6 Reference Data Collection

Collection of the reference data for use in an accuracy assessment is a key component of any assessment. Failure to collect appropriate reference data produces erroneous results, dooming the assessment. The collection of accuracy assessment data requires completing the following three steps while considering both the reference data being collected and the map being assessed:

- First, the accuracy assessment sample sites must be accurately located both on the reference source and on the map. This can be a relatively simple task in an urban area, or a far more difficult one in a wildland area where few recognizable landmarks exist. While the arrival of GPS has greatly increased our ability to locate accuracy sites, it is still possible to misidentify the location of a site.
- Next, the sample unit must be delineated. Sample units must represent exactly the same area on both the reference data and the map. Usually they are delineated once, either on the reference source data or on the map, and then transferred to the other. However, if the source of the reference data is not accurately coregistered to the map being assessed, slivers of unaligned assessment sites may be created that can greatly confuse the assessment.
- Finally, the reference and map labels must be assigned to each sample unit based on the map classification scheme. The reference labels may be collected from a variety of sources, and may be captured either through observation or measurement.

Serious oversights and problems can arise at each step of data collection. To adequately assess the accuracy of the remotely sensed classification, each step must be implemented correctly on each and every sample. If the reference labels are inaccurate, then the entire assessment becomes meaningless. Four basic considerations drive all reference data collection:

1. What should the source be for the reference data samples? Can existing maps or existing field data be used as the reference data? Should the information be collected from remotely sensed data or are new field visits required?
2. What type of information should be collected for each sample? Should measurements be taken or are observations adequate?

3. When should the reference data be collected? Should it be collected during initial field investigations when the map is being made, or should it only be collected after the map is completed? What are the implications of using old data for accuracy assessment?
4. How can we ensure that the reference data are collected correctly, objectively, and consistently?

There are many methods for collecting reference data, some of which depend on making observations (qualitative assessments) and some which require detailed, quantitative measurements. Given the varied reliability, difficulty, and expense of collecting reference data, it is critical to know which of these data collection techniques are valid and which are not for any given project.

We all understand that maps are rarely 100% correct. Each remote sensing project requires trade-offs between the remotely sensed data used to create the map and the scale and level of accuracy required by the project. We accept some level of map error as a trade-off for the cost savings inherent in using remotely sensed data to create the map. However, accuracy assessment reference labels must be correct if they are to constitute a fair assessment of the map. Thus, reference labels must be derived using source data or methods that are assumed to be more reliable than those used to make the map.

WHAT SHOULD BE THE SOURCE OF THE REFERENCE DATA?

The first decision in data collection requires determining what source data will be used for the determination of reference labels. The type of source data required will depend on the complexity of the map classification scheme and the budget. It is best to keep in mind this general rule: the simpler the classification scheme, the simpler and less expensive the reference data collection.

Sometimes, previously existing maps or ground data can be used as the reference data. More often, the reference source data are newly collected information that is at least one level more accurate than the remotely sensed data and methods used to make the map. Thus, aerial photography is often used to assess the accuracy of maps made from moderate-resolution satellite imagery (e.g., SPOT and Landsat TM), ground visits are often used to assess the accuracy of maps created from high-resolution airborne imagery, and manual image interpretation is often used to assess the accuracy of automated classification methods.

Using Existing versus Newly Collected Data

When a new map is produced, the usual first reaction may be to compare the map to some existing source of information about the mapped area. Using previously collected ground information or existing maps for accuracy assessment is tempting because of the cost savings resulting from avoiding new data collection. While this can be a valuable qualitative tool, existing data are rarely acceptable for accuracy assessment because:

1. The classification schemes employed to create existing maps usually differ from the one being used to create the new map. Comparisons between the two maps can result in the error matrix expressing merely differences between the reference data and the map data classification schemes, rather than map error. Developing a crosswalk that specifically translates the map classification scheme to the new map classification scheme can sometimes solve this problem. However, this method rarely produces a perfect crosswalk and, therefore, some error is unavoidable.
2. Existing data are older than those being used to create the new map. Changes on the landscape (e.g., fire, urban development, etc.) will not be reflected in the existing data. Therefore, differences in the error matrix caused by these changes will incorrectly be assumed to be caused by map error.
3. Errors in existing maps are rarely known (it is unlikely that an accuracy assessment has been performed on the existing map). Often, the differences caused by existing map errors are then blamed on the new map, thereby wrongly lowering the new map's accuracy.
4. Existing field inventory data usually were collected for a purpose other than accuracy assessment. Often the size of the inventory plot is too small (e.g., a 1 m ecology site cannot be used to assess a map with a 4 m minimum mapping unit) or the measurements made on the plots cannot be transformed into measurements that are useful for the accuracy assessment.

