2 The History of Map Accuracy Assessment

HOW MAPS ARE MADE

Before the invention of aircraft, maps were created from human observations made on the earth's surface using survey equipment and the most basic, yet sophisticated, remote sensing devices: the human eyes and the analytical capabilities of the human brain. By the early sixteenth century Portuguese navigators were able to map the coast of Africa (see Figure 2.1) by relying on measurements taken at sea from astrolabes, quadrants, cross-staffs, and other early navigation tools. During their exploration of the American Northwest, Lewis and Clark were able to produce the remarkably detailed map in Figure 2.2. Indian pundits secretly mapped the Himalayas to high precision in the mid-1800s by pretending to be Buddhist pilgrims (Hopkirk, 1992), keeping count of their paces using holy beads and concealing compasses and other instruments in their clothing and walking sticks. However, all of these maps were not without error, and when observations on the earth's surface were unobtainable, map makers often interpolated between field observations with questionable results, as illustrated in one of the earliest, and obviously incorrect, maps of California, displayed in Figure 2.3.

One of the most notorious examples of reliance on an erroneous map created from field observations was the disastrous Donner party in 1846, which chose to follow Hasting's cutoff rather than the established Oregon–California trail during their migration from the Midwest. As a result, they added miles to their journey (the positional accuracy was in error), were forced to cross unexpected steep mountains and expanses of waterless desert (the thematic accuracy was in error), as shown in Figure 2.4, and ended up attempting to cross the Sierra Nevada mountains in late fall rather than during the summer. The group ended up stranded in 20 ft of snow just below the summit for the entire winter, and lost almost half of their party to starvation, hypothermia, and cannibalism (Stewart, 1960). Regardless of how the map is made, not knowing the accuracy of maps can have catastrophic results!

Today, most map makers use remote sensing† rather than field observations as the main source of spatial information. While field observations are still important, they are ancillary to the remote sensing data, providing information at sample locations instead of a total enumeration of the area to be mapped. Since the first aerial photograph was captured from a balloon in 1858, data collected using remote sensing has supplanted ground observations for map making. Satellites and aircraft offer humans

† Remote sensing is defined as the collection and interpretation of information about an object from a distant vantage point. Remote sensing systems involve the measurement of electromagnetic energy reflected or emitted from an object, and include instruments on aircraft and satellites.

