class\ J) = \widehat{p_{+J}} = \sum_i \widehat{p_{iJ}}$$

$$= \sum_i (n_{iJ}/n_{i+})(A_i/A)$$

Because this estimate is the sum of estimated terms, the variance of this estimate is simply the sum of the variances of those terms (Thompson 2002):

$$est\left(Var(\widehat{p_{+J}})\right) = \hat{V}(\widehat{p_{+J}}) = \sum^i var(\widehat{p_{iJ}})$$

The producers' accuracies can be estimated as the ratio of the estimates $\widehat{p_{iJ}}$ and $\widehat{p_{+J}}$:

$$est(Producer's Accuracy\ of\ vegetation\ class\ J) = \widehat{p_{JJ}}/\widehat{p_{+J}}$$

The variance of the estimate of the producers' accuracy is complicated, as the estimate is the ratio of two estimates. A reasonable approximation of the variance is:

$$est(Var(Producer's\ Accuracy\ Class\ J)) = \left(\frac{\widehat{p_{iJ}}^2}{\widehat{p_{+J}}^4}\right)\hat{V}(\widehat{p_{+J}}) - \left(\frac{2\widehat{p_{iJ}} - \widehat{p_{+J}}}{\widehat{p_{+J}}^3}\right)\hat{V}(\widehat{p_{iJ}})$$

As for the users' accuracies, 90% confidence limits are calculated by subtracting or adding 1.645 times the square root of the variance from/to the point estimate of the producers' accuracies.

An estimate of the true area of each map class (the area of the map class had the map been created using the response design [reference data] methodology), denoted as $\hat{A}_{+J}$, may be calculated by multiplying the producers' accuracy for the class by the total inference area (known from GIS):

$\hat{A}_{+J} = p_{+J} A_{++}$

4.6.3 Computation of Measures of Overall Accuracy

The estimated overall accuracy is similar to the form of the estimate of the proportion of the target area in vegetation class *J*, with the difference that for the proportion in vegetation class *J* the sum is a column of $\widehat{p_{+J}}$, while for overall accuracy the sum is the diagonal of map class *i* = vegetation class *J* $\widehat{p_{i=J}}$

$$est(Overall\ Accuracy) = \sum_{i=J} \widehat{p_{iJ}} = (n_{iJ}/n_{i+})(A_i/A)$$

$$est\big(Var(Overall\ Accuracy)\big) = \sum_{i=J} var\big(\widehat{p_{iJ}}\big)$$

As for the users' accuracies, 90% confidence limits for the overall accuracy is calculated by subtracting or adding 1.645 times the square root of the variance from/to the point estimate of the overall accuracy.

Finally, Kappa ($\hat{K}$) can be estimated from the estimated overall accuracy, the proportions of the target area mapped as each vegetation class ($p_{i+}=A_i/A$), and the estimated proportion of the target area in each field vegetation class:

$$\widehat{kappa} = \frac{\sum_{i=J} \widehat{p_{iJ}} - \sum_{i=J} p_{i+} \widehat{p_{+J}}}{1 - \sum_{i=J} p_{i+} \widehat{p_{+J}}}$$

Both the numerator and the denominator contain terms that are sums of estimates, so the variance of the estimate of Kappa is quite complex. From Congalton and Green (2009), an estimate of the variance of $\hat{K}$ is:

$$\text{variance}(\hat{K}) = \frac{1}{n}\left\{ \frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2^2)}{(1-\theta_2)^4} \right\}$$

The individual variance terms (θ_1, θ_2, θ_3, and θ_4) are:

$$\theta_1 = \frac{1}{n_{tot}} \sum_i n_{ii}$$

$$\theta_2 = \frac{1}{n_{tot}^2} \sum_i n_{i+} n_{+i}$$

$$\theta_3 = \frac{1}{n_{tot}^2} \sum_i n_{ii} (n_{+i} + n_{i+})$$

$$\theta_4 = \frac{1}{n_{tot}^3} \sum_i \sum_J n_{ii} (n_{i+} + n_{+J})^2$$

As for the users' accuracies, 90% confidence limits for the Kappa estimate is calculated by subtracting or adding 1.645 times the square root of the variance from/to the point estimate of Kappa.

4.7 Lumped Classes

One common modification in response to unacceptable measures of individual class accuracy is to combine two or more similar and thus often confused classes into a single map class. Following such lumping, the overall accuracy, kappa, and users' and producers' accuracies for the new combined class must be calculated. Note that the users' and producers' accuracies for the other vegetation classes are unchanged by this lumping. The complication introduced by lumping classes is that after lumping, the sample selection within the new lumped map class is no longer simple random sampling within that class. Rather, it is stratified random sampling within that lumped class, with strata within the lumped class defined by the original map classes.

However, this stratification is accounted for in the estimated population contingency table $\hat{p}$. Because the elements $\widehat{p_{iJ}}$ are each proportions of the total target area, lumping two rows is simply summing the elements in those rows, and lumping two columns is simply summing the elements in those columns. And, because the variance of a sum of estimates is the sum of the variances of the estimates, the variances of the estimates from those lumped rows are simply the sums of the variances of the rows being lumped. Therefore, once the population contingency table with its elements ($\hat{p}_{iJ}$) and the corresponding matrix of variances of those elements is computed for the full set of map classes, recomputing measures of accuracy only requires summing the appropriate rows & columns in the full matrices.

A stratified design is also the case when map classes are lumped or aggregated to report accuracy at higher levels or at coarser levels of an *ad hoc* or of a formal hierarchy. Table 11 portrays an example of individually mapped associations lumped to the higher group level.

Table 11. Calculation of users' accuracy point estimate and 90% confidence intervals for three forest types at Delaware Water Gap National Recreation Area that are lumped to a coarser thematic class (data from Perles et al. 2007).

MAP CLASS NAME		Users' Accuracy, with 90% Confidence Interval	n_{i+}	Map Class Area (A_{i+}) (hectares)	Producers' Accuracy, with 90% Confidence Interval	n_{+J}	Estimated Map Class Area ($\hat{A}_{+J}$) (hectares)
ASSOCIATION	Eastern White Pine Forest	87.5% (73.9%-100.0%)	26	395.1	33.2% (27.3%-39.1%)	30	682.7
ASSOCIATION	Eastern Hemlock Forest	73.1% (58.8%-87.4%)	16	1290.0	65.3% (60.4%-70.2%)	30	1357.5
ASSOCIATION	Eastern Hemlock – Northern Hardwood Forest	68.8% (49.7%-87.8%)	16	1404.1	61.2% (53.5%-68.9%)	31	1239.3
NORTHERN HARDWOOD – HEMLOCK - EASTERN WHITE PINE FOREST GROUP		81.8% (67.6%-95.9%)	58	3089.2	63.8% (59.3%-68.6%)	91	3679.5

The Eastern White Pine Forest, the Eastern Hemlock Forest, and the Eastern Hemlock-Northern Hardwood Forest are members of the Northern Hardwood – Hemlock – Eastern White Pine Forest Group.

Another common situation involving stratified estimation within map classes is spatial stratification in the sampling design. For instance, the Vegetation Inventory project for the NPS National Capital Region mapped 10 contiguous or proximate NPS units together as a single project. Accuracy assessment observations for each map class were stratified across units, with different sampling intensities across units due to minimum and maximum numbers of observations per unit and mapped vegetation class, and different proportions of each mapped class in different units. (This approach also allowed estimates of unit-specific error rates, albeit less precise [with wider confidence intervals] than the estimates of overall project error rates).

Another situation for needing to treat accuracy measures for an individual map class as a stratified random sampling design would result from having stratified the inference area by access cost, and allocating more effort to the more accessible strata within each map class (see section 2.3.3). In these cases, the general approach is to first use all combinations of map class and stratum to compute the estimated population contingency table, then use the simple additive forms for the estimated proportions and variances of the pooled terms, then computing the accuracy measures from the pooled values. Note that multiple stratifications can easily go too far, as dividing a map class into multiple sub-strata reduces the sample size of accuracy assessment observations in each stratum, and thus increases the within-stratum variances that subsequently get summed.

4.8 Design and Formatting for Contingency Tables

It is helpful to balance the rows and columns of the contingency tables to the extent possible, so that that every row is balanced by a corresponding column that defines a vegetation type represented by the map class. This symmetry helps the user interpret to the table patterns.

It is also helpful to order the map classes and vegetation classes so that classes with higher rates of mutual confusion (i.e., higher cell counts in the raw contingency table and higher proportions in the proportional contingency table that lie off the main diagonal of n_{ii} cells) are positioned near one another in the tables. For example, Table 12 shows some confusion between B and A (7 cases in which B [on the ground] was mapped as A, and 7 cases in which A [on the ground] was mapped as B). High incidences of mutual confusion reflect difficulty in distinguishing among the confused classes (this difficulty may be in the mapping process, ecological classification process, ecological interpretation process, other causes, or some combination of these) and tends to occur between thematically similar classes. High rates of confusion, combined with thematic similarity, may be a basis for merging the classes (either permanently in the database or as an *ad hoc* user application that requires more accuracy than thematic resolution).

In some cases, vegetation types that are represented by map classes that are arranged in rows and to which some observations have been assigned may have no reference data (i.e., the class was mapped but was never found in the field); all sample data for such a class are errors of commission. The row total for this class will be zero, and a corresponding column should be added to the table (also totaling zero). This situation is represented by hypothetical Vegetation Type/Map Class F in Table 12.

In other cases, vegetation types that were not considered by the mapper to have occurred in large enough stands to create a map class (and to which no observations based on sample data values were made) may be found, in fact, on the ground in patches larger than the size of the minimum mapping unit. These are recorded as reference data values for observations placed in other map classes, and all reference data for such a class are errors of omission. Although this class would have no observations that were assigned to the map class for evaluation in the sample (row) data, a row should be added (row totals will be zero) to balance the reference data column for this class, which will have some observations. This situation is represented by hypothetical Vegetation Type/Map Class G in Table 12.

Table 12: Numerical example of a sample contingency table for seven hypothetical sample data classes, represented by nine hypothetical reference data classes.

	Reference Data (Observations Classified as Vegetation Type on Ground)									
Sample Data (Observations Selected Within the Map Class)	**A**	**B**	**C**	**D**	**E1**	**E2**	**E3**	**F**	**G**	**Row Total (n_{i+})**
A	20	7	1	0	1	1	0	0	0	30
B	7	15	7	0	0	0	0	0	1	30
C	2	0	27	0	1	0	0	0	0	30
D	1	3	7	8	2	4	2	0	1	28
E	1	0	1	1	3	3	2	0	0	11
F	0	0	0	0	1	1	0	0	0	2
G	0	0	0	0	0	0	0	0	0	0
Column Total (n_{+J})	31	25	43	9	8	9	4	0	2	
Column Total by Map Class	31	25	43	9	21			0	2	**131**

This table largely represents the data of Table 8. Vegetation Type/Map Class F (not mapped as a sample data value, but found in the reference data) and Vegetation Type/Map Class G (mapped as a sample data value, but never found in the reference data) have been added. Additionally, Map Class E is represented in this table as a mapping complex of three Vegetation Types (E1, E2, and E3) that could not be individually distinguished from one another by the mapping methodology, but can be discerned individually on the ground (in the reference data). For brevity, these distinctions are shown here only in the raw counts table.