If existing information is the only available source of reference data, then consideration should be given to not performing quantitative accuracy assessment. Instead, a qualitative comparison of the new map and existing map or field data should be performed, and the differences between the two should be identified and scrutinized. If a quantitative assessment is performed with existing data, it is vital to document the issues with the reference data so as to allow the potential user of the map to understand the limitations of such an assessment.

Photos versus Ground

If new data are to be collected for reference samples, then a choice must be made between using ground visits and using aerial imagery, video, or reconnaissance as the source of the reference data. The accuracy assessment professional must assess the reliability of each data type for obtaining an accurate reference site label.

Simple classification schemes with a few general classes can often be reliably assessed from air reconnaissance or interpretation of aerial imagery or video. As the level of detail in the map classification scheme increases, so does the complexity of the reference data collection. Eventually, even the very largest-scale photography cannot provide valid reference data. Instead, the data must be collected on the ground.

In some situations, the use of image interpretation or videography for generating reference data may not be appropriate. For example, aerial photo interpretation is often used as reference data for assessing a land cover map generated from satellite imagery such as Landsat Thematic Mapper. The photo interpretation is assumed to be correct because the photos have greater spatial resolution than the satellite imagery and because photo interpretation has become a time-honored skill that is accepted as accurate. Unfortunately, errors do occur in photo interpretation and air recognizance depending on the skill of the photo interpreter and the level of detail required by the classification system. Inappropriately using photo interpretation as reference data could severely impact the conclusions regarding the accuracy of the satellite-based land cover map. In other words, one may conclude that the satellite-based map is of poor accuracy when actually it is the photo interpretation that is in error.

In such situations, actual ground visitation may be the only reliable method of data collection. At the very least, a subset of data should be collected on the ground and compared with the airborne data to verify the reliability of the reference labels interpreted from airborne imagery. Even if the majority of reference labels will come from image interpretation or videography, it is critical that a subsample of these areas be visited in the field to verify the reliability of the interpretation. Much work is yet to be done to determine the proper level of effort and collection techniques necessary to provide this vital information. When the labels developed from image interpretation and the ground begin to disagree regularly, it may be time to switch to ground-based reference data collection. However, the collection of ground reference data is extremely expensive and, therefore, the collection effort must be sufficient to meet the needs of the accuracy assessment while being efficient enough to stay within the budget.

For example, Biging et al. (1991) compared photo interpretation to ground measurements for characterizing forest structure. These characteristics included forest species, tree size class, and crown closure, which were photo-interpreted by a number of expert photo interpreters. The ground data used for comparison were a series of measurements made in a sufficient number of ground plots to characterize each forest polygon (i.e., stand). The results showed that the overall accuracy of photo interpretation of species ranged in accuracy between 75 and 85%. The accuracy of size class was around 75%, and the accuracy of crown closure was less than 40%. This study reinforces the need to be careful if assuming that the results of the photo interpretation are sufficient for use as reference data in an accuracy assessment.

HOW SHOULD THE REFERENCE DATA BE COLLECTED?

The next decision in reference data collection involves deciding how information will be collected from the source data to obtain a reliable label for each reference site. Reference data must be labeled using the same classification scheme that was used to make the map. In many instances, simple observations/interpretations are sufficient for labeling a reference sample. In other cases, observation is not adequate, and actual measurements in the field are required.

The purpose of collecting reference data for a sample site is to derive the "correct" reference label for the site for comparison to the map label. Often, the reference label can be obtained by merely observing the site from an airplane, car, or aerial imagery. For example, in most cases, a golf course can be accurately identified through observation from quite a distance away.

Whether or not accuracy assessment reference data should be obtained from observations or measurements will be determined by the complexity of the landscape, the detail of the classification scheme, the required precision of the accuracy assessment, and the project budget. Reference data for simple classification schemes that distinguish homogeneous land cover or use types from one another (e.g., water versus agriculture) usually can be obtained from observations and/or estimations either on the ground or from larger-scale remotely sensed data. For example, distinguishing conifer forest from an agricultural field from a golf course can be determined from observation alone. Collecting reference data may be as simple as looking at aerial imagery or observing sites on the ground.

However, complex classification schemes may require some measurements to determine the precise (i.e., nonvarying) reference site labels. For example, a more complex forest classification scheme may involve collecting reference data for tree size class (related to the diameter of the trunk). Tree size class is important both as a determinant of endangered species habitat and as a measurement of wood product merchantability. Size class can be ocularly estimated on aerial imagery and on the ground. However, different individuals may produce different estimations, introducing variability into the observation. Not only will this variation exist between individuals, but also within the same individual. The same observer may see things differently depending on whether it is Monday or Friday; or whether it is sunny or raining; or especially depending on how much coffee he or she has consumed. To avoid variability in human estimation, size class can be measured, but a great many trees will need to be measured to precisely estimate the size class for each sample unit. In such instances, the accuracy assessment professional must decide whether to require measurement (which can be time consuming and expensive) or to accept the variation inherent in human estimations.