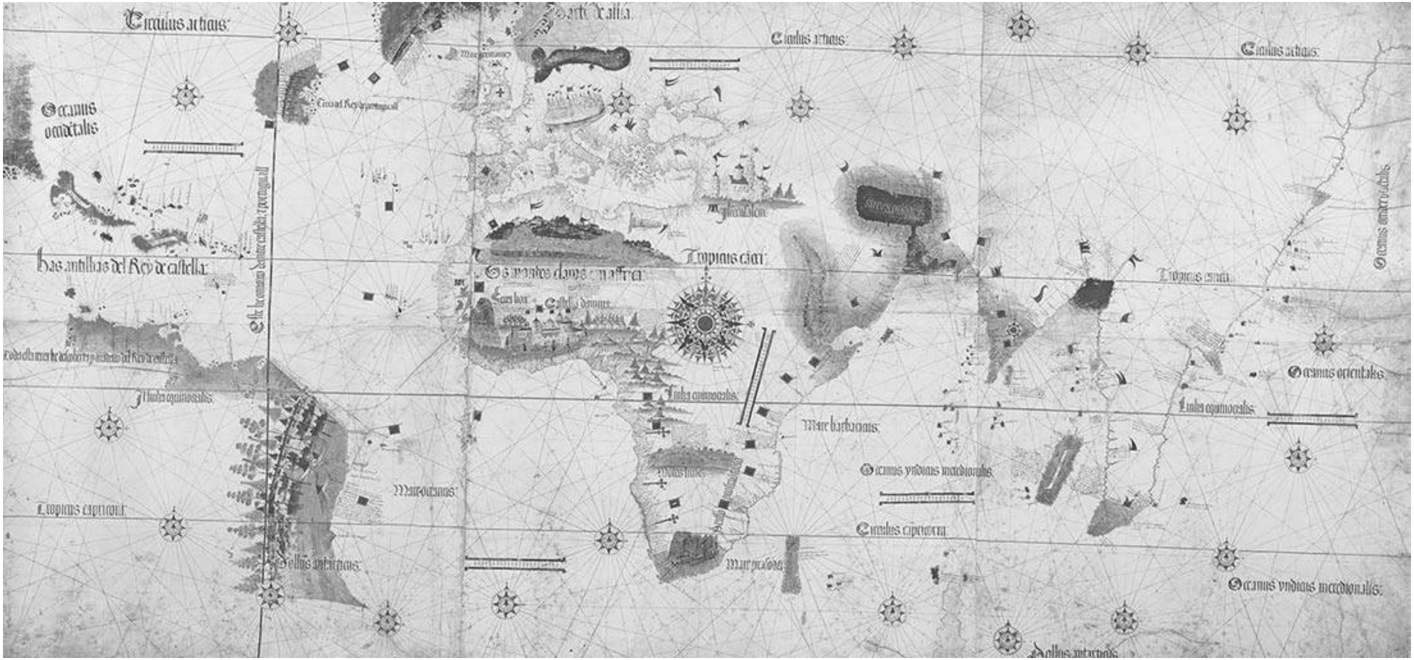


FIGURE 2.1 ***(Color version follows page 112)*** The Cantino World Map, which is a map of the known coastlines of the world, created by sixteenth century navigators.