In some cases, row and column categories will need to be unbalanced as when the field observations can record vegetation classes that are mapped together as a map complex (a many to one relationship of field classes to map classes) or, in the rarer case when field class to map class relationship is one to many (multiple map classes represent a single vegetation type). Such a situation is depicted in Table 12 by the Map Class E, a complex which is comprised of Vegetation Types E1, E2, and E3, so that a reference data value of either E1, E2, or E3 constitutes a match with a sample data value of E. In these cases, the raw counts (in the sample contingency table) and cell and proportions (in the population contingency table) should be

reported at the finest level of thematic resolution (map class or vegetation class). However the accuracy statistics (point estimate and confidence limits) as reported in the proportional contingency table should reflect that the intended vegetation type to map class relationship, with the finer thematic units pooled into their respective coarser thematic units. This may be represented in the contingency tables by an additional marginal row or column (e.g., the bottom row of Table 12) and for these calculations to reflect a match with their coarser thematic units.

Table 13 is an example of a population contingency table, based on the sample data in Table 8. Users', producers', overall, and Kappa accuracies, with associated 90% confidence limits, as calculated using the formulae of Sections 4.4 and 4.6. An estimate of the true map class areas (the estimated areas of each map class had the map been created using the response design (reference data) methodology has been calculated as described in Subsection 4.6.2..

Table 13: Numerical example of a population contingency table for the five hypothetical sample data and reference data vegetation classes of Table 8.

	Reference Data (Observations Classified as Vegetation Type on Ground)									
Sample Data (Observations Selected Within the Map Class)	**A**	**B**	**C**	**D**	**E**	**Row Total (p_{i+})**	**Map Class Area (Hectare) (A_{i+})**	**Point Estimate, Users' Accuracy**	**Lower Limit, 90% Conf. Int.**	**Upper Limit, 90% Conf. Int.**
A	0.353	0.124	0.018	0.000	0.035	0.530	258.6	66.7%	52.3%	81.0%
B	0.026	0.055	0.029	0.000	0.000	0.110	53.8	50.0%	43.1%	56.9%
C	0.015	0.000	0.207	0.000	0.008	0.230	112.2	90.0%	77.8%	100.0%
D	0.003	0.014	0.024	0.028	0.028	0.096	47.0	28.6%	23.5%	33.7%
E	0.003	0.000	0.003	0.003	0.024	0.034	16.4	72.7%	65.1%	80.4%
Column Total (p_{+J})	0.401	0.193	0.281	0.031	0.095	**1.000**	**488.0**	**Point Estimate, Overall Accuracy (with 90% Confidence Interval):** **66.7% (54.5% - 78.9%)** **Point Estimate, Overall Accuracy Kappa Statistic (with 90% Confidence Interval):** **50.0% (41.2% - 58.7%)**		
Estimated True Map Class Area ($\hat{A}_{+J}$) (Hectares)	195.6	94.0	137.2	14.9	46.3					
Point Estimate, Mean Producers' Accuracy	88.1%	28.6%	73.6%	90.0%	25.7%					
Lower Limit, 90% Conf. Int.	82.3%	19.9%	68.9%	89.9%	21.1%					
Upper Limit, 90% Conf Int.	94.0%	37.4%	78.3%	90.4%	30.3%					

5.0 Reporting Requirements

Each NPS Vegetation Inventory project will report the methods and results of the thematic accuracy assessment in a thematic accuracy assessment report. The report may also be presented as one or more sections or chapters in an overall project report.

At a minimum, the thematic accuracy assessment report will describe:

1. Project methodology for the thematic accuracy assessment (both sampling design and field methods), including a clear description of the inference area, the sampling design, the response design, and other project-specific decisions made within the latitude of these guidelines. Methodology should be clear enough to allow the assessment to be repeated in a reasonably comparable manner, so that the results from a second campaign would be reasonably comparable with the original campaign.

2. A sample contingency table (see Tables 8 and 9 of Chapter 4 and Table 15 of Exhibit F).

3. A population contingency table (see Tables 10 and 13 of Chapter 4 and Table 16 of Exhibit F). This table should display:

   a. individual sample data value by reference data value cell proportions.

   b. users' and producers' accuracies for each map class (presented as a point estimate of class accuracy, with 90% confidence interval limits for each class).

   c. map class inference areas in hectares

   d. estimated true area of each map class (from reference data observations) (in hectares).

   e. overall map (inference area) accuracy (presented as a point estimate of overall accuracy, with 90% confidence interval limits).

   f. overall kappa statistic accuracy (presented as a point estimate of overall accuracy, with 90% confidence interval limits).

4. Where National Vegetation Classification (NVC) classes are map classes, versions of sample and population contingency tables (as in 2 and 3 above) that are lumped and/or crosswalked to the NVC Group and Macrogroup levels.

5. Data reporting formats should follow NPS Vegetation Inventory guidelines (http://science.nature.nps.gov/im/inventory/vg/index.cfm) and other guidance from Inventory and Monitoring networks as appropriate.

Literature Cited

Brewer, C., B. Schwind, R. Warbington, W. Clerke, P. Krosse, L. Suring, and M. Schanta. 2005. Section 3: Existing Vegetation Mapping Protocol. Pages 3-1 to 3-60 in Brohman, R. and L. Bryant, editors. Existing Vegetation Classification and Mapping Technical Guide. General Technical Report WO-67. U.S. Department of Agriculture Forest Service, Ecosystem Management Coordination Staff, Washington, DC, USA.

Cochran, W. G. 1977. Sampling Techniques, 3rd ed. John Wiley and Sons, New York, NY, USA.

Cogan, D. 2007. Vegetation Classification and Mapping Project Report, Washita Battlefield National Historic Site, National Park Service, Southern Plains Network. Natural Resource Technical Report NPS/SOPN/NRTR—2007/075. National Park Service, Fort Collins, CO, USA. (http://biology.usgs.gov/npsveg/waba/wabarpt.pdf).

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. 2009. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, 2nd ed. CRC Press, Boca Raton, FL, USA.

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

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 49:1671–1678.

Czaplewski, R. L. 1992. Misclassification bias in areal estimates. Photogrammetric Engineering and Remote Sensing 58:189–192.

Czaplewski, R. L. 2003. Accuracy Assessment of Maps of Forest Condition: Statistical Design and Methodological Considerations. Pages 115 to 140 in Wulder, M. A. and S. E. Franklin, editors. Remote Sensing of Forest Environments: Concepts and Case Studies. Kluwer Academic Publishers, Boston, MA, USA.

Environmental Systems Research Institute, National Center for Geographic Information and Analysis, and The Nature Conservancy. 1994. Accuracy Assessment Procedures: NBS/NPS Vegetation Mapping Program. Report prepared for the National Biological Survey and National Park Service, Redlands, CA Santa Barbara, CA, and Arlington, VA, USA.

Faber-Langendoen, D., N. Aaseng, K. Hop, M. Lew-Smith, and J. Drake. 2007. Vegetation classification, mapping, and monitoring at Voyageurs National Park, Minnesota: An application of the U.S. National Vegetation Classification. Applied Vegetation Science 10:361-374.

Federal Geographic Data Committee. 1997. Vegetation Subcommittee. Vegetation Classification Standard. FGDC-STD-005. Federal Geographic Data Committee, U.S. Geological Survey, Reston, VA, USA. (http://www.fgdc.gov/standards/documents/standards/vegetation/vegclass.pdf).

Federal Geographic Data Committee. 2008. Vegetation Subcommittee. National Vegetation Classification Standard, Version 2. FGDC-STD-005-2008. Federal Geographic Data Committee, U.S. Geological Survey, Reston, VA, USA. (http://www.fgdc.gov/standards/documents/standards/vegetation/vegclass.pdf).

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. University of California, National Center for Geographic Information and Analysis, Santa Barbara, CA, USA.

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

Grossman, D. H., D. Faber-Langendoen, A. S. Weakley, M. Anderson, P. Bourgeron, R. Crawford, K. Goodin, S. Landaal, K. Metzler, K. Patterson, M. Pyne, M. Reid, and L. Sneddon. 1998. International Classification of Ecological Communities: Terrestrial Vegetation of the United States. Volume I: The National Vegetation Classification System: Development, Status, and Applications. The Nature Conservancy, Arlington, VA, USA. (http://www.natureserve.org/library/vol1.pdf).

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

Lea, C. 2008. A report card for the National Vegetation Classification: Is it a standard in practice? Presentation at Organized Oral Session OOS-19. Approaches for linking vegetation classification, vegetation mapping, and remote sensing technologies. Ecological Society of America Annual Meeting. Milwaukee, WI, USA. August 7, 2008.

Lea, C. 2009. New approaches to applying the National Vegetation Classification. Poster presentation at Biennial Conference of the George Wright Society. Portland, OR, USA. March 2-6, 2009.

Milliken, J., D. Beardsley, and S. Gill, S. April 1998. Accuracy Assessment of a Vegetation Map of Northeastern California Using Permanent Plots and Fuzzy Sets. Natural Resources Management Using Remote Sensing. Pages 218-229 *in* Proceedings of the Seventh Forest Service Remote Sensing Applications Conference. American Society for Photogrammetry and Remote Sensing, Bethesda, MD, USA.

Moeller, J., D. Shaw, J. Lund, N. Roper, and T. Loveland. 1998. U.S.G.S. Biological Resources Division Review. Report of the Vegetation Mapping Program review team on the U.S. G. S. Biological Resources Division Vegetation Mapping Program. Agency internal reports. U.S. Geological Survey, Denver, CO, USA. (http://biology.usgs.gov/npsveg/reviews/peer2.html).

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

National Park Service. 1996. Program Review of National Park Service Vegetation Mapping Program. Agency internal report. National Park Service, Denver, CO, USA. (http://biology.usgs.gov/npsveg/reviews/peer96.html).

National Park Service. 1998. National Park Service Review of the Vegetation Mapping Program. Agency internal report. National Park Service, Denver, CO, USA. (http://biology.usgs.gov/npsveg/reviews/peer1.html).

National Park Service. 2008. Alaskapak v2.2 for ArcGIS 9.x. Software application. (http://science.nature.nps.gov/nrgis/applications/gisapps/gis_tools/8x/alaskapak.aspx).

Patterson, K. D. 2008. Vegetation Classification and Mapping at Appomattox Court House National Historical Park, Virginia. Technical Report NPS/NER/NRTR—2008/125. National Park Service, Philadelphia, PA, USA. (http://biology.usgs.gov/npsveg/apco/apcorpt.pdf).

Perles, S. J., G. S. Podniesinski, E. Eastman, L. A. Sneddon, and S. C. Gawler. 2007. Classification and Mapping of Vegetation and Fire Fuel Models at Delaware Water Gap National Recreation Area: Volume 1 of 2. Technical Report NPS/NER/NRTR—2007/076. National Park Service, Philadelphia, PA, USA. (http://biology.usgs.gov/npsveg/dewa/dewarpt.pdf).

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

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

Stehman, S. V. 1995. Estimating standard errors of accuracy assessment statistics under cluster sampling. Remote Sensing of Environment 60:258–269.

Stehman, S. V. 2009. Sampling designs for accuracy assessment of land cover. International Journal of Remote Sensing 30:5243-5272.

