Accuracy Assessment

Purpose

The accuracy assessment estimates thematic errors in the data, providing users the information needed to assess data suitability for a particular application. At the same time, data producers are able to learn more about the nature of errors in the data. Thus, there are actually two views to an accuracy assessment: users' accuracy", which is the probability that an accuracy assessment point has been mapped correctly (also referred to as errors of commission) and "producers' accuracy," which checks to see if the map actually represents what was found on the ground (also referred to as errors of omission). With users' accuracy, the number of correctly classified samples of a map class is divided by the total number of field samples that were classified in that map class. The emphasis here is on the reliability of the map, or how well the map represents what is really on the ground. With producers' accuracy, the number of correctly classified samples of a map class is divided by the total number of field samples of that map class. The emphasis here is on the probability that the ground field samples have been correctly classified. Both users' and producers' accuracy can be obtained from the same set of data using different analyses. Errors occur when map classes are not the same as the classes observed in the field. A major assumption of accuracy assessment is that the process of mapping and the process of the assessment (i.e., the application of the classification system) are identical, so that a false error is not detected because of procedural differences.

Results of the accuracy assessment are presented in an error or misclassification matrix (also referred to as a contingency or confusion matrix). The accuracy numbers are interpreted as the probability of encountering a particular map class when visiting a particular spot, or point, not a particular polygon. Accuracy requirements for the project specify 80% overall (the proportion of correctly assessed sites) accuracy for each vegetation map class.

Sampling Design

The objectives of collecting samples for the accuracy assessment is to obtain a measure of the probability with which a particular location has been assigned its correct vegetation class. We used a stratified random sampling approach that covered most park fee areas (and some easement areas). For logistical reasons, we did not include in our sampling approach the numerous smaller islands within Penobscot Bay that encompass many of the Park's easement lands (discussed later in more detail). Because of access constraints, we did not include in the design areas mapped outside the park. Maximum and minimum number of samples per map class theme followed program recommendations (The Nature Conservancy et al. 1994), as suggested in the following scenarios:

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

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

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

Scenario D: The class is rare. It has more than 5 but fewer than 30 polygons and covers less than 50 ha of the area. In this case, the recommended number of samples is 5. The rationale for reducing the sample size is that the class consists of small polygons and the frequency of the polygons is low. Specifying more than 5 sample sites will therefore probably result in multiple sample sites within the same (small) polygon. Collecting 5 sample sites will allow an accuracy estimate to be computed, although it will not be very precise.

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

The recommendations above take into account both the statistical and operational aspects of sampling. The accuracy estimate associated with rare classes cannot be stated with the same level of confidence as that associated with classes that are more abundant. For example, with a sample size of 5, the level of error in the estimate is closer to 25% at a 90% confidence level, as opposed to 10% with a sample size of 27. This has implications for our ability to accept a given point estimate as meeting accuracy requirements. Whether or not a given accuracy estimate is accepted as meeting requirements depends on the width of the confidence interval associated with the point estimate and the outcome of a hypothesis test that determines if a given point estimate is equivalent to or exceeds requirements.

We randomly stratified all accuracy assessment site locations across the vegetation map data that are within lands that could be accessed by the field crew. We determined accessible lands for accuracy assessment by park ownership and ease of access. We determined that all Acadia NP lands on Mount Desert Island, Schoodic Peninsula, Isle au Haut, and Long Island were accessible. In contrast, we determined that the numerous small islands in the ocean (most under Park easement) were too remote and difficult to access, requiring considerable more time and logistical maintenance. In consolation, these islands express vegetation communities that are quite extensive throughout the areas we determined accessible (e.g., maritime spruce-fir forest). We had determined areas for accuracy assessment early on in the mapping process, which allowed us to prioritize our mapping efforts for the accuracy assessment field season.

While we had completed our initial photointerpretation process prior to the field season, our subsequent digital mapping of the interpreted data extended into the field season. As we continued with our digital mapping, focusing on the areas for accuracy assessment, we provided the field crew locations (GPS coordinates and maps) in segments as we continued our mapping. We separated the park accessible lands into four major segments (phases), sending site location data to the field crew shortly after we completed them. These four phases are as follows:

- Phase I covered the western third of Mound Desert Island
- Phase II covered the Schoodic Peninsula, Isle au Haut, and Long Island
- Phase III covered the eastern third of Mound Desert Island
- Phase IV covered the central third of Mound Desert Island

As the field season continued, our digital mapping of the assessment area concluded. We were able to provide the field crew all site locations in time for successful data collection during the 1999 field season.