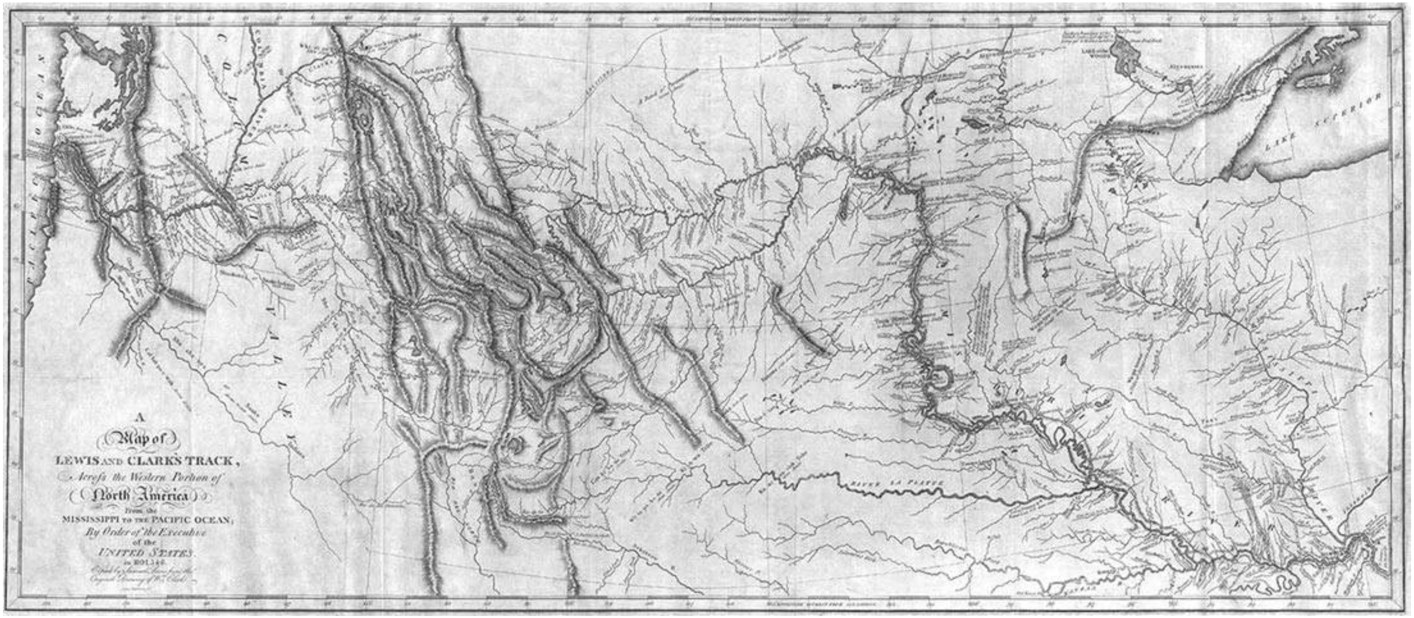


FIGURE 2.2 Map of the American Northwest created by Lewis and Clark. (From Lewis, Meriwether, William Clark, Nicholas Biddle, Paul Allen. 1814. *Maps of Lewis and Clark's Track across the Western Portion of North America*. Bradford and Inskeep, Philadelphia.)