Stehman, S. V. and R. L. Czaplewski 1998. Design and analysis for thematic map accuracy assessment: fundamental principles. Remote Sensing of Environment 64:331–344.

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, Santa Barbara, CA, USA..

Story, M., and R. G. Congalton. 1986. Accuracy assessment: a users' perspective. Photogrammetric Engineering and Remote Sensing 52:397–399.

Thomas, K., T. Keeler-Wolf, J. Franklin, and P. Stine. 2004. Mojave Desert Ecosystem Program: Central Mojave Vegetation Database. U. S. Geological Survey, Sacramento, CA, USA.

Thompson, S. K. 2002. Sampling, 2nd ed. John Wiley and Sons, New York, NY, USA.

U.S. Geological Survey. 1999. Responses to recommendations by the review team. USGS Biological Resources Division Vegetation Mapping Program. Agency internal reports. Denver, CO, USA. (http://biology.usgs.gov/npsveg/reviews/recommendations.html).

Young, J., G. Fleming, W. Cass, and C. Lea. 2009. Vegetation of Shenandoah National Park in Relation to Environmental Gradients, Version 2.0. U.S. Geological Survey, Virginia Natural Heritage Program, and National Park Service, Leetown, VA, USA.

Yuan, D. 1997. A simulation comparison of three marginal area estimators for image classification. Photogrammetric Engineering and Remote Sensing 63:385–392.

Zar, J. H. 1996. Biostatistical Analysis, 3rd ed Prentice Hall, Upper Saddle River, NJ, USA.

Appendix A. Minimum Mapping Unit (Point) Versus Polygon Sampling Designs for Thematic Accuracy Assessment

Sampling for the accuracy assessment for thematic data may have a minimum mapping unit-based (standard area based) or polygon-based design. Minimum mapping unit-based sampling often is regarded as synonymous with point-based sampling (Environmental Systems Research Institute et al. 1994). This has caused confusion in interpretation. Both minimum mapping unit and polygon-based sampling evaluate a specific area on the earth's surface. In the former method, this area is a standard-sized area; in the latter, the size of this area is defined by the mapper for each observation. "Point-based" refers to the site selection methods. Determination of the siting of the observation area for a minimum mapping unit-based design can either be based on selection of an infinite number of points that define the center of the observation area or on selection from a finite number of standard sized pixels variation of a point-based design for geographic data that require that a pre-defined area be observed. Thus, a minimum mapping unit-based design may be either point-based or area-based, with the former usually more rigorous in a statistical sense and logistically advantageous. In contrast, observation selection for a polygon-based design can only be area-based.

With a point (or minimum mapping unit)-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 independently from any boundaries that the mapper may have used to delineate vegetation stands in the spatial data. Typically, this type of assessment requires the random selection of a number of coordinates from the map data. Field investigators then visit the site and observe the conditions to determine its reference data class. The area observed is theoretically the site itself (a dimensionless point), and, for some geographic data, the reference data value of a site can be assessed on the ground from an accordingly small area (nearly a point). However, for many types of geographic data (e.g., vegetation),observers must assess an area of a minimally sufficient size to accommodate most or all of the attributes that must be observed and evaluated in order to establish class of the site to for the classification (the minimum thematic area). Observations of these attributes must be sufficiently near to the site (point) itself in order to be representative of the site. In order to be meaningful as an assessment of remotely sensed (mapped) data, the minimum thematic area for field assessment must be no larger than the minimum thematic area for remote sensing classification (the minimum mapping unit), and, ideally, is the same size as the minimum mapping unit.

In contrast, in a polygon-based approach, the sampling design randomly selects a number of sample polygons from an inference population of all possible polygons. Each selected polygon represents an individual observation which will then be visited in order to determine whether the polygon as a whole has been correctly classified. While most minimum mapping unit (point)-based designs permit the observation of a "point area" in its entirety (i.e., without further subsampling to determine its reference data value), individual vegetation polygons usually are too large to be classified from an observation at a single site. Therefore, to field-check the reference data value of an entire polygon usually requires using a subsampling technique, such as plot or transect sampling within each polygon.

For thematic accuracy assessment of vegetation, a minimum mapping unit based approach usually is preferable to a polygon-based approach from both a map users' perspective and from the perspective of the efficiency of data collection.

A minimum mapping unit based approach to thematic accuracy assessment for vegetation is preferable from a user perspective because, although data representation may be in the form of polygons, many data uses involve raster data. By defining a minimum mapping unit size for the polygon data (the minimum required size for a homogeneous vegetation class), the accuracy assessment defines a resolution scale for the vector data and enables the user to assess the data at a resolution similar to a pixel scale as used in raster data. A minimum mapping unit-based accuracy assessment is often more meaningful to the user because users tend to have implicit "point-based" views of the meaning of an accuracy estimate. As such, accuracy estimates tend to be interpreted as the probability of encountering that class when visiting any spot on the Earth's surface, not when visiting a particular area. A minimum-mapping unit approach also allows the user to make inferences about areas of the map class that are less constrained by the mapper's methodology; polygons are of variable sizes. Finally, a minimum mapping unit-based design allows the probability inclusion of any observation to be known on an area scale (rather than the non-specific scale of a "polygon"). This allows the calculation of marginal areal proportions of map classes and allows the accuracy assessment results to be applied to the map as a more accurate estimator of actual map class areas than is the unassessed map alone (Czaplewski 1992, Yuan 1997, Congalton and Green 2009). This function would either be impossible, or would be, at best, computationally daunting using a polygon-based accuracy assessment design.

A minimum mapping unit-based approach is preferable from the perspective of data collection because it is difficult to draw inferences about polygons as a whole, without observing the entire polygon or by subsampling the polygon using plots or transects intensively enough to determine class. Using plots would almost certainly be prohibitively expensive. Although more cost-effective than plots, the use of transects 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, there can be a substantial inherent level of doubt in assigning a class to a polygon on the basis of the results of transect sampling.

The requirement for the NPS Vegetation Inventory for Validation and for Accuracy Assessment is a minimum mapping unit based approach. An area that is pre-defined as the size of the minimum mapping unit for the map class will serve as the observation area to be used for classifying the point.

Appendix B. Sampling Design for Thematic Accuracy Assessment of Shenandoah National Park Vegetation Map (as modified from Young et al. 2009)

Site Considerations:

Shenandoah National Park (SHEN) is a National Park unit that is both large (79,246 hectares) (= 195,821 acres) and fairly complex (35 map classes, mostly equivalent to the National Vegetation Classification association level).

Because of the linear shape of the park, the presence of the Skyline Drive traversing the long axis of SHEN, a number of access roads along the boundary, and a fairly well-developed trail and fire road system, 98% of all sites within SHEN are within 600 meters in horizontal distance of either a road or trail, and the most remote sites are no more than 2 kilometers (1.2 miles) in horizontal distance from a road or trail. While, few, if any, sites are truly inaccessible on foot, off-trail travel can be extremely slow and difficult due to the steep and, often, rocky, slopes and the sometimes dense and "unfriendly" vegetation that may occur at ground level in response to death of forest canopy trees from recent fires, ice storms, and insect (gypsy moth, hemlock woolly adelgid) infestations. Thus, a "cost surface" in the form of a maximum distance from a road or trail was used to modify the accuracy assessment inference area. From previous experience, 250 meters from the nearest road or trail was judged to be a reasonable maximum distance for most map classes that would balance off-trail travel costs with representativeness (at least 50% of the map class would be considered).

Not all map classes at SHEN are equally distant from an access route. Because the Skyline Drive traverses the crest of the Blue Ridge, most of the map classes representing high elevation vegetation types are never far from a road. Because hiking trails usually follow stream valleys or ridges, map classes representing riparian or ridge top vegetation tend to be concentrated near trails, although sometimes a lengthy hike from a road. The most remote map classes tend to represent vegetation types that occur on steep, rocky slopes, middle slopes (e.g., boulder fields) because roads and trails usually avoid such terrain. In other cases, some rarer map classes may have the majority of their area away from roads and trails simply because of distributional chance.

For purposes of balancing map class representation with costs of conducting the accuracy assessment campaign, we deemed it desirable to include a minimal percentage of each map class in the inference area. The 250 meter access buffer would allow over 50% of the map class to be considered for site selection for 29 of the 35 map classes. For the additional 6 classes (only), extending this default buffer distance by 50 meters (to 300 meters) would allow at least 50% of the total area of these classes to also be included for consideration.

Because of the long driving distances (more than 160 kilometers between the ends of the park) and trail hiking distances sometimes involved, we briefly considered also buffering for driving and trail hiking costs. It was apparent that a driving distance buffer would cause some geographic imbalance (the south and east sides of the park are further from SHEN headquarters than the north and west sides), and a trail hiking buffer would further reduce representativeness in many types. Additionally, once the cost of a long driving and/or trail hiking trip is overcome,

many sites near these access routes can be visited on a single trip. Thus, we elected not to consider driving and trail hiking time in our cost buffer. Clearly, off-trail hiking from an access route to individual sites and a return to the access route tended to be the largest access time sink (highest cost per observation).

Determination of Map Class Sample Size Allocation:

Using ArcView ® 3.2, we calculated areas (in hectares) for all map classes prior to the classes being reduced in size through buffering for site allocation requirements of (1) field navigation error and positioning of entire plot within the map class (2) access convenience (see below). We allocated 30 accuracy assessment (AA) sites to each map class that occupied more than 50 hectares in total area. A site is a specific location (point) defined by a set of x (easting) and y (northing) coordinates and represents the center point coordinates of individual AA sample units or observations. For map classes with 50 hectares or less of total area, we allocated AA sites equal to the total map class area (in hectares) divided by 1.67. We allocated at least 5 sites to every map class, except in cases in which the map class area (as reduced by buffering) was too small to accommodate at least 5 spatially separate (non-overlapping) observation areas of the minimum mapping unit (MMU) size prescribed for the class, in which case the map class received the maximum possible number of non-overlapping sites.

Determination of Map Class Site Observation Size:

We used a 0.5 hectare observation area (usually a circular plot, 40 meters in radius), centered around each AA site (point) for large patch or matrix-forming map classes (Table 14). For small patch and linear map classes, a 0.25 hectare observation area (usually a circular plot, 30 meters in radius) was used (Table 14). The area reduction from the normative 0.5 hectare observation area (Environmental Systems Research Institute et al. 1994) was necessary in order to accommodate multiple AA observation plots within stands mapped as small or narrow polygons. The vegetation representing most of these small patch or linear map classes is comprised primarily nonvascular and herbaceous vegetation types, in which smaller observation areas can capture an adequate number of attributes to establish class (vegetation type).

Preparation of Map Classes for Site Selection:

First, the raster file representing the vegetation map was converted to a polygon (shape) file, suitable for geoprocessing and buffering.

We used ArcView ® 3.2 geoprocessing wizard to prepare the sampling population for each map class, as follows:

To prepare the cost surface (access buffer) for each map class:

(1) Using the SHEN roads (http://nrdata.nps.gov/shen/shendata/roads.e00; as accessed November 8, 2009) and trails (http://nrdata.nps.gov/shen/shendata/trails.e00; as accessed November 8, 2009) themes, we created a buffer theme consisting of 12 progressively more remote areas buffered from any road or trail as represented by these line themes, with buffers in increments of 50 meter distances. This produced a polygon theme with map classes consisting of areas within 50 meters of a road or trail, from 50 to 100 meters from a road or trail, from 100 to 150 meters from a road or trail, etc., up to from 550 to 600 meters from a road or trail.