Whether or not measurements are required depends on the level of precision required by the map users and on the project budget. For example, information on spotted owl habitat requirements indicates that the owls prefer older, multistoried stands that include large trees. "Large" in this context is relative, and precise measurements of trees will probably not be needed as long as the map accurately distinguishes between stands of single-storied small trees and multistoried large trees. In contrast, many wood product mills can only accept trees within a specific size class. Trees that are one inch smaller or larger than the prescribed range cannot be accepted by the machinery in the mill. In this case, measurements will probably be required.

Observer variability is especially evident in estimates of vegetation cover that cannot be precisely measured from aerial imagery. In addition, ground verification of aerial estimates of vegetative cover is problematic as estimates of cover from the ground (i.e., below tree canopies) are fundamentally different from estimates made from above the canopy. Spurr (1960) asserts that forest crown closure is overestimated from aerial imagery and underestimated from the ground. Therefore, using ground estimates as reference data for aerial cover estimates can be like comparing apples and oranges.

The trade-offs inherent between observation and measurement are exemplified in a pilot study conducted to determine the level of effort needed to collect appropriate ground reference data for use in forest inventory. The objective of this study was to determine if visual calls made by trained experts walking into forest polygons are sufficient or whether actual ground measurements need to be made. There are obviously many factors influencing the accuracy of ground data collection, including the complexity of the vegetation itself. A variety of vegetation complexities were represented in this study. The results are enlightening to those remote sensing specialists who routinely collect forest ground data only by visual observation. The pilot study was part of a larger project aimed at developing the use of digital remotely sensed data for commercial forest inventory (Biging and Congalton, 1989).

Commercial forest inventory involves much more than creating a land cover map derived from digital remotely sensed data. Often the map is used only to stratify the landscape; a field inventory is conducted on the ground to determine tree volume statistics for each type of stand of trees. A complete inventory requires that the forest type, the size class, and the crown closure of a forested area be known in order to determine the volume of the timber in that area. If a single species dominates, the forest type is commonly named by that species (Eyre, 1980). However, if a combination of species is present, then a mixed label is used (e.g., the mixed conifer type). The size of the tree is measured by the diameter of the tree at 4.5 ft above the ground (i.e., diameter at breast height, DBH) and then is divided into size classes such as poles, small saw timber, and large saw timber. This measure is obviously important because large-diameter trees contain more volume of high-quality wood (i.e., valuable timber) than small-diameter trees. Crown closure, as measured by the amount of ground area the tree crowns occupy (canopy closure), is also an important indicator of tree size and numbers. Therefore, in this pilot study, it was necessary to collect ground reference data not only on tree species/type, but on crown closure and size class as well.

Ground reference data were collected using two approaches. In the first approach, a field crew of four entered a forest stand (i.e., polygon), observed the vegetation, and came to a consensus regarding a visual call of dominant species/type, size class of the dominant species, crown closure of the dominant size class, and crown closure of all tree species combined. Dominance was defined as the species or type comprising the majority of forest volume. In the second approach, measurements were conducted on a fixed-radius plot to record the species, DBH, and height of each tree falling within the plot. A minimum of two plots (1/10th or 1/20th acre) were measured for each forested polygon. Because of the difficulty of making all the required measurements (precise location and crown width for each tree in the plot) to estimate crown closure on the plot, an approach using transects was developed to determine crown closure. A minimum of four 100-ft-long transects randomly located within the polygon were used to collect crown closure information. The percentage of crown closure was determined by the presence or absence of tree crown at 1-ft intervals along the transects. All the measurements were input into a computer program that categorized the results as dominant species/type, the size class of the dominant species/type, the crown closure of the dominant size class, and the crown closure of all tree species for each forested area. The results of the two approaches were compared by using an error matrix.

Table 6.1 shows the results of field measurement versus visual call as expressed in an error matrix for the dominant species. This table indicates that species can be fairly well determined from a visual call because there is strong agreement between the field measurements and the visual call. Of course, this conclusion requires one to assume that the field measurements are a better measure of ground reference data, a reasonable assumption in this case. Therefore, ground reference data collection of species information can be maximized using visual calls, and field measurements appear to be unnecessary.