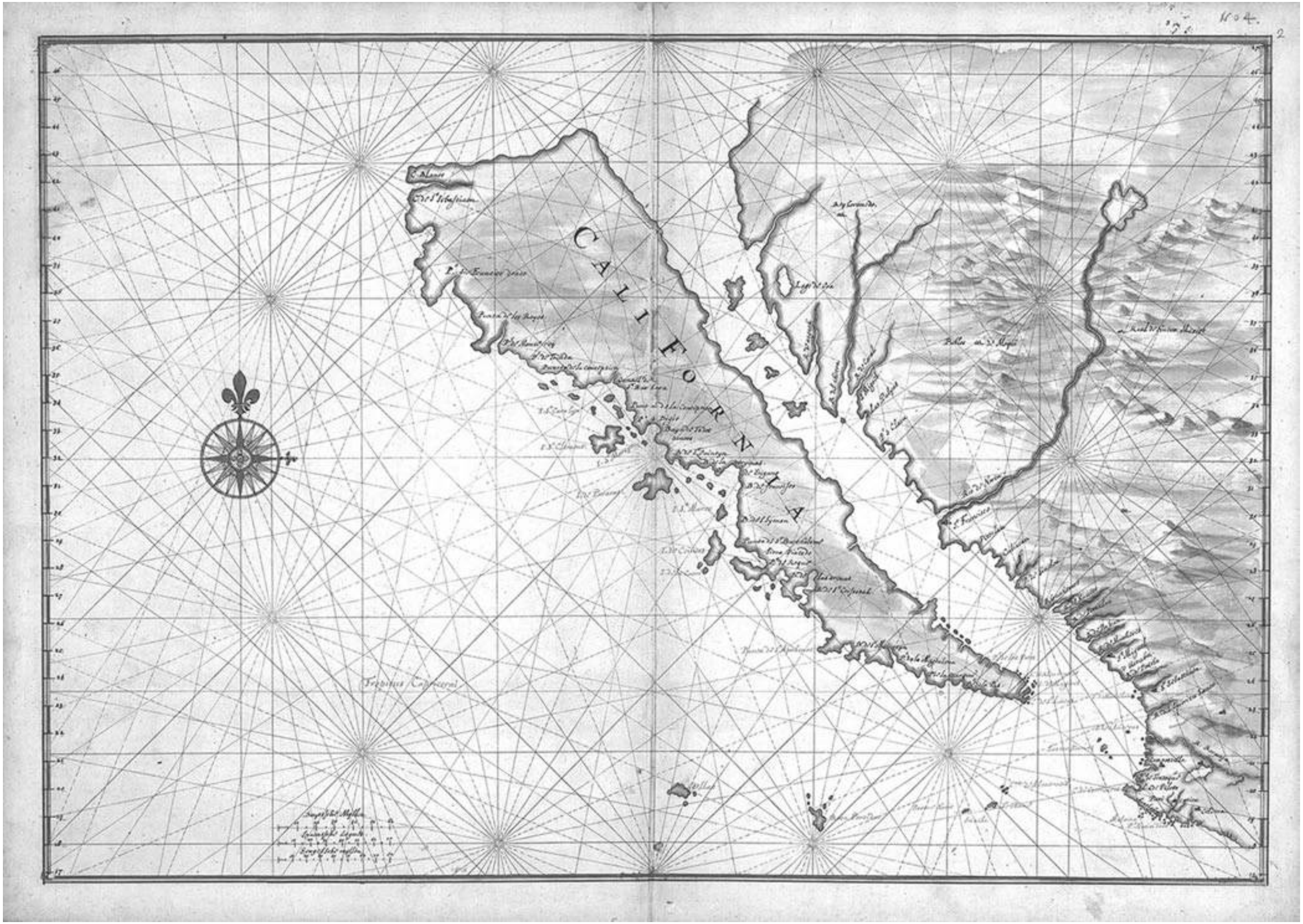


FIGURE 2.3 ***(Color version follows page 112)*** A seventeenth century map of California.

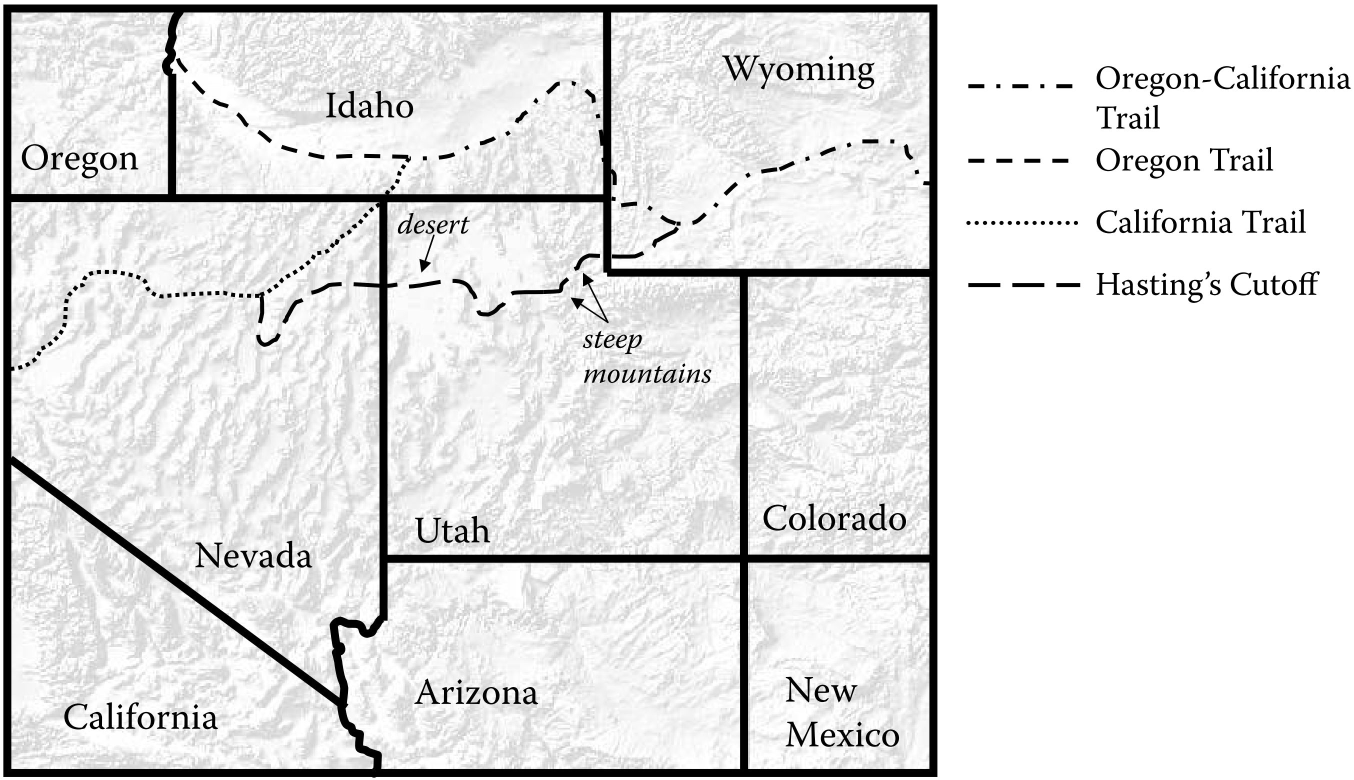


FIGURE 2.4 ***(Color version follows page 112)*** Hasting's Cutoff versus the safer California and Oregon trails used in 1846 by emigrants.