(2) We created a union of this theme and the vegetation theme, so that each vegetation map class was divided into 12 distance classes, each represented that portion that occurred within each 50 meter buffer increments, and a 13th class that represented that portion that was more remote than 600 meters (600 to 2000 meters) from a road or trail. This theme had fields of buffer distance class and vegetation type (map class).

(3) We imported the attribute (.dbf) table from this union theme into Microsoft Excel ®. We created a pivot table of the buffer distance class membership (in columns) against the vegetation class membership (in rows), and the cell values the sum of the total map areas for each combination of these vegetation classes and buffer distance classes. The cell values were first converted to a percentage of the [row] total for each vegetation class. These percentages were converted to cumulative percentiles (beginning with buffer distance class of 0 to 50 meters) of the vegetation class that occurred within each buffer distance class or all less remote buffer distance classes.

(4) In ensuring that, minimally, the most accessible 50th percentile of each vegetation map class would be included in the sampling population, we found that a 250 meter buffer would satisfy this condition for 29 of the 35 vegetation map classes and that a 300 meter buffer would satisfy this condition for the other six vegetation map classes. We designated a 300 meter buffer to be used to constrain sampling for these six classes plus four additional classes that were limited in overall area and would be further reduced in area by edge buffering. Most of these classes represented small patch types that, in total, would receive a very small proportion of all AA sites. A 250 meter buffer was designated for all other (25) vegetation map classes. The percentiles considered for each vegetation map class ranged from 51% to 100% (Table 14). Approximately 56% of the area of SHEN occurs within 250 meters of a road or a trail, while about 65% occurs within 300 meters of at least one of these features.

A minimum buffer distance from the edge of all polygons is needed in order to ensure that global positioning system (GPS) navigation and positional recording area does not create uncertainty about the map class membership of individual AA sites. To establish this buffer:

(1) We created a 40 meter buffer inside the boundary of all polygons.

(2) We created a union theme of the original map classes and this buffer theme.

(3) We selected the portions of all polygons that were within the 40 meter buffer and deleted them from this union theme. The result of this geoprocessing step was a polygon theme (sampling population) that was comprised of the interiors of all polygons (all areas more than 40 meters from a boundary with a different map class). This buffer was to be applied to all map classes to be observed in the field at the scale of 0.5 hectare.

(4) We repeated the above process using a 30 meter buffer. The result of this geoprocessing step was a polygon theme that was comprised of the interiors of all polygons (all areas more than 30 meters from a boundary with a different map class). This buffer was to be applied to all map classes to be observed in the field at the scale of 0.25 hectare.

To apply the appropriate access buffer to these edge buffer themes:

(1) We created a buffer theme of all SHEN areas within 300 meters of any road or trail.

(2) We created a buffer theme of all SHEN areas within 250 meters of any road or trail.

To create the final inference area (sampling population) for each individual vegetation map class:

(1) We used the intersection (geoprocessing) command with each combination of the two road/trail access buffer themes (250 meters and 300 meters) with each vegetation polygon edge buffer theme (30 meters and 40 meters) to create four themes representing the sampling population to be applied to the vegetation map classes. The third and fourth columns of Table 14 list which combination of access and polygon edge buffer distance was used for each map class.

Selection of Individual AA sites from the Prepared Map Classes:

(1) Using the appropriate inference area for each map class (Table 14), we located individual AA sites (plot centers) by allocating the specified number of points for each map class to the modified union theme derived from the operations above, using the "Select Random Features" function in the National Park Service Alaskapak tools package for ArcView® 3.2 (National Park Service 2002). This achieves the requirement of simple random sampling within each individual inference area (map class, as modified above).

(2) When two or more AA sites were near enough to one another to produce overlapping observation areas (i.e., within 80 meters of one another for classes to be observed at the 0.5 hectare scale or within 60 meters of one another for classes to be observed at the 0.25 hectare scale), one site at a time was selected randomly (using the random numbers function in Microsoft® Excel) and deleted. A replacement site was generated for each site so deleted, using the "Select Random Features" function (as above). If the replacement site were near enough to a previously located site, so as to produce overlapping observation areas (as above), it was rejected, and the process repeated, until either (1) the full complement of sites for the map class was assigned or (2) it was determined that the map class was saturated (could accommodate no more sites without observation area overlap between one or more AA plots).

Table 14. Accuracy assessment observation area size and results of access buffer application for individual Shenandoah National Park vegetation map classes (sample data classes).

Map Class (Vegetation Type)	Observation Area Size (hectares)	Access Buffer in meters[1]	Percentile of Class Considered[2]
Virginia Pine Successional Forest	0.25	300	51.1%
High-Elevation Greenstone Barren	0.25	250	88.1%
Central Appalachian Basic Woodland	0.25	250	92.4%
High-Elevation Outcrop Barren	0.25	250	100.0%
Central Appalachian High-Elevation Boulderfield Forest	0.25	250	100.0%
Central Appalachian Basic Woodland	0.25	250	100.0%
Central Appalachian Mafic Barren	0.25	300	68.1%
Central Appalachian Circumneutral Barren	0.25	250	72.0%
Cent. App. Xeric Chestnut Oak - Virginia Pine Woodland	0.25	300	77.7%
Central Appalachian Acidic Boulderfield	0.25	250	65.8%
Central Appalachian Mafic Boulderfield	0.25	300	62.5%
Northern Blue Ridge Mafic Fen	0.25	250	100.0%
Cultural Meadow	0.25	250	81.2%
Central Appalachian Northern Hardwood Forest	0.50	250	71.6%
Northern Blue Ridge Montane Alluvial Forest	0.50	250	87.9%
Central Appalachian Pine - Oak / Heath Woodland	0.50	250	68.4%
Sweet Birch - Chestnut Oak Talus Woodland	0.50	300	52.6%
Central Appal. / Northern Piedmont Chestnut Oak Forest	0.50	250	51.5%
Low-Elevation Mixed Oak / Heath Forest	0.50	300	53.7%
Cent. App. Dry-Mesic Chest. Oak – North. Red Oak For.	0.50	300	56.4%
Hemlock - Northern Hardwood Forest	0.50	300	55.5%
Northern Red Oak Forest	0.50	250	76.7%
Southern Appalachian Cove Forest	0.50	250	76.8%
Acidic Cove For. (White Pine - Hemlock - Mixed Hardwds)	0.50	300	58.5%
Successional Tuliptree Forest	0.50	250	65.5%
Central Appalachian Basic Boulderfield Forest	0.50	250	53.5%
Central Appalachian Rich Cove Forest	0.50	250	51.0%
Central Appalachian Basic Montane Oak - Hickory Forest	0.50	250	54.4%
Central Appalachian Acidic Montane Oak - Hickory Forest	0.50	250	63.1%
Central Appalachian Acidic Oak - Hickory Forest	0.50	250	75.0%
Central Appalachian Basic Oak - Hickory Forest	0.50	250	53.2%
Northeastern Modified Successional Forest	0.50	250	56.0%
Cent. App. Dry Chest. Oak – North. Red Oak / Heath For.	0.50	250	63.3%
Acidic Cove For. (Hemlock - Hardwood / Mountain Laurel)	0.50	300	57.9%
Catastrophically Disturbed Forest	0.50	250	63.1%

1 - Buffer distance from road or trail that was applied to the map class to constrain sampling to more accessible areas.

2 - Resulting percentile of entire map class area at SHEN that was considered for sampling, using the buffer distance applied (250 or 300 meters).

Appendix C. Examples of Providing Field Maps Suitable as Navigation Aids for Thematic Accuracy Assessment

These figures are examples of both acceptable and incorrect ways to display map and image data to field teams conducting thematic accuracy assessment field observations. These visual aids are intended to be a supplement to sufficiently accurate Global Positioning System navigation to the sites and site selection methods adequate to place most sites accurately within the target map classes.

Figures 3A through 3D are examples from Thomas Stone National Historical Site (Maryland) (digital orthophoto quarter quadrangle image from USGS Port Tobacco 7.5 quadrangle Southeast quarter quadrangle and hypsography from U.S. Geological Survey Port Tobacco 7.5 minute quadrangle).

Figures 4A though 4D (digital raster graphic from U.S. Geological Survey Conejo Wells (California) 7.5 minute quadrangle), 5A, and 5B (digital raster graphic from U.S. Geological Survey Yucca Valley (California) 7.5 minute quadrangle) are examples from Joshua Tree National Park.

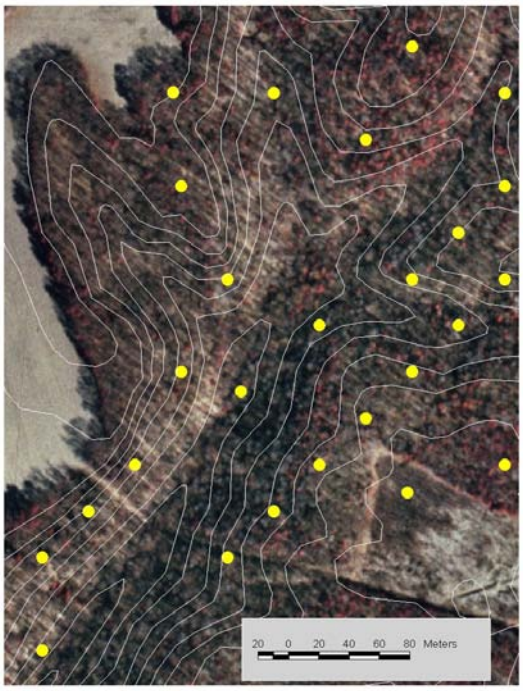

Figure 3A. Acceptable. Observation sites (plot centers) are shown as yellow points on an image of the study area, with elevation contour lines (white).

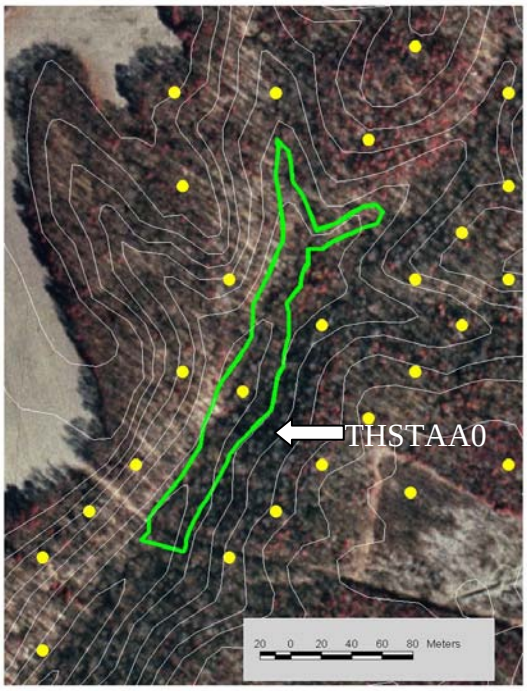

Figure 3B. Acceptable. A single polygon (only) enclosing a single observation site is depicted (in green) to aid navigation to THSTAA015.