TABLE 6.1
Error Matrix for the Field Measurement versus Visual Call for Dominant Species

		Field Measurement							Row
		TF	MC	LP	DF	PP	PD	OAK	Total
Visual Call	TF	14	0	0	0	0	0	0	14
	MC	0	10	0	0	0	2	0	12
	LP	0	0	1	0	0	0	0	1
	DF	0	1	0	8	0	0	0	9
	PP	1	1	0	0	0	0	0	2
	PD	0	0	0	1	0	0	0	1
	OAK	0	0	0	0	0	0	0	0
	Column Total	15	12	1	9	0	2	0	39

Species

TF = true fir
MC = mixed conifer
LP = lodgepole pine
DF = Douglas fir
PP = Ponderosa pine
PD = PP and DF
OAK = oaks

OVERALL ACCURACY = 33/39 = 85%

PRODUCER'S ACCURACY

TF = 14/15 = 93%
MC = 10/12 = 83%
LP = 1/1 = 100%
DF = 8/9 = 89%
PP = 0/0 = —
PD = 0/2 = 0%
OAK = 0/0 = —

USER'S ACCURACY

TF = 14/14 = 100%
MC = 10/12 = 83%
LP = 1/1 = 100%
DF = 8/9 = 89%
PP = 0/2 = 0%
PD = 0/1 = 0%
OAK = 0/0 = —

Source: Reproduced with permission from the American Society for Photogrammetry and Remote Sensing, from Congalton R. and G. Biging. 1992. A pilot study evaluating ground reference data collection efforts for use in forest inventory. *Photogrammetric Engineering and Remote Sensing*. 58(12): 1669–1671.

TABLE 6.2
Error Matrix for the Field Measurement versus Visual Call for Dominant Size Class

Visual Call	Field Measurement 1	2	3	4	Row Total
1	1	0	0	0	1
2	1	3	1	0	5
3	0	0	17	5	22
4	0	0	1	11	12
Column Total	2	3	19	16	40

Size Classes

1 = 0–5″ dbh
2 = 5–12″ dbh
3 = 12–24″ dbh
4 = >24″ dbh

OVERALL ACCURACY
= 32/40 = 80%

PRODUCER'S ACCURACY

1 = 1/2 = 50%
2 = 3/3 = 100%
3 = 17/19 = 89%
4 = 11/16 = 69%

USER'S ACCURACY

1 = 1/1 = 100%
2 = 3/5 = 60%
3 = 17/22 = 77%
4 = 11/12 = 92%

Source: Reproduced with permission from the American Society for Photogrammetry and Remote Sensing, from Congalton R. and G. Biging. 1992. A pilot study evaluating ground reference data collection efforts for use in forest inventory. *Photogrammetric Engineering and Remote Sensing*. 58(12): 1669–1671.

Table 6.2 presents the results of comparing the two ground reference data collection approaches for the dominant size class. As in species, the overall agreement is relatively high, with most of the confusion occurring between the larger classes. The greatest inaccuracies result from visually classifying the dominant size class (i.e., the one with the most volume) as size class three (12–24 in. DBH) when, in fact, size class four (>24 in. DBH) trees contained the most volume. This visual classification error is easy to understand. Tree volume is directly related to the square of DBH. There are numerous cases when a small number of large trees contribute the majority of the volume in the stand, while there may be many more medium-sized trees present. The dichotomy between prevalence of medium-sized trees, but dominance in volume by a small number of large trees can be difficult to assess visually. It is likely that researchers and practitioners would confuse these classes in cases where the size class with the majority of volume was not readily evident. In cases like this, simply improving one's ability to visually estimate diameter would not improve one's ability to classify size class. The ability to weigh numbers and sizes to estimate volume requires considerable experience and would certainly require making plot and tree measurements to gain and retain this ability.

Tables 6.3 and 6.4 show the results of comparing the two collection approaches for crown closure. Table 6.3 presents the crown closure of the dominant size class results, while Table 6.4 shows the results of overall crown closure. In both matrices,

TABLE 6.3
Error Matrix for the Field Measurement versus Visual Call for Density (Crown Closure) of the Dominant Species

Visual Call	Field Measurement: O	L	M	D	Row Total
O	10	8	3	0	21
L	2	8	1	1	12
M	0	3	1	1	5
D	0	1	0	0	1
Column Total	12	20	5	2	39

Density Classes

O = Open (0–10% crown closure)
L = Low (11–25% crown closure)
M = Medium (26–75% crown closure)
D = Dense (> 75% crown closure)

OVERALL ACCURACY
= 19/39 = 49%

PRODUCER'S ACCURACY

O = 10/12 = 83%
L = 8/20 = 40%
M = 1/5 = 20%
D = 0/2 = 0%

USER'S ACCURACY

O = 10/21 = 48%
L = 8/12 = 67%
M = 1/5 = 20%
D = 0/1 = 0%

Source: Reproduced with permission from the American Society for Photogrammetry and Remote Sensing, from Congalton R. and G. Biging. 1992. A pilot study evaluating ground reference data collection efforts for use in forest inventory. *Photogrammetric Engineering and Remote Sensing*. 58(12): 1669–1671.

there is very low agreement (46–49%) between the observed estimate and the field measurements. Therefore, it appears that field measurements may be necessary to obtain precise measures of crown closure and that visual calls, although less expensive and quicker, may vary at an unacceptable level.