a view of their surroundings that humans cannot obtain on their own. Well before the first human went aloft in a balloon in 1783, humans had long been fantasizing about flight. Once humans invented successful flying machines, it was an easy step to put cameras in flying machines so that the pilot's perspective could be shared with those on the ground.

We use remotely sensed data to make maps because it:

- Is significantly less expensive and more efficient than creating maps from observations on the earth's surface,
- Offers a perspective from above (the "bird's-eye or synoptic view"), improving our understanding of spatial relationships, and
- Permits capturing imagery and information in electromagnetic wavelengths that humans cannot sense, such as the infrared portions of the electromagnetic spectrum.

Remotely sensed imagery is irresistible because it provides a view that can be readily understood, is inimitably useful, and yet is impossible to obtain without the use of technology. The innovation of air and space remote sensing has fundamentally changed the way we conduct war, manage inventory and resources, perform research, and respond to disasters.

Map making with remotely sensed data requires:

1. Precise linkage of the distances in the remote sensing imagery to distances on the ground so that spatial features can be accurately located, and
2. Understanding what causes variation in the features to be mapped and understanding how the remotely sensed data and ancillary information respond to those variations, so that the spatial features can be labeled.

Remotely sensed data provide an excellent basis for making maps because (1) remote sensing instruments and platforms are highly calibrated, and (2) a high correlation exists between variation in remotely sensed data and variation across the earth's surface.

However, there is never a complete one-to-one correlation between variation in remotely sensed data and variation on the earth's surface. Aircraft movement, topography, lens distortions, clouds, shadows, and a myriad of other factors can combine to weaken the relationship between the imagery and the earth's surface. Thus, much judgment, analysis, and interpretation are required to turn remotely sensed data into maps, and as a result, errors can occur during the many steps throughout any mapping project. As illustrated in Figure 2.5, the possible sources of error are multiple and compounding. Error can derive from the acquisition of imagery, to its rectification and classification, through its presentation as a map, and the application of the map in a decision-making process. Also, of course, error can also occur in the accuracy assessment itself. Accuracy assessment estimates, identifies, and characterizes the impact that arises from all of the sources of error.