We had determined number of samples needed per map class (taking in account all phase areas) prior to the site selection process. We extrapolated the number of sample sites needed per phase area by analyzing area reports of map phase areas already complete coupled with the photo interpreter's knowledge of map class distribution for the phase areas that we were digitally mapping at the time. Based on these results, we distributed each map theme's number of samples across each of the four phase areas. Three times the number of sites needed was randomly generated using a software program. We did this for two reasons. One is that, in our experience, PLGR units often express up to 10 m in reading errors, particularly in dense conifer forests (a signature for Acadia NP). By eliminating random generated GPS coordinates that fall near polygon boundaries, we anticipated fewer GPS field collected coordinates displace into neighboring polygons. The second reason was to reduce any other accessibility issues that our pre-processing of access areas did not address. An example of this is a remote site distanced far from other sites or access points (e.g., roads, trails) requiring high investment of time, energy, and logistical planning to access. After reducing the over-selection back to the designated number of sites per theme (map class), we had selected 728 sites.

To prepare the field team with locating assessment sites, we plotted 1:12,000-scale orthophoto quadrangle hardcopy maps (from USGS 3.75-minute digital orthophoto quadrangle images) showing locations of the accuracy assessment sites, the unlabelled polygon boundaries of the vegetation map, and the park boundary. We sent Acadia NP staff the field site coordinates (projection in UTM, Zone 19, and datum in NAD83), which they in turn uploaded into a PLGR GPS unit. We also provided the field crew with written instructions for general navigational and data collection methods and with data sheets.

Data Collection Methods

The accuracy assessment team used the PLGR GPS unit to navigate to each site. They also used the hard copy orthophoto maps showing the accuracy assessment site, along with the Project's aerial photographs, to navigate around environmental barriers (e.g., lakes, ponds, deep marshes). Once the sampling site was reached, they evaluated the plant community within a 0.5-hectare radius (the minimum mapping unit or MMU) using the key to vegetation types (see Appendix B: Dichotomous Keys to the Vegetation Communities at Acadia National Park). They also assigned a provisional vegetation community name to the site and recorded the field GPS coordinate location, dominant species, environmental data, and pertinent comments (see Appendix C: Example of an Accuracy Assessment Form for a sample data sheet). If the area was not homogeneous (containing more than one vegetation association), the other associations were also listed on the data sheet. The field team collected data for 724 sites using this method (Figure 16), nearly all of the 728 selected sites (an outstanding achievement).