In conclusion, it must be emphasized that this is only a small pilot study. Further work needs to be conducted in this area to evaluate ground reference data collection methods and to include the validation of aerial methods (i.e., image interpretation and videography). The results demonstrate that making visual calls of species are relatively easy and accurate, except where many species occur simultaneously. Size class is more difficult to assess than species, because of the implicit need to estimate the size class with the majority of volume. Crown closure is by far the toughest to determine. It is most dependent on where one is standing when the call is made. Field measurements, such as the transects used in this study, provide an alternative means of determining crown closure. This study has shown that at least some ground data must be collected using measurements, and it has suggested that a multilevel effort may result in the most efficient and practical method for collection of ground reference data.

Table 6.5 presents the pros and cons of the different sources of reference data.

TABLE 6.4
Error Matrix for the Field Measurement versus Visual Call for Overall Density (Crown Closure)

Visual Call	Field Measurement: O	L	M	D	Row Total
O	0	1	1	0	2
L	1	3	7	0	11
M	0	0	8	10	18
D	0	0	0	6	6
Column Total	1	4	16	16	37

Density Classes

O = Open
L = Low
M = Medium
D = Dense

OVERALL ACCURACY
= 17/37 = 46%

PRODUCER'S ACCURACY

O = 0/1 = 0%
L = 3/4 = 75%
M = 8/16 = 50%
D = 6/16 = 38%

USER'S ACCURACY

O = 0/2 = 0%
L = 3/11 = 27%
M = 8/18 = 44%
D = 6/6 = 100%

Source: Reproduced with permission from the American Society for Photogrammetry and Remote Sensing, from Congalton R. and G. Biging. 1992. A pilot study evaluating ground reference data collection efforts for use in forest inventory. *Photogrammetric Engineering and Remote Sensing*. 58(12): 1669–1671.

WHEN SHOULD THE REFERENCE DATA BE COLLECTED?

The world's landscape is constantly changing. If change occurs between the date of capture of the remotely sensed data used to create a map and the date of the reference data collection, accuracy assessment reference sample labels may be affected. When a crop is harvested, a wetland drained, or a field developed into a shopping mall, the error matrix may show a difference between the map and the reference label that is not caused by map error, but rather by landscape change.

As previously noted, aerial or high-resolution satellite imagery is often used as reference source data for accuracy assessment of forest type maps created from Landsat TM or SPOT satellite data. Because aerial imagery is relatively expensive to obtain, existing imagery that may be 5–10 years old may be used. If an area has changed because of fire, disease, harvesting, or growth, the resulting reference labels in the changed areas will be incorrect. Harvests and fire are clearly visible on most satellite imagery, making it possible to detect the changes by looking at the imagery.†

† Using satellite imagery to correct the reference information collected from the photos seems a little convoluted since the photos are supposedly being used to assess the accuracy of a map produced from the imagery.

TABLE 6.5
Comparison of Sources of Reference Data

Source of Reference Data	Pros	Cons
Existing maps/data	Least expensive and quickest.	Can be out of date if change has occurred on the landscape. Must ensure that the minimum mapping unit and classification scheme used to label the existing data are identical to the scheme used to label the map.
New office interpreted data from remote sensing	Less expensive and time consuming than field collected data. Provides the same perspective as the remotely sensed data used to make the map (i.e., view from above).	Less accurate for vegetation species identification than field-collected data. Can be out of date if change has occurred on the landscape since the capture of the remotely sensed reference data.
New field collected data	More accurate for vegetation species identification.	Most expensive. Does not offer the same perspective as captured by the remotely sensed data (i.e., view from below versus view from above). Often difficult to establish because of terrain or access restrictions.

However, stand growth and partial defoliation from disease or pests are not as readily observable on the imagery, making the use of older photos especially problematic in the northwest and southeast portions of the U.S., where trees can grow through several size classes in a 10-year period.

In general, accuracy assessment reference data should be collected as close as possible to the date of the collection of the remotely sensed data used to make the map. However, trade-offs may need to be made between the timeliness of the data collection and the need to use the resulting map to stratify the accuracy assessment sample. In most, if not all, remote sensing mapping projects, it is necessary to go to the field to become familiar with what causes variation in the classes to be mapped, to calibrate the eye of the image analyst, and to collect information for training the classifier (i.e., supervised classification or object-oriented classification) or to aid in labeling the clusters (i.e., unsupervised classification). If the reference data for accuracy assessment can be collected independently, but simultaneously, during this trip,