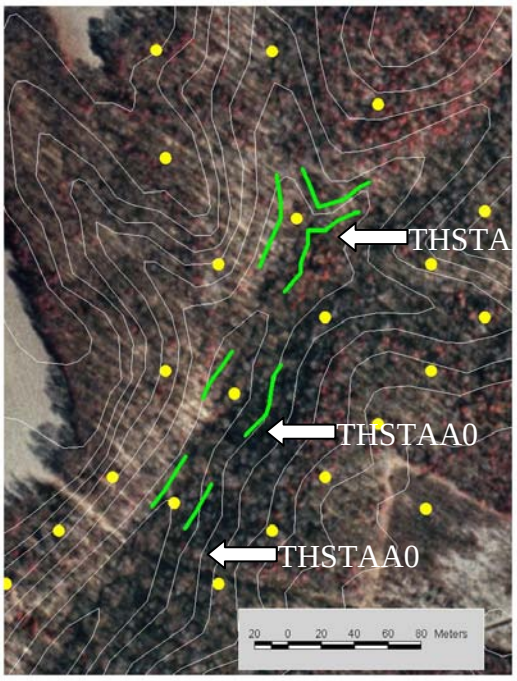

Figure 3C. Acceptable. Map class boundaries are revealed only in vicinity of individual sites in small or linear polygon(s). For this example, sites THSTAA014 and THSTAA0016 were added to the array.

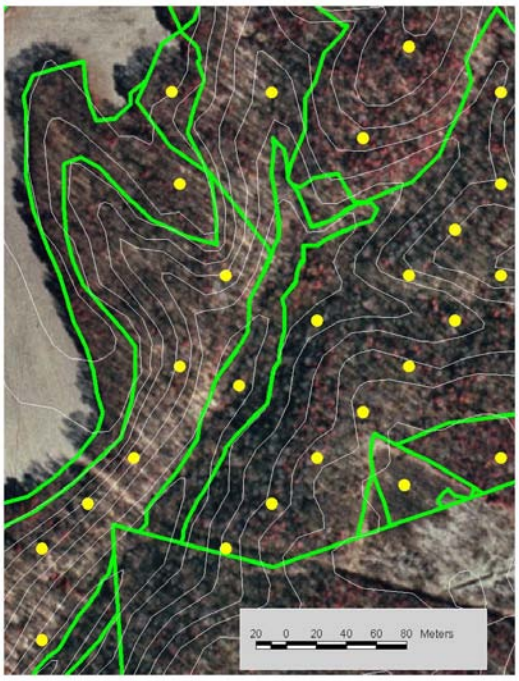

Figure 3D. Incorrect. Showing all map class boundaries (green lines) for multiple sites within same and/or adjoining polygons reveals map class membership relationships.

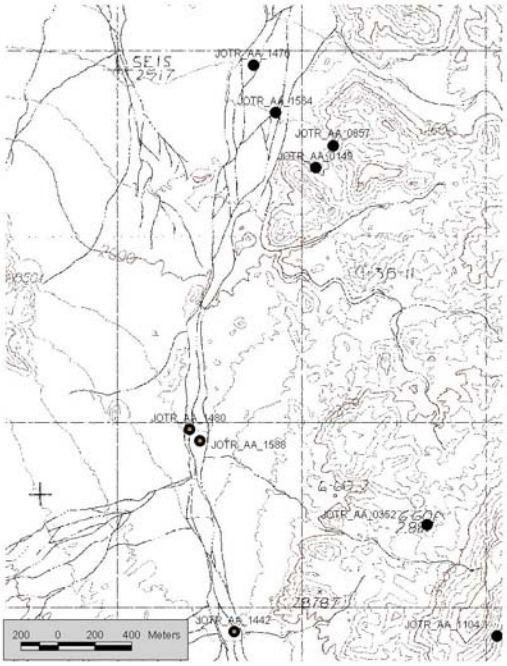

Figure 4A. Acceptable. Observation sites (plot centers) are depicted on digital raster graphic (7.5' USGS) image of the study area.

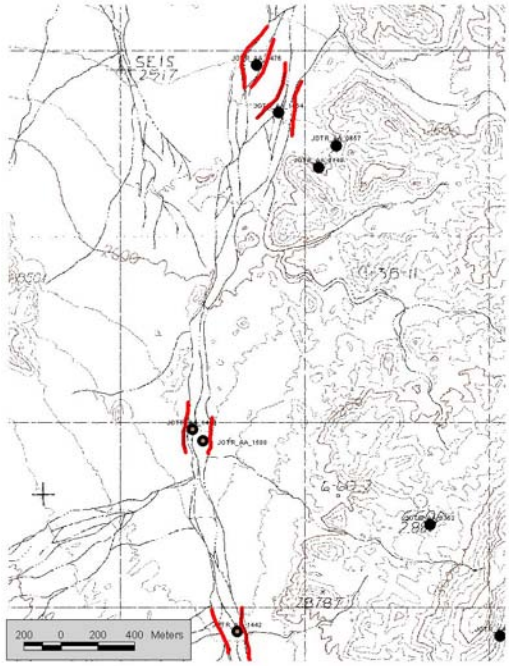

Figure 4B. Acceptable. Map class boundaries (red lines) are revealed only in vicinity of individual sites to aid field navigation.

Figure 4C. Incorrect. Showing all map class boundaries (dark black lines) for multiple sites within same and/or adjoining polygons reveals map class relationships.

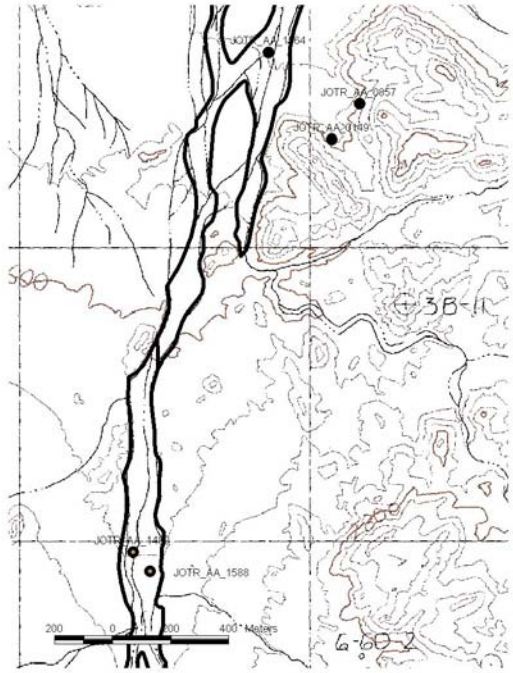

Figure 4D. Incorrect. Showing all map class boundaries (dark black lines) for multiple sites within same and/or adjoining small/linear polygons (only) still reveals map class relationships.

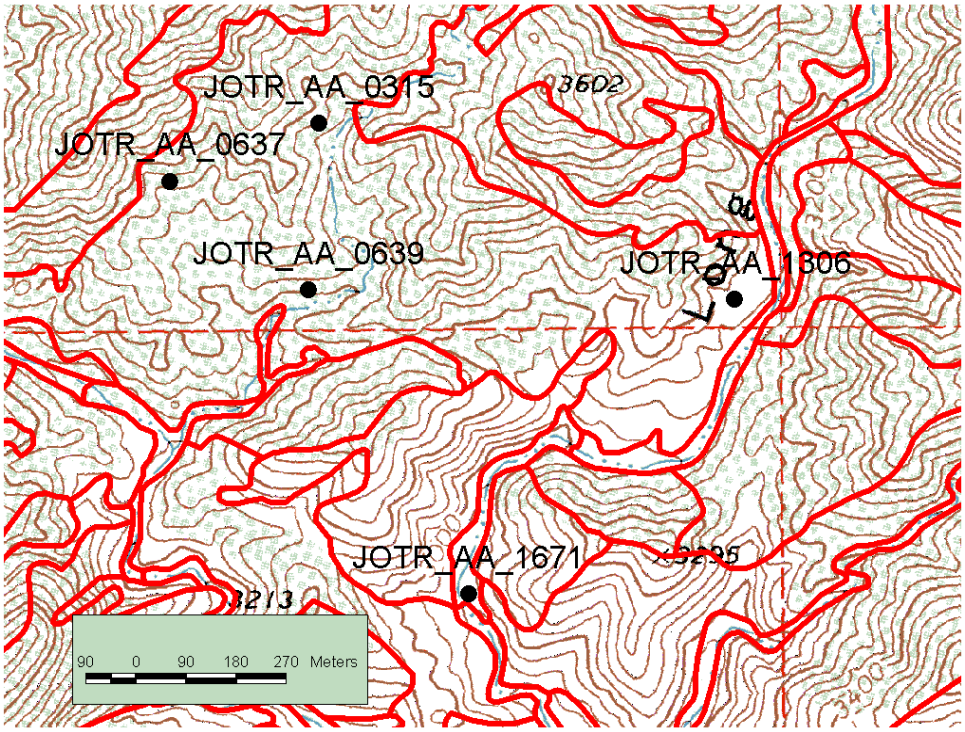

Figure 5A. Incorrect. Showing all map class boundaries (red lines) for multiple sites within same and/or adjoining small/linear polygons reveals map class relationships.

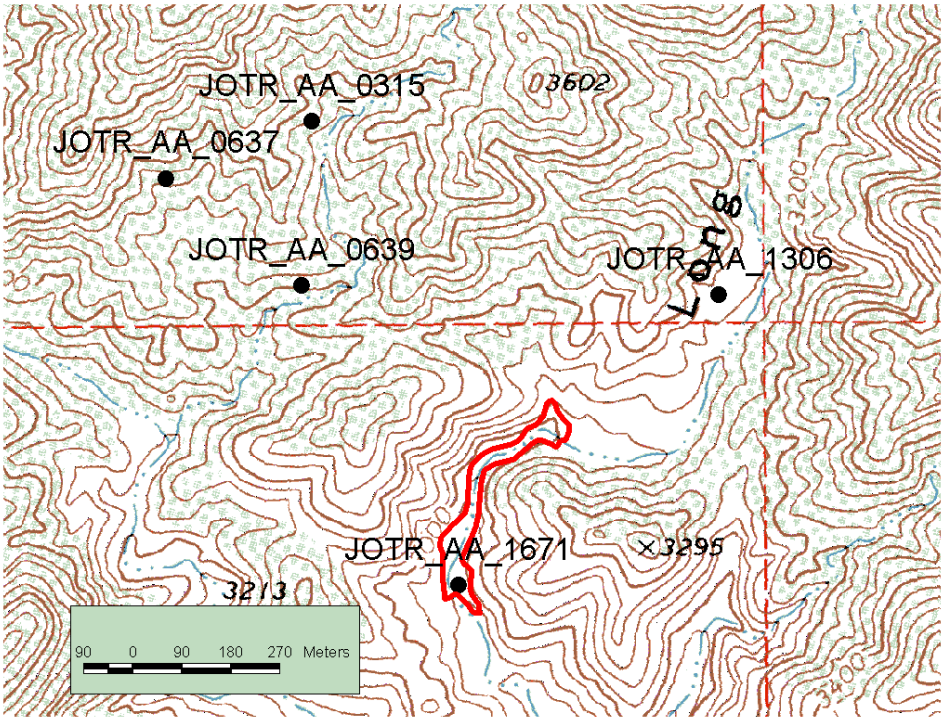

Figure 5B. Acceptable. Revealing small or linear polygon boundary (red lines) for a polygon that contains single observation site reveals little about map class relationships.