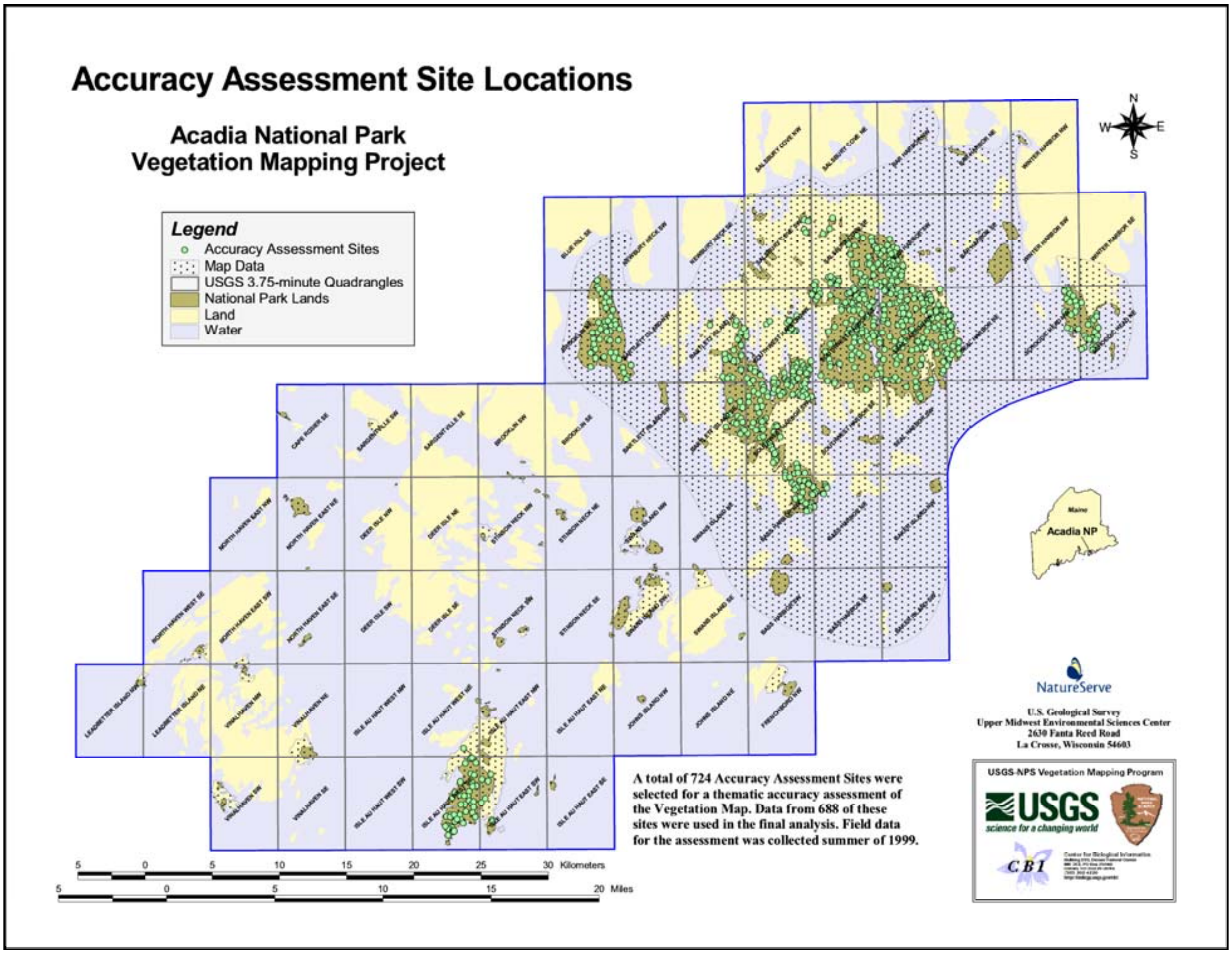

Figure 16. Locations of assessment sites sampled for the Acadia National Park Vegetation Mapping Project.

Data Analysis

The accuracy assessment data were entered into the PLOTS database (The Nature Conservancy 1997) and subsequently reviewed for data entry errors. The analysis of the map accuracy using the field data includes the following steps:

- Initial regrouping of vegetation types and map classes to make a 1:1 relation
- Initial comparison analysis of field and map data
- Initial review of all disagreements and making data adjustments as necessary
- Further classification adjustments and comparison analysis of field and map data
- Final output of results into a contingency matrix

Initial Classification Relations

Although some map classes and vegetation communities (associations) have a 1:1 relation to each other, some do not. Some map classes represent two or more vegetation types. In contrast, some vegetation types are represented with more than one map class (e.g., to map variations of a community, subtype). More on this is discussed in the Results and Discussion, Map Classification section. To properly compare the map class data with the field assessed vegetation types, we regrouped map classes and vegetation communities such that a 1:1 relation exists.

Initial Comparison Analysis

With the 1:1 relations between the two classifications in place, we intersected the field point data with the map polygon data. This allowed us to compare each field accuracy assessment call to the corresponding polygon map class code. PROC FREQ (SAS 1996) was used to compare and tabulate the total number of field assessment sites and map polygons that were in agreement.

Initial Review of Disagreements

All mismatches (disagreements) were subsequently reviewed to see if there were any "false errors." A false error is defined as a mismatch between the map polygon and an accuracy assessment call if caused by any of the following: (1) error in GPS field coordinate, (2) map agreement to an alternate field call, (3) misapplied field call (e.g., from misapplication of the vegetation key), or (4) field site assessment area smaller than the polygon minimum mapping unit (MMU). This review process involved looking at every polygon and its corresponding accuracy assessment site on the photos. We used both the accuracy assessment site and the vegetation map coverages in ArcView GIS to help us locate the sites on each photo. The field data sheet was usually reviewed to gain a fuller context of the ground data. From this process, disagreements that were deemed "false errors" were corrected, resulting in either a match or a true error.