TABLE 6.6
Error Matrix Showing Number of Samples in Each Crop Type

MAP DATA \ REFERENCE DATA	A	C	SG	CN	L	M	BG	CS	T	SU	O	CR	F	D	S	Total
A	157		8				3						3			171
C		1			1	1										3
SG	3		163		6						12	2	1			187
CN																0
L			4		3					1		1				9
M						5							1			6
BG	1						10									11
CS								69								69
T																0
SU																0
O			1		3						7					11
CR												2				2
F													224			224
D														11		11
S																0
Total	161	1	176	0	13	6	13	69	0	1	19	5	229	11	0	704

LEGEND		Producer's Accuracy	User's Accuracy
A = Alfalfa	A	98%	92%
C = Cotton	C	100%	33%
SG = Small Grains	SG	93%	87%
CN = Corn	CN	—	—
L = Lettuce	L	23%	33%
M = Melons	M	83%	83%
BG = Bermuda Grass	BG	77%	91%
CS = Citrus	CS	100%	100%
T = Tomatoes	T	—	—
SU = Sudan Grass	SU	0%	—
O = Other Veg.	O	37%	64%
CR = Crucifers	CR	40%	100%
F = Fallow	F	98%	100%
D = Dates	D	100%	100%
S = Safflowers	S	—	—

then a second trip to the field is eliminated, saving costs and ensuring that reference data collection is occurring close to the time the remotely sensed data are captured.

However, if accuracy assessment reference data are collected at the beginning of the project before the map is generated, then it is not possible to stratify the samples by map class since the map has yet to be created. It is also not possible to have a proportional to area allocation of the samples since the total area of each map class is still unknown.

An example helps illustrate these points. The USDI Bureau of Reclamation maps the crops of the lower Colorado River region four times a year using Landsat TM data. Farmland in this region is so productive and valuable that growers plant three to four crops per year and will plow under a crop to plant a new one in response to the futures market. With so much crop change, ground data collection and accuracy assessment must occur at the same time the imagery is collected. The Bureau sends a ground data collection crew to the field for two weeks surrounding the date of image acquisition. A random number generator is used to determine the fields to be visited and the same fields are visited during each field effort, regardless of the crops being grown. Therefore, the accuracy assessment sample is random, but not stratified by crop type. As Table 6.6 illustrates, some crops are oversampled and others are undersampled each time. The Bureau believes it is more important to ensure correct crop identification than it is to ensure that enough samples are collected in rarely occurring crop types.

Table 6.7 compares and contrasts the trade-offs required when deciding when to collect reference data.

TABLE 6.7
Pros and Cons of the Timing of Reference Data Collection

When Should the Reference Data Be Collected?	Pros	Cons
When the remotely sensed data are collected	Eliminates any chance of landscape change between the date of the acquisition and the date of the reference data. Cost effective as information needed to make and assess the map are collected at the same time.	Because the map has not been made, there is no way to ensure that enough samples will be taken in each map class.
After the map has been made	Because the map has been made, it is poosible to ensure that enough samples for each map class are collected.	Can be more expensive. Introduces the possibility of landscape change occurring between the date of the map and the date of the reference data collection.

ENSURING OBJECTIVITY AND CONSISTENCY

For accuracy assessment to be useful, map users must have faith that the assessment is a realistic representation of the map's accuracy. They must believe that the assessment is objective and the results are repeatable. Maintaining the following three conditions will ensure objectivity and consistency:

1. Accuracy reference data must always be kept independent of any training data.
2. Data must be collected consistently from sample site to sample site.
3. Quality control procedures must be developed and implemented for all steps of data collection.

DATA INDEPENDENCE

It was not uncommon for early accuracy assessments to use the same information to assess the accuracy of a map as was used to create the map. This unacceptable procedure obviously violates all assumptions of independence and biases the assessment in the favor of the map. Independence of the reference data can be assured in one of two ways. First, the reference and training data collection can be performed at a completely different time and/or by different people. Collecting information at different times is expensive and can introduce landscape change problems as discussed earlier. Using different people can also be expensive, as more personnel need to be completely trained in the details of the project.

The second method of ensuring independence involves collecting reference and training data simultaneously, and then using a random number generator to select and remove the accuracy assessment sites from the training data set. The accuracy assessment sites are not reviewed again until it is time to perform the assessment. In both cases, accuracy assessment reference data must be kept absolutely independent (i.e., separate) of any training/labeling data, and it must not be accessible during manual map editing.

DATA COLLECTION CONSISTENCY

Data collection consistency can be ensured through personnel training and the development of objective data collection procedures. Training should occur simultaneously for all personnel at the initiation of data collection. One to three days of intensive training is often necessary and must include reference collection on numerous example sites that represent the broad array of variation between and within map classes. Trainers must ensure that reference data collection personnel are (1) applying the classification scheme correctly and (2) accurately identifying characteristics of the landscape that are inherent in the classification scheme. For example, if a classification scheme depends on the identification of plant species, then all reference data personnel must be able to accurately identify species on the reference source data. The classification scheme must use the same minimum mapping unit as was applied to create the map.