Appendix D. How to Represent Accuracy Assessment Data when Map Corrections are Made

Generally, the validation process, which is conducted by the project oversight team and prior to accuracy assessment, will address problems of low thematic accuracy as evaluated across the entire map. By the time the accuracy assessment is conducted, it is assumed that the map, as a whole, will have been assessed to be of acceptable accuracy. Additionally, verification, a process conducted by the map production team that seeks to identify and correct class-specific causes of thematic error, will have corrected a number of individual map class deficiencies. Nevertheless, thematic accuracy assessments will almost invariably reveal that some individual map classes have considerably lower accuracy than the overall map accuracy process.

Low accuracy for individual map classes may result from inaccurate or improper procedures during the compilation of the map but may also be due to limitations of time and/or other resources that are inherent in the vegetation classification and mapping process itself. If excessive error rates result from procedural issues, the classification and/or mapping process may have to be revised or repeated, starting at the point in the process where the procedural error occurred. This is possible only if enough data are collected to isolate the error source. While NPS Vegetation Inventory procedures allow for such data to be collected during map production (i.e., as an internal process of verification), the data collection procedures for accuracy assessment often will not allow these problems to be identified. This reality should emphasize the need for producers to conduct adequate verification for individual classes and for project oversight team to conduct an adequate validation for the entire map. As a general rule, the NPS Vegetation Inventory accuracy assessment process is designed to be primarily a means for the user to understand the properties and limitations of individual map classes; it is not an appropriate process for correcting major deficiencies in the map (i.e., it is not meant to serve as an internal quality assurance function for map production).

In most cases of low accuracy for individual map classes, it is probably best to simply report thematic accuracy of all map classes in a contingency table and to leave it to the user to decide how best to use the map. Thematic accuracy and thematic resolution are generally inversely related, and both qualities may be of value to users, with some sets of user applications favoring the former and others the latter. The reporting of accuracies at the finest thematic level attempted allows for more diverse map applications. Consider the case in which an individual map class A representing vegetation type A has 33% users' accuracy in a map in which the overall accuracy at the finest thematic level attempted is 65% and which has 50 classes. In the context of the entire map, this class may seem to have low accuracy, and consideration might be given to permanently merging it with map class B with which it has been frequently confused. However, an accuracy rate of 33% for a single class out of 50 would be considerably higher than could be achieved by chance. If it were important to find this vegetation type at the finest thematic level in the field, the map user would understand from the contingency table that there is a one in three chance of finding it when using the map (and that the vegetation type occupies only about a third of the map class that represents it). Conversely, if map class A were permanently merged with map class B to create a thematically coarser map class with higher users' accuracy than that of either individual class individually, it would be considerably more difficult for a user to locate individual stands of vegetation type A. Moreover, if map class B were considerably more

abundant than map class A, a permanent merge of the classes might make it impossible, for practical purposes, for the user to locate vegetation type A by using the map. Using the information in the contingency table, the user may employ either the properties of individual map classes or those of *ad hoc* (application specific) classes formed by the merging of user-selected map classes.

In some cases, two or more map classes may show high levels of internal confusion (i.e., errors of commission and/or omission among one another), but have relatively high accuracy (little confusion with other classes) when considered as a map complex comprised of the finer thematic classes. If the internal error rates are more than to little less than would be expected to occur by chance (i.e., the mapper were to "guess" at the identities of individual types within the difficult complex), it may be desirable to represent them always as a map complex. The results in the contingency table should be modified to reflect this permanent merging, taking into account the effect of a stratified random sampling design with unequal densities of observations within individual strata on the reporting point estimates and confidence intervals for thematic accuracy.

Where field reference sites document a different vegetation type from that represented by the map class, the question may arise as to whether to incorporate this information into the map in order to increase accuracy, especially if the field observations are deemed unquestionably to be a more accurate representation of the vegetation, rather than a measure of correspondence between the assertion by the map and that made by a typical field user (where either assertion might be incorrect). Except for projects at the smallest parks, the area occupied by the mapped class will be much more abundant than the area of the individual [sample unit] observations within it, for most map classes. Strictly speaking, only the area observed at each reference site (usually 0.5 hectare) can be assumed to correspond to the reference data value (an assumption of the minimum mapping unit based assessment used by the NPS Vegetation Inventory), rather than the entire area represented by the mapper's polygon within which the site occurs (which would be the assumption if a polygon-based design were used – see Appendix B). In cases (usually in rarer map classes) in which the area represented by the mapper's polygon around the site is not much bigger than that of the observed area and in which the vegetation type represented by both the map class and by the field observation is known to be both distinct from other types and internally homogeneous, an assumption of homogeneity might be reasonably extended to the entire polygon. However, this assumption would less likely be valid for vegetation types that are mapped more abundantly and as larger polygons, for types that are represented by individual polygons that may be internally heterogeneous, and/or for types for which the locations of stand boundaries are mapped with less certainty. The importance of the individual accuracy assessment observations is their contribution as a part of a larger sample to an understanding about the map classes in which they occur and about the vegetation classes to which they belong, and not so much in what they reveal about the particular location to which they happened to be assigned and where they were observed. In most cases, the use of accuracy assessment observations to improve the map classes is an inefficient process, at best.

Nevertheless, there may be cases during the accuracy assessment in which the number of stands of a rare or undiscovered vegetation type significantly increases or newly documents the known amount of the type. It is often of management interest to document these few locations in which these types occur. In these cases, correcting the map to the vegetation type identified in the individual observed area (or extending that correction to a part or all of a larger polygon, taking

into account the considerations above) may be warranted. However, these types of corrections essentially create a different “edition” of the map, and other tools (field keys) used in the assessment may change, as well. These changes would reflect a different set of assumptions than those under which the accuracy assessment campaign and results for individual map classes were derived. It is suggested that the accuracy assessment (contingency table) results be reported as though the map had not been corrected (i.e., these corrections still be counted as errors in the table), since this error rate more accurately represents the properties of the map classes and vegetation types involved and is a more conservatively appropriate representation of the limitations of the map classes.

Example: At Grand Portage National Monument, a vegetation type (the Jack Pine – Aspen / Bush Honeysuckle Forest) that was novel to the project was identified at a site within a map polygon not much larger than the minimum mapping unit. The original sample data value for the polygon was White Pine / Mountain Maple Mesic Forest, a relatively abundant class. The map class value of this polygon was edited from White Pine / Mountain Maple Mesic Forest to Jack Pine – Aspen / Bush Honeysuckle Forest. In the contingency table, the row (sample data) value of the observation was retained as a White Pine / Mountain Maple Mesic Forest and the column (reference data) value was recorded as Jack Pine – Aspen / Bush Honeysuckle Forest (added to the table.

Appendix E. Glossary of Terms, Definitions, and Acronyms

Accuracy	In the context of "thematic accuracy assessment," the measure of the absence of error. In a statistical sense, it is also defined as the measure of how close an estimate is to its true value.
Binomial Distribution	Probability distribution used to describe discrete random variables.
Calibration	As used by the NPS Vegetation Inventory, a process used by production team (internal to the mapping process) to fit the ecological classification model to an initial mapping model. Calibration is map development, rather than map checking and adjusting (verification).
Cluster Sampling	A sampling design in which the inference population is divided into primary sampling units, and the primary sampling units are, in turn divided into secondary sampling units. A subset of the primary units are selected (e.g., by simple random sampling), and all secondary units within the selected primary units are selected.
Community	In the context of vegetation science, vegetation occupying a specific area on the Earth's surface that is internally homogeneous (the entire stand is classified as a single vegetation type). A community is a real, observable entity, as opposed to a community type, which is an abstraction used to describe communities.
Community Type (Vegetation Type)	In the context of vegetation science, an abstract class unit applied to a set of [real] plant communities that are related in a formal taxonomic scheme by some common and, often, defining attributes. Community type or vegetation type often are applied as to be synonymous with the values of the finest level of a classification scheme (e.g., the association of the National Vegetation Classification). However these terms may be and are applied more broadly to values representing any recognized level of a taxonomic hierarchy.
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%).
Contingency Table	Table constructed for classifying count data. The entries in the cells show the number of observations falling into a particular intersection of values for two categories. 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.

Continuous Cover Measurement of vegetation cover as a continuous variable (e.g., 0-100%), rather than as a discrete variable (e.g., in cover classes). Note that, as a practical matter, the measurement of cover as a continuous variable is limited to the precision of the measurement method (e.g., ocular estimates are often recorded in 1% increments for small values and 5% increments for larger values), so that, in a strict sense, continuous cover is usually recorded as in discrete classes with very small class increments.

Continuous Variable A random variable with an infinite set of outcomes.

Correspondence The degree of agreement between two variables. If one variable is regarded as being a higher source of accuracy than the other, then correspondence is defined as absence of error (correctness).

Cost Surface (Access Buffer) A Geographic Information Systems (GIS) method of restricting study sites to the most accessible (least costly in access) subset of the area. In a vector setting, it is typically accomplished by establishing a maximum distance threshold (this distance can be established and computed in units of distance, access time, or access effort), placing a buffer around access sites or corridors and restricting study sites to areas within the buffer. In a raster setting, the resistance to travel is computed for individual raster cells and the limits of the study area are defined by which cells fall below a cumulative maximum cost threshold.

Cover (Vegetation Cover) The percentage of total area within a predefined observation area (e.g., a plot) that is occupied by the two-dimensional above ground area of all vegetation or some subset of the vegetation (e.g., of a vegetation life form or of all members of an individual plant species) within that observation area. Cover is used as a measure of abundance of vegetation or of a species within a prescribed area.

Differential GPS Global Positioning System (GPS) methodology that improves the accuracy of raw (field) GPS positions by comparing them to reference GPS data measured simultaneously at a known (well-surveyed) fixed ground position. The comparison may be made in the field (real-time), if the GPS receiver is so enabled, but is usually done after the fact (in which case the more accurate position can be applied to a data record after the fact, but cannot assist in navigation). The need for differential GPS for vegetation mapping largely has become much less due to discontinued use of Selective

	Availability by the NAVSTAR system, by more accurate commercial receivers, and by the use of WAAS (Wide Area Augmentation System) – enabled GPS.
Discrete 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 site in the database and the class that would be derived from reference data of a higher source of accuracy (e.g., as might be observed in the field). If the reference data are considered not [necessarily] to be a higher accuracy source, then error can be viewed as lack of correspondence. Errors may be thematic or positional in nature.
Error of Commission	An error for an individual observation in which the attribute category applied by the mapper in the sample (test) data fails to match the attribute category that is applied as reference data by the observer employing the higher source of accuracy (the response design), usually by ground observation. The error of commission is applied to the sample data category.
Error of Omission	An error for an individual observation in which the attribute category that is applied by the observer employing the higher source of accuracy, the reference data category, fails to match the attribute category applied by the mapper in the sample (test) data. The error of omission is applied to the reference data category.
Estimate	See point estimate
Global Positioning System (GPS)	A geographic positioning system employing radio signals emitted by a constellation of navigation satellites in precise geo-synchronous orbit that are received by a user with an autonomous receiver on the earth's surface. Calculation of the signal timing accurately places the location of the receiver on the earth's surface. The GPS used in North America employs the United States Navigation Satellite Timing and Ranging (NAVSTAR) 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 table.