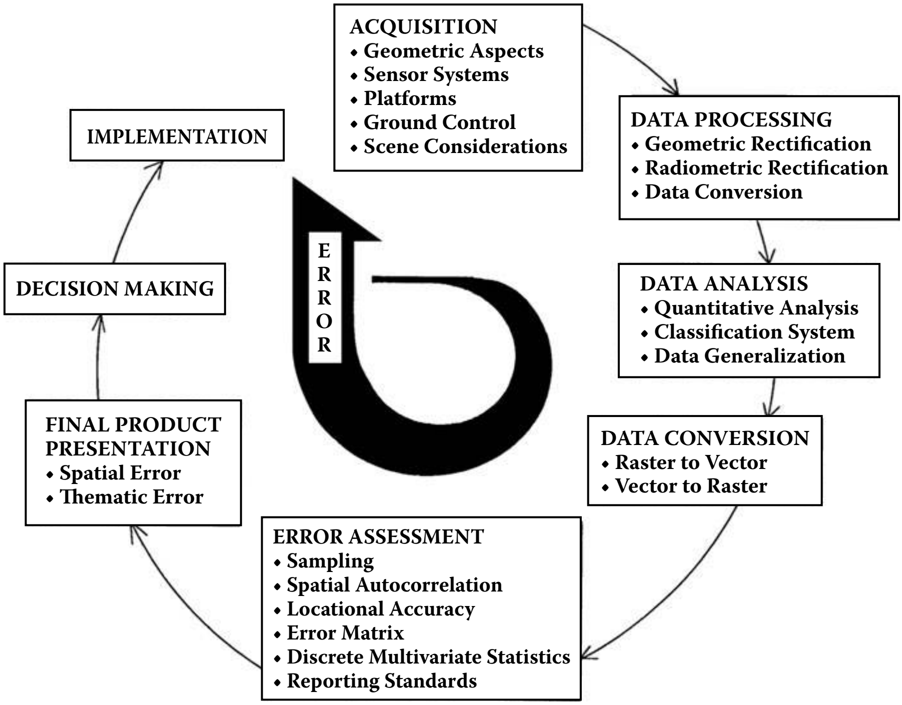


FIGURE 2.5 Sources of error in remotely sensed data. (Reproduced with permission from the American Society for Photogrammetry and Remote Sensing, from Lunetta, R., R. Congalton, L. Fenstermaker, J. Jensen, K. McGwire, and L. Tinney. 1981. Remote sensing and geographic information system data integration: error sources and research issues. *Photogrammetric Engineering and Remote Sensing*. 57(6): 677–687.)

HISTORY OF ACCURACY ASSESSMENT

The widespread acceptance and use of remotely sensed data have been and will continue to be dependent on the quality of the map information derived from it. As we learned in the previous section, the history of using remotely sensed data for mapping and monitoring the earth is a relatively short one. Aerial photography (analog or film-based remote sensing) has been used as an effective mapping tool only since the early 1900s. Digital image scanners and cameras on satellites and aircraft have an even shorter history beginning in only the mid-1970s. The following two sections briefly review the history of positional and thematic accuracy assessment of maps created from remotely sensed data.

Positional Accuracy Assessment

Photogrammetry, the science of determining the physical dimensions of objects from measurements on aerial photographs or imagery, was first implemented in 1849 using terrestrial photographs taken on the earth's surface (McGlone, 2004). Aerial photogrammetry, which utilizes images taken from aerial or satellite platforms, followed soon after the first photographs were taken from aircraft. Adoption of aerial photographs to create maps exploded with:

- The need to rebuild Europe following World War I,
- Development of roll film by George Eastman (founder of Kodak),
- Reduction of camera lens distortion,
- Improvements in camera bodies including increased sturdiness, permanently mounted lenses, techniques for holding the film flat, and inclusion of a mechanism for aligning the camera axis,
- Employment of fiducial marks for the definition of the image plane,
- Development of analytical photogrammetry equations, and
- Invention of the stereo plotter. (Ferris State University, 2007)

From the very first days of aerial photogrammetry, positional accuracy has been assessed by comparing the coordinates of sample points on a map against the coordinates of the same points derived from a ground survey or some other independent source deemed to be more accurate than the map. In the early twentieth century, mapping scientists focused on map production and attempted to characterize each different contributor to positional error. Now, positional error assessment is more user-focused, emphasizing the estimation of the total error, regardless of the source.

In 1937, the American Society of Photogrammetry (now the American Society for Photogrammetry and Remote Sensing or ASPRS) established a committee to draft spatial accuracy standards. Soon after, the U.S. Bureau of the Budget published the *United States National Map Accuracy Standards* (NMAS) in 1941. The current version of the National Map Accuracy Standards was published in 1947 (U.S. Bureau of the Budget, 1947) and is included in the following text:

1. "Horizontal accuracy. For maps on publication scales larger than 1:20,000, not more than 10% of the points tested shall be in error by more than 1/30th inch, measured on the publication scale; for maps on publication scales of 1:20,000 or smaller, 1/50th inch. These limits of accuracy shall apply to positions of well-defined points only. Well-defined points are those that are easily visible or recoverable on the ground, such as the following: monuments or markers, such as bench marks, property boundary monuments; intersections of roads and railroads; corners of large buildings or structures (or center points of small buildings). In general, what is well defined will also be determined by what is plottable on the scale of the map within 1/100th inch. Thus, while the intersection of two roads or property lines meeting at right angles would come within a sensible interpretation, identification of the intersection of such lines meeting at an acute angle would not be practicable within 1/100th inch. Similarly, features not identifiable upon the ground within close limits are not to be considered as test points within the limits quoted, even though their positions may be scaled closely upon the map. This class would cover timber lines and soil boundaries.
2. Vertical accuracy, as applied to contour maps on all publication scales, shall be such that not more than 10% of the elevations tested shall be in error by more than one-half the contour interval. In checking elevations

taken from the map, the apparent vertical error may be decreased by assuming a horizontal displacement within the permissible horizontal error for a map of that scale.