(1) Spatial GPS coordinate errors occur when the field collected GPS coordinate has slight inaccuracies in geo-positional placement, moving the coordinate just inside an adjacent polygon and acquiring a map class different from that intended for the actual area assessed in the field. Through our sampling design (selecting sites more than 10 m from polygon edges), we were able to reduce these errors. There are limitations to the design approach, however, especially with narrow corridor shaped polygons. For sites determined to have spatial GPS field coordinate displacement, we adjusted accordingly for the analysis to reflect the intended polygon's map class. (We left the accuracy assessment database intact, preserving the actual field coordinate locations.)

Some GPS coordinate errors are due merely to incorrect database entry. We assessed these types of errors by reviewing the field data sheets, complimented with accessing the original selected site coordinates using GIS (as an additional measure to assertain proper site coordinate lcation). Some coordinates could not be successfully recovered and thus dropped from the analysis. Of those that could be recovered, the accuracy assessment database was updated to reflect the correction.

(2) Alternate vegetation communities were often recorded on the field data sheets when the site being assessed was not clear between closely related vegetation types. With these alternate calls entered into the database in a secondary field, they were not included with the comparison analysis (only the primary or initial field call was used). Upon manual review of the field data sheets, if the alternate vegetation community matched the vegetation map, the assessment was adjusted to give the map the benefit, an approach approved by the VMP. (In future, comparing the map data to both the primary field call and all subsequent alternate calls using computerized automation techniques might expedite the review process and reduce the tedious manual approach taken. However, this did encourage us to look deeper into vegetation community concepts, understand how they relate to other closely related types, or understand how those relations correspond to the vegetation map.)

(3) In some instances, the analysis team might question the field assessment call based on the final vegetation key and final community descriptions. During 1999, the vegetation key was in draft, and in one sense, being tested with the accuracy assessment. Vegetation community

descriptions of Acadia NP had not yet been written as the vegetation analyses was not yet complete. In these cases, vegetation classifiers reviewed the data sheets. We updated our analysis tables to reflect any changes in the classified community type in preparation for the second comparison analysis (the project's vegetation database was updated, too).

(4) The area of which some sites are assessed in the field might fall below the MMU for mapping (termed as an inclusion). We discovered instances where, after reviewing the aerial photographs, the site was found to be an inclusion to the surrounding vegetation type. Certain vegetative features can be quite apparent from each other while viewing the aerial photographs (e.g., sparse vegetation on rock outcrop versus dense stand of conifer trees), allowing easy assessment in the lab of site inclusions. In these cases, the map again was given the benefit.

Additional Classification Adjustments and Comparison Analysis

As a side benefit to the in depth review of all disagreements between the accuracy assessment sites and the vegetation map, we began to notice consistent diverging patterns between the map and field assessment data. At this point, we began adjusting the map through a series of "global" changes, digitally changing the classification in the map (that is, globally changing the classification of entire groups of like-classified map polygons) to better align with the final version of vegetation classification (final version of the classification was completed prior to final accuracy assessment). For example, we combined selected wetland forested map classes into one group to account for several conceptual differences between the vegetation classification and the map classes. Another example, we collapsed two alder map classes into one.

Also, from the detailed review, we recognized additional map classes that merely represent an expression (or, in part) of vegetation types. We wanted these expressions preserved in the vegetation map database, while at the same time combine the subtype mapping for the accuracy assessment. To do so, we combined those map classes only for this analysis, leaving the map classes intact in the spatial database. For example, with our analysis we combined the conifer dominant spruce - fir forest map class with that of the mixed conifer-deciduous expression (two map classes representing different expressions of the same vegetation types).

Of the 724 accuracy assessment sites originally collected, we dropped 36 from the analyses. 17 were due to irresolvable GPS errors. The other 19 were due to a disjointed map class that had to be eliminated (19 cases). We discovered at this time that our defining concepts for this map class were incompatible with its counterpart in the vegetation classification (we reinterpreted areas originally mapped with this class to other various other valid map classes). We considered another 72 sites to be inclusions and corrected these to reflect the surrounding area that was of mappable size (adjusted for this analysis only and not in the Project's accuracy assessment database). A total of 688 accuracy assessment sites were used for the final analysis.

With each "false" discrepancy now reflecting proper assignments (whether now a match, or remains a disagreement), and revisions made to the vegetation map to better reflect the vegetation classification, we once performed a comparison analysis of the field data and vegetation map data, once again using PROQ FREQ (SAS 1996).

Contingency Table

We transferred the final set of numbers generated from this last analysis into a contingency table (matrix), where we calculated user and producer accuracy percentages for each map class (theme). The matrix shows both the frequency of agreement and the placement (and frequency thereof) of disagreements.