Inclusion	A distinct vegetation community whose extent is less than the minimum mapping unit and is heterogeneous with the surrounding vegetation. In a minimum mapping unit based design for thematic accuracy assessment, inclusions are either avoided in the siting of an observation or they are absorbed into the larger observation area, depending on the amount of heterogeneity with the surrounding vegetation and the potential effect of this heterogeneity on the response design objectives.
Inference Population (Inference Area)	The elements to be studied or described in a given experiment. In the case of accuracy assessment, the inference population is often referred to as an inference area and is that area to which at least part of the sample was assigned (i.e., the probability of inclusion for an observation is greater than zero). Ideally, the inference population is all vegetation that has been mapped (see the management population), but cost limitations may require a reduction in the inference area to a more accessible subset of the management population. In such a case, the results of the inference population sampling may be applied to the management population by assumption, but not by statistical inference. Inference population is equivalent to sampling population or population (in a statistical sense).
Management Population	The population of management interest. In a vegetation mapping project, it is typically equivalent to all the vegetation mapped by the project.
Map Class	In a spatial database, a value in a data field for a descriptive categorical attribute that is potentially assigned to a data field for a subset of records in the database. In a vegetation map, a map class might be "ponderosa pine forest." Equivalent to, but a preferable term for, map unit.
Map (Vegetation) Thematic Complex	A map class that is intended to represent more than one distinct vegetation type because the individual vegetation types are thematically too finely resolved to map.
Map (Vegetation) Mosaic Complex	A map class that is intended to represent more than one distinct vegetation type because the individual vegetation types are spatially too finely resolved to map. Equivalent to a mosaic-complex in the sense of Mueller-Dombois and Ellenberg (1974).
Map Unit	Equivalent to map class. Although used as a convention by many in this context, the guidance of the NPS Vegetation Inventory uses the term "map class" rather than "map unit," to refer to a database field value. This avoids linguistic confusion with the term "mapping unit,"

which refers to a database record. See Map Class.

Mapping Unit — In a spatial database, an individual record (e.g., a polygon, line, or point). This is contrasted with "map class" or "mapping unit" above. In a vegetation map, a mapping unit might be an individual polygon representing a particular stand of a "ponderosa pine forest."

Minimum Mapping Unit (MMU) — For a map class vegetation type, a description of the smallest size of a stand of the vegetation type that is represented on a map. Equivalent to minimum map feature of Brewer et al. (2005). For NPS Vegetation Inventory purposes, the minimum mapping unit size is equivalent to the size of the observation area to be used in thematic accuracy assessment. For a minimum mapping unit size of 0.5 hectare for a type, it is expected that all vegetation stands that are 0.5 hectare or more in size will be mapped as a distinct and homogeneous vegetation type (exceptions are considered to be mapping errors) and that stands that are less than 0.5 hectare in size may or may not be mapped (if not mapped, they will be treated as inclusions in other map classes and are not considered to be mapping errors).

Minimum Thematic Unit — For a vegetation type, the minimum area size required to contain an adequate representation of the individual species elements of the type for field evaluation. Equivalent to the "minimum area" of Mueller-Dombois and Ellenberg (1974).

Multistage Sampling — A sampling design in which the inference population is divided into primary sampling units, and the primary sampling units are, in turn divided into secondary sampling units (stages). These, secondary units may be subdivided further (multiple stages). A selection of a subset of each individual sampling unit is made at each stage. This differs from cluster sampling, in which only the primary units are subjected to the selection process, and all secondary units included in the primary units that are selected are incorporated into the sample.

National Map Accuracy Standards (NMAS) — Standards maintained by the U.S. Geological Survey that define accuracy standards for published maps, including horizontal and vertical accuracy, accuracy testing method, accuracy labeling on published maps, labeling when a map is an enlargement of another map, and basic information for map construction as to latitude and longitude boundaries.

National Vegetation Classification (NVC) — A relatively recently developed standardized scheme for classifying vegetation within the United States maintained by the Federal Geographic Data Committee (2008) and contributed by a wide variety of vegetation ecologists.

Normal Distribution A bell-shaped probability distribution, the height of the curve at any single point of which is defined by a specific relationship with the standard deviation and the mean of the distribution (see Zar 1996). Observations of interval or ratio scale data are often approximated by this distribution. Also known as the Gaussian distribution.

Observation Individual members of an inference population that have been selected in a sampling exercise, that are intended to represent that population, and from which variable values are derived to estimate a value for a population parameter. Equivalent to sample unit. Individual observations may have multiple variables assigned to them (e.g., for thematic accuracy assessment, each observation, has, minimally, the variables of sample data value, vegetation type observed in the field (reference data value), and a geographic position. Equivalent to Sample Unit.

Observation Area (Observation Plot) In thematic accuracy assessment, an area associated with an individual sample unit (observation) and its site and over which field (reference) data are collected. The observation area is normally centered on the site. See Site.

Offset In Global Positioning System (GPS) or survey operation, a method used to determine or estimate the position of a site that cannot be physically reached (usually accomplished by estimating a bearing and a distance from a position that can be recorded by the GPS or by recording the site position, using a second surveying tool such as a laser rangefinder, to calculate the position from that recorded by the GPS). Also, the values for the bearing and distance of this methodology.

Parameter A numerical population descriptor.

Point Estimate Value computed from sample data that approximates a population parameter (also called estimate).

Point-based Sampling A sampling scheme for thematic accuracy assessment whereby inference population for a map class is a infinite set of dimensionless points that may be located within the a map class. For practical purposes, the population is more often treated as a finite set of all possible non-overlapping areas of a fixed observation area (the minimum mapping unit). Thus, it is often referred to as Minimum Mapping Unit based sampling. It differs from a polygon (essentially a vector –based approach) based scheme in that it employs a fixed observation area size (essentially a raster-based approach).

Polygon	In a spatial database, a record representing a [two-dimensional] area on the Earth's surface that is homogeneous in at least one attribute. In vegetation mapping, a polygon is an abstract representation of a real vegetation stand on the Earth's surface.
Polygon-based Sampling	A sampling scheme for thematic accuracy assessment whereby the inference population for a map class is a finite set of individual polygons (records) of varying area within the map class.
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. A measure of dispersion of the probability distribution associated with a measurement and expresses the degree of repeatability of a measurement.
Producers' Accuracy	The probability that an observation in an individual reference category has been classified correctly. Producers' accuracy is represented in a sample as 1.0 (or 100%) minus the probability of 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.
Reference Data Value	In thematic accuracy assessment, the attributes (thematic values) that have been applied by the source of higher accuracy (typically ground observation) to a set of sample observations. In a ground-based accuracy assessment, these are often informally referred to as "field calls." The reference data are compared to the sample data to determine whether they correspond (i.e., are accuracy) or not (i.e., represent a thematic error).
Response Design	In thematic accuracy assessment, the methodology used to develop the reference data values (usually, the field identification methods).
Sample	A collection of sample units or observations selected together from an inference population for which parameters are estimated. The term sometimes is misapplied to the individual sample units.
Sample Data Value	In thematic accuracy assessment, the attributes (thematic value) that have been applied by the mapping process to a set of observations that has been selected to be tested for accuracy and is compared

against the reference data.

Sample Unit See observation.

Sampling Design In thematic accuracy assessment, the methodology used to assign individual observations within the management population (and inference population, if different from the management population.

Sampling Population See Inference area.

Simple Random Sampling A method of selecting observations (sample units) from an inference population such that every possible sample unit in the inference population has an equal probability of being selected.

Site (Observation Site) In thematic accuracy assessment, the geographic location associated with an observation, represented as a point and defined by a set of x and y coordinates. The observation area is normally centered on the site coordinates. See Observation Area.

Stand In the context of vegetation science, essentially synonymous with plant community"

Stratified Random Sampling A method of selecting observations from an inference population that involves first dividing the inference population into subpopulations (strata) that have similar characteristics, and then employing simple random sampling within each individual stratum. While every possible sample unit has an equal probability of being selected within its stratum, sampling units in different strata may have a different probability of being selected because of different stratum sizes and different sampling densities. Stratification is often performed to (1) describe the subpopulations, as well as the inference population and (2) to reduce the effects of heterogeneity within the inference population, based on known or expected stratum effects on the variable of interest, and (3) to account for differences in costs or logistics between strata.

Stratum In vegetation ecology, a distinct layer comprised of individual plants that share a common height and, often, a common growth form. In sampling theory, a division of an inference population to which a subsample of observations of that population is allocated.

Thematic Accuracy Absence of 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.

Thematic Accuracy Assessment	As used by the NPS Vegetation Inventory, a process used by the project oversight team (external to the mapping process) to check the accuracy of individual map classes. Unlike verification, the objective is to inform the user of map class limitations, rather than to address errors in the mapping process.
Type	See Vegetation Type
Users' Accuracy	The probability that an observation in an individual sample data category has been classified correctly (that a sample from the classified data actually represents that category on the ground). Users' accuracy is represented in a sample as 1.0 (or 100%) minus the probability of an error of commission.
Validation	As used by the NPS Vegetation Inventory, a process used by the project oversight team (external to the mapping process) to check the accuracy of the entire map against an overall accuracy standard. Unlike verification or accuracy assessment, the population of interest is the entire map, rather than individual map classes.
Variable	A specific value for a parameter.
Vegetation	The collective plant cover of an area or the collective defined categories of plant cover of an area.
Vegetation Type	A named category of plant community or vegetation defined on the basis of shared floristic and/or physiognomic characteristics that distinguish it from other kinds of plant communities or vegetation.
Verification	As used by the NPS Vegetation Inventory, a process used by production team (internal to the mapping process) to test the fit of the mapping model or the diagnostic user materials (field keys) against the ecological classification, with the objective of making adjustments or improvements prior to external evaluation steps (validation or accuracy assessment). Unlike calibration, verification requires a map to test.

Exhibit F. Examples of Contingency Tables

Table 15. Sample contingency table for thematic accuracy assessment, Thomas Stone National Historic Site.

	Upland Depression Swamp	Red Maple-Blackgum Seep. Swamp	Small Stream Sweetgum-Tuliptree Forest	Mesic Mixed Hardwood Forest	Successional Tuliptree Forest	Low-Elev Mixed Oak / Heath Forest	Successional Virginia Pine Forest	Successional Sweetgum Upland Forest	Dry Meadow	Lawn/Unvegetated	Undifferentiated Storm Residue	GRAND TOTALS, SAMPLE DATA VALUES ($n_{i\bullet}$)
Upland Depression Swamp	1											**1**
Red Maple - Blackgum Seepage Swamp		1										**1**
Small Stream Sweetgum - Tuliptree Forest			4									**4**
Mesic Mixed Hardwood Forest				11								**11**
Successional Tuliptree Forest		1	3	5	20	1	1	1				**32**
Low-Elevation Mixed Oak / Heath Forest				3	2	23		2				**30**
Successional Virginia Pine Forest						3	2					**5**
Successional Sweetgum Upland Forest								4				**4**
Dry Meadow									29			**29**
Lawn/Unvegetated										5		**5**
Undifferentiated Storm Residue											11	**11**
GRAND TOTALS, REFERENCE DATA VALUES ($n_{\bullet j}$)	**1**	**2**	**7**	**19**	**22**	**27**	**3**	**7**	**29**	**5**	**11**	**Total Observations (n): 133**

In this case, the map classes represent individual vegetation types, and the sample (row) data and reference (column) data classes are equivalent.