In addition to personnel training, objective data collection procedures are key to consistent data collection. The more measurement (as opposed to estimation) involved in reference data collection, the more consistent and objective the collection. However, measurement increases the cost of accuracy assessment, so most assessments rely heavily on ocular estimation. If ocular estimates are to be used, then the variance inherent in estimation must be accepted as an unavoidable part of the assessment, and some method of assessing it must be included in the assessment. Several of these methods are discussed in Chapter 9.

An important mechanism for imposing objectivity is the use of a reference data collection form to force all data collection personnel through the same collection process. The complexity of the reference data collection form will depend on the level of the complexity of the classification scheme. The form should lead the collector through a rule-based process to a definitive reference label from the classification scheme. Forms also provide a means of performing a quality assessment/quality control check on the collection process. Figure 6.1 is an example data collection form for a relatively simple classification scheme. An important portion of this form is the dichotomous key that leads data collection personnel to the land cover class label solely on the basis of the classification scheme rules.

Reference data collection forms, regardless of their complexity, have some common components. These include (1) the name of the collector and the date of the collection, (2) locational information about the site, (3) some type of table or logical progression that represents what the collector is seeing, (4) a place to fill in the actual reference label from the classification scheme, and (5) a place to describe any anomalies, any variability, or interesting findings at the site.

These days it may be more common to have the form on your laptop computer, data logger, or PDA, rather than as a piece of paper. Regardless of how the form is represented, it is vital to make use of some form to ensure objectivity.

Quality Control

Quality control is necessary at every step of data collection. Each error in data collection can translate into an incorrect indication of map accuracy. Data collection errors result in both over- and underestimations of map accuracy.

The following text discusses some of the most common quality control problems in each step of accuracy assessment data collection. Because accuracy assessment requires collecting information from both the reference source data and the map, each step involves two possible occasions of error: during collection from the map and during collection from the reference source data.

1. *Location of the accuracy assessment sample site.* It is not uncommon for accuracy assessment personnel to collect information at the wrong location because inadequate procedures were used to locate the site on either the map or the reference data. As discussed in Chapter 3, any errors in the position of either the location of the accuracy assessment sample site on the reference data or on the map will result in a thematic error. Positional accuracy cannot be ignored when conducting a thematic accuracy assessment.

Site Name: ______________________ Site ID #: ________

Quad Name: ______________________ Satellite Scene ID #: ________

Latitude: __________ Longitude: __________ Elevation Range: ________

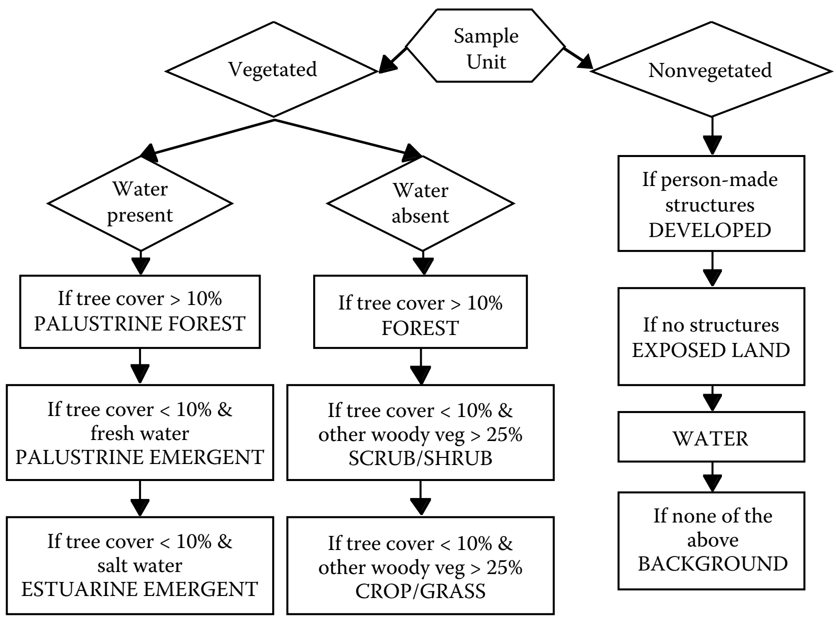
Method for Determining Position: ______________________

Comments on Position: ______________________

Comments on Weather: ______________________

Crew Names: ______________________ Date: __________

CATEGORIES		
Background	**Forest**	**Palustrine Forest**
Developed	**Scrub/Shrub**	**Estuarine Emergent**
Crop/Grass	**Exposed Land**	**Palustrine Emergent**
	Water	



Actual category as determined from flowchart: ______________________

Comments on anomalies, variability, or interesting finding: ______________________

FIGURE 6.1 Example of a reference data collection form for a simple classification scheme.