3. The accuracy of any map may be tested by comparing the positions of points whose locations or elevations are shown upon it with corresponding positions as determined by surveys of a higher accuracy. Tests shall be made by the producing agency, which shall also determine which of its maps are to be tested, and the extent of such testing.
4. Published maps meeting these accuracy requirements shall note this fact in their legends, as follows: "This map complies with National Map Accuracy Standards."
5. Published maps whose errors exceed those aforestated shall omit from their legends all mention of standard accuracy.
6. When a published map is a considerable enlargement of a map drawing (manuscript) or of a published map, that fact shall be stated in the legend. For example, 'This map is an enlargement of a 1:20,000-scale map drawing,' or 'This map is an enlargement of a 1:24,000-scale published map.'
7. To facilitate ready interchange and use of basic information for map construction among all federal map-making agencies, manuscript maps and published maps, wherever economically feasible and consistent with the use to which the map is to be put, shall conform to latitude and longitude boundaries, being 15 minutes of latitude and longitude, or 7.5 minutes, or 3.75 minutes in size."

Establishment of the standards was a critical step in implementing consistency in positional accuracy across the United States. However, NMAS focuses on errors measured at the map instead of ground scale, which became problematic over the years as maps migrated from paper to digital formats that can be printed at variable map scales. Additionally, the standards state the requirements for spatial accuracy, but only briefly discuss procedures for collecting samples to determine whether or not those standards have been met. Thus, while the accuracy percentage was standardized, the procedures for measuring accuracy were not.

In the 1960s a precursor to the present-day National GeoSpatial-Intellegence Agency (NGA), the Aeronautical Chart and Information Center, printed a report entitled *Principles of Error Theory and Cartographic Applications* (Greenwalt and Schultz, 1962, 1968) that meticulously laid the statistical foundation for estimating the distribution of positional map error from a sample of reference points. The basic concepts of the report derive from the probability theories developed in the 1800s to predict the probable distribution of artillery shells fired at a target. Relying on the root-mean-square error (RMSE)[†] as the parameter to be estimated in characterizing positional map accuracy, the report became, and has remained, the foundation for all other publications that stipulate the calculation of map error from a set of sample points

[†] RMSE is the square root of the average squared differences between accuracy assessment sample map and reference locations. The equations for calculating RMSE are presented in Chapter 3.

(ASPRS, 1990; DMA, 1991; FGDC, 1998; MPLMIC, 1999; Bolstad, 2005; Maune, 2007). However, unlike later publications, the report focused only on how to calculate error and did not address how the sample points should be chosen or measured.

In the late 1970s, the American Society for Photogrammetry and Remote Sensing's (ASPRS) Specifications and Standards Committee started a review of the 1947 standards with the goal of updating them to include standards for both hardcopy and digital maps. The result was the 1990 publication of *ASPRS Interim Accuracy Standards for Large-Scale Maps* (ASPRS, 1990), which stipulated that accuracy be reported at ground scale rather than map scale, thereby allowing the consideration of digital as well as hardcopy maps. The standards established the maximum RMSE (measured at ground distances) permissible for map scales from 1:60 to 1:20,000. It also cited Greenwalt and Schultz (1962, 1968) in establishing RMSE as the pivotal map accuracy parameter. Finally, it provided guidance on how accuracy sample points should be identified, measured, and distributed across the map and how these points should be collected.

Soon after the release of the ASPRS Standards, the Ad Hoc Map Accuracy Standards Working Group of the Subcommittee on