Table 16. Population contingency table for thematic accuracy assessment, Thomas Stone National Historic Site.

1 - Confidence limits are reported as 0%, if calculated value is less than 0%; confidence limits are reported as 100%, if calculated value exceeds 100%.	Upland Depression Swamp	Red Maple-Blackgum Seep. Swamp	Sweetgum-Tuliptree Ravine Forest	Mesic Mixed Hardwood Forest	Successional Tuliptree Forest	Mixed Oak / Heath Forest	Successional Virginia Pine Forest	Successional Sweetgum Forest	Dry Meadow	Lawn/Unvegetated	Undifferentiated Storm Residue	GRAND TOTALS ($p_{i\bullet}$)	MAP AREAS (A_i) (hectares)	POINT ESTIMATE OF MAP CLASS USERS' ACCURACY	LOWER LIMIT, 90% CONFIDENCE INTERVAL[1]	UPPER LIMIT, 90% CONFIDENCE INTERVAL[1]
Upland Depression Swamp	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0008**	**0.11**	**100.0%**	**95.2%**	**100.0%**
Red Maple-Blackgum Seep. Swamp	0.0000	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0002**	**0.03**	**100.0%**	**97.4%**	**100.0%**
Sweetgum–Tuliptree Ravine Forest	0.0000	0.0000	0.0332	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.0332**	**4.29**	**100.0%**	**85.3%**	**100.0%**
Mesic Mixed Hardwood Forest	0.0000	0.0000	0.0000	0.1066	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	**0.1066**	**13.77**	**100.0%**	**84.7%**	**100.0%**
Successional Tuliptree Forest	0.0000	0.0067	0.0200	0.0334	0.1334	0.0067	0.0067	0.0067	0.0000	0.0000	0.0000	**0.2135**	**27.58**	**62.5%**	**52.6%**	**72.4%**
Mixed Oak / Heath Forest	0.0000	0.0000	0.0000	0.0235	0.0157	0.1803	0.0000	0.0157	0.0000	0.0000	0.0000	**0.2352**	**30.38**	**76.7%**	**65.1%**	**88.2%**
Successional Virginia Pine Forest	0.0000	0.0000	0.0000	0.0000	0.0000	0.0088	0.0059	0.0000	0.0000	0.0000	0.0000	**0.0147**	**1.89**	**40.0%**	**34.4%**	**45.6%**
Successional Sweetgum Forest	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0387	0.0000	0.0000	0.0000	**0.0387**	**5.01**	**100.0%**	**84.1%**	**100.0%**
Dry Meadow	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2724	0.0000	0.0000	**0.2724**	**35.19**	**100.0%**	**86.4%**	**100.0%**
Lawn/Unvegetated	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0371	0.0000	**0.0371**	**4.79**	**100.0%**	**86.1%**	**100.0%**
Undifferentiated Storm Residue	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0477	**0.0477**	**6.16**	**100.0%**	**89.4%**	**100.0%**
GRAND TOTALS($\hat{p}_{\bullet j}$)	**0.0008**	**0.0068**	**0.0533**	**0.1635**	**0.1491**	**0.1957**	**0.0125**	**0.0611**	**0.2724**	**0.0371**	**0.0477**	**133**				
TRUE AREA ESTIMATE (A_j) (hectares)	**0.11**	**0.88**	**6.89**	**21.12**	**19.26**	**25.29**	**1.62**	**7.89**	**35.19**	**4.79**	**6.16**	**OVERALL ACCURACY: 85.6%** **90% CONF. INTERVAL: 79.1%-92.1%**				
POINT ESTIMATE, PRODUCERS' ACCURACY	**100.0%**	**3.5%**	**62.4%**	**65.2%**	**89.5%**	**92.1%**	**46.8%**	**63.4%**	**100.0%**	**100.0%**	**100.0%**					
LOWER LIMIT, 90% CONFIDENCE INTERVAL, PRODUCERS' ACCURACY[1]	**100.0%**	**2.2%**	**61.8%**	**64.6%**	**88.3%**	**90.6%**	**46.6%**	**62.8%**	**98.2%**	**99.8%**	**99.7%**	**KAPPA ACCURACY: 80.4%** **90% CONF. INTERVAL: 74.1%-86.6%**				
UPPER LIMIT, 90% CONFIDENCE INTERVAL, PRODUCERS' ACCURACY[1]	**100.0%**	**4.8%**	**62.9%**	**65.8%**	**90.7%**	**93.6%**	**46.9%**	**64.0%**	**100.0%**	**100.0%**	**100.0%**					

Exhibit H. Decision Tree for Relocating or Reshaping a Thematic Accuracy Assessment Reference Data Observation

1. Is a circular observation area (of radius as prescribed by the MMU size for that individual site) homogeneous (e.g., represents just one NVC Macrogroup)?

Yes……Identify the vegetation in a circular observation area around the site of the appropriate radius for the MMU size that has been specified for that site and record the plot center position in the field.

No…….Go to Step 2.

2. Can the observation area be made homogeneous at the prescribed MMU size for the vegetation represented at the site center by reconfiguring the plot shape?

Yes……..…….Reconfigure the plot shape and move the site center (as necessary) a minimal distance to make the observation area homogeneous for the vegetation type at the original plot center. Identify and record the vegetation type. Record the position center of the plot center in the field (whether relocated or not) and also the observation area limits (e.g., corners, if a rectangular plot is used).

No……..……. Go to Step 3.

3. Does the stand of the vegetation type that is represented at the site center extend to at least the size of the MMU designated for that site?

Yes…….Move the site (and reconfigure the observation area shape, as needed) minimally to make the observation area homogeneous. Record the new observation area center (if circular) and also the observation area limits, if not circular (e.g., corners, if a rectangular plot is used).

No…….……. Locate the stand of the next vegetation type in the original observation area that is nearest the original site. Move the site to the closest point within that stand and return to Step 2. Repeat as necessary.

If, after repeating Steps 2 and 3, it is found that no stand within the original observation area is the size of the MMU (i.e., all vegetation occurs in a mosaic smaller than the designated MMU size), identify each individual vegetation type within the original observation area and estimate the percentage of the observation area that each occupies.

Revised June 2010

Exhibit I: Accuracy Assessment Form

NPS Vegetation Inventory

1. PLOT (WAYPOINT) #: ____-____-_____ 2. DATE: _____

3. OBSERVER (DETERMINING ASSOCIATION) _____

4. Observer (assisting) _____

5. ACCURACY OF NAVIGATION (METERS) _____

6. How Determined: _____

7. UTM EASTING _____ 8. UTM NORTHING_____

9. UTM Zone _____ 10. Datum _____

11. Shape and Size of Observation Area (if not standard for project) _____

12. If GPS Position is an *intentional* offset from the waypoint, circle the explanation:

a. Mosaicing scenario (too heterogeneous to key because of two or more clearly distinct types within observation area)
b. Physical constraints in reaching waypoint
c. Other (explain as needed): _____

_____

_____

13. VEGETATION ASSOCIATION (Primary call): _____

_____

14. Problems with identifying vegetation type (#12):

Yes	No
☐	☐

15. Explanation for #13 (if "Yes" is checked) (Other possible vegetation type calls): _____

_____

_____

OPTIONAL FIELDS (USE AS NEEDED):

16. Dominant/characteristic species in tree layer (~ 1 – 5 species, where layer is present)

17. Dominant/characteristic species in shrub layer (1 – 5 species, where layer is present)

18. Dominant/characteristic species in herbaceous layer (1 – 5 species, where layer is present)

19. Other comments (if needed)

Accuracy Assessment Form Instructions

FIELDS DENOTED BY ALL CAPITAL LETTERS ARE MANDATORY FOR EACH PLOT (either digital or written or both). Other fields should be filled out as applicable and/or as needed. It may be possible to enter some items denoted in blue at the end of the field day to save time.

This form may be modified, as necessary, or convenient for field work and/or data entry.

1. Enter unique number for waypoint representing the accuracy assessment site (center of the observation area or plot). A suggested format is xxxx-mmm-AA-nnnn, where x represents the [official NPS] 4 letter park unit code, m denotes the size of the area to be observation for the individual site (where these vary within a project), AA designates the observation as a thematic accuracy assessment observation, and n represents a unique identifying number for the observation. Using this format, JOTR-050-AA-1012 represents site (observation) number 1012 in the Joshua Tree national Park thematic accuracy assessment, and observers are to evaluate this site at the scale of 0.5 hectares. For a campaign within a single park unit, the park acronym can be pre-printed on all sheets.
2. Enter Date (can enter in all sheets at end of day).
3. Enter Observer (full name) making final decision on determining dominant association for plot.
4. Enter other observer(s) who were present, if applicable.
5. Enter estimated accuracy of GPS navigation to real waypoint position (can enter on all sheets at end of day if same method was used for all waypoints; enter individually in field if a point-by-point estimate was made).
6. Enter how #6 was determined (e.g., EPE reading at waypoint, 99% precisions for recorded GPS point, etc.).
7. Enter UTM Easting (in meters to nearest meter).
8. Enter UTM Northing (in meters to nearest meter).
9. UTM Zone (if project occurs within a single zone, this can be pre-printed on the sheet).
10. Datum (normally, this would be NAD-83 (North American Datum 1983 and can be pre-printed).
11. If the plot is moved because of a (a) mosaicing situation or spatial complexity: two or more types in original observation area that are too different to evaluate as one observation, (b) physical constraints (e.g., waypoint is in middle of a cliff), or (c) other reasons, note this and why.
12. Record vegetation association dominating (occupying most area of) plot. Land use categories are acceptable, if one fits better than an association in the key. Try to decide on just one association (the “best fit”), but if the site is truly transitional between two or more intergrading types and you are not sure to which association it best fits, list the most likely association here and check “Yes” box in 13. Explain rationale for other type possibilities (refer to key, as necessary) in #14. May use “Other” if nothing makes sense, but explain in #14 if you do.
13. Check “Yes” or “No.” See instruction for #12.
14. Explanation for #12 and #13, if needed. See instructions for #12 and #13.

OPTIONAL FIELDS:

15. List the most abundant, frequent, and/or characteristic species (Latin names) in the tree layer (suggest 0-5 species). Mark names of species that are significantly more abundant than the others (i.e., strongly dominant) with an asterisk (*).
16. List the most abundant, frequent, and/or characteristic species (Latin names) in the shrub layer (suggest 0-5 species). Mark names of species that are significantly more abundant than the others (i.e., strongly dominant) with an asterisk (*).
17. List the most abundant, frequent, and/or characteristic species (Latin names) in the herbaceous layer (suggest 0-5 species). Mark names of species that are significantly more abundant than the others (i.e., strongly dominant) with an asterisk (*).
18. Use this space to make comments on other items that are confusing or require more detailed explanation.

The Department of the Interior protects and manages the nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its special responsibilities to American Indians, Alaska Natives, and affiliated Island Communities.

NPS 999/102741, June 2010

National Park Service
U.S. Department of the Interior

Natural Resource Program Center
1201 Oakridge Drive, Suite 150
Fort Collins, CO 80525

www.nature.nps.gov