A common method for locating accuracy assessment sites on reference aerial imagery is to view the site on the map and then “eyeball’ the location onto the photos based on similar patterns of land cover and terrain in both the map and the reference data. In this situation, it is critical to provide the reference personnel with as many tools and as much information as possible to help them locate the site. GPS equipment has become critical to assuring location during fieldwork. Helpful information includes digitized flight line maps and other ancillary data such as stream, road, or ownership coverages. A laptop computer linked to a GPS and loaded with GIS software; the imagery, and ancillary data can reduce field time and increase reference location accuracy immeasurably.

Field location is always problematic, especially in wildlands (e.g., tundra, open water, wilderness areas, etc.) with few recognizable landscape characteristics. GPS is extremely helpful and should always be used to ensure the correct location of field sample sites.

2. *Sample unit delineation.* Both the reference site and the map accuracy assessment sites must represent exactly the same location. Thus, not only must the sites be properly located, they must also be delineated precisely and correctly transferred to a planimetric base. For example, if an existing map is used as the reference source, and the map was not registered correctly, then all accuracy assessment reference sites will not register with the new map being assessed, and a misalignment will occur when the reference site and the map site are compared. This was not an uncommon situation when aerial photography was used to create the existing map, and the transfer from the photo to the map was performed ocularly without the use of photogrammetric equipment.

   Another common error in accuracy assessment occurs when the reference and the map sites are in the same general location, but are of different sizes or shapes. For example, if map polygons constitute the sampling unit, and the reference data are aerial photographs, the selected polygons will often fall across two or more photos as depicted in Figure 6.2.

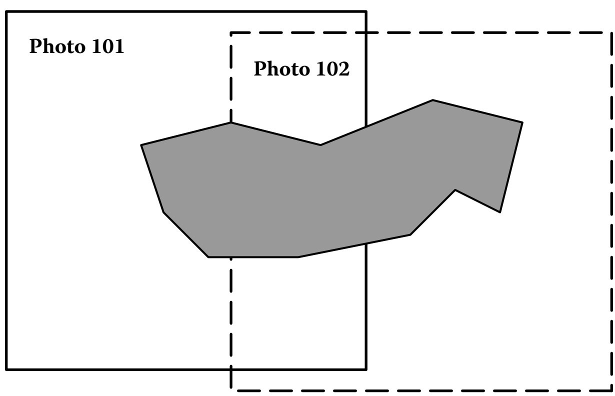


FIGURE 6.2 The shaded accuracy assessment site polygon falls across two aerial photographs.

In this case, the analyst must either collect reference data across all the photos (which can be time consuming), or the sample site must be reduced in size (though not below the mmu of the classification scheme) on both the map and the reference data so that it fits on one photo.

Also, given the need and ability to automate this entire process today, it is possible to use map coordinates to represent both the reference data samples and map locations and to automate the process of creating the error matrix such that the analyst is completely removed from the process. It happens entirely inside the computer, and only the final error matrix is revealed. While it is very appealing for this process to be automated, it is also very dangerous if left completely unverified because no visual comparison of the sample unit points is performed and none of the potential problems discovered. Historically, many of these issues would have been discovered during the manual comparison of reference sample sites to map sites during the error matrix generation. Therefore, even if using a fully automated method, it is recommended that some sample of these points be visually examined to check for errors.

3. *Data collection and data entry* are the most common sources of quality control problems in accuracy assessment. Data collection errors occur when measurements are done incorrectly, variables of the classification scheme are misidentified (e.g., species), or the classification scheme is misapplied. In addition, weak classification schemes will also create ambiguity in data collection. Unfortunately, the first indication of a weak classification system often occurs during accuracy assessment, when the map is already completed, and refinement of the classification scheme is not possible.

   Data collection errors are usually monitored by selecting a subsample of the accuracy assessment sites and having reference data on them collected simultaneously by two different personnel. Usually, the most experienced personnel are assigned to the subsample. When differences are detected, the sources of the differences are immediately identified. If data collection errors are the source of the differences, then personnel are either retrained or removed from reference data collection.

   When aerial imagery is used as the reference source data, it is critical that a ground assessment of the imagery interpretation be conducted. In addition, reference samples chosen from an existing map must also be assessed for accuracy.

   Data entry errors can be reduced by using digital data entry forms that restrict each field of the form to an allowable set of characters. Data can also be entered twice and the two data sets compared to identify differences and errors. Data entry errors also can occur when the site is digitized. Quality control must include a same-scale comparison of the digitized site to the source map.

Finally, although no reference data set may be completely accurate, it is important that the reference data have high accuracy or else the assessment is not a fair characterization of map accuracy. Therefore, it is critical that reference data collection be carefully considered in any accuracy assessment. Much work is yet to be done to determine the proper level of effort and collection techniques necessary to provide this vital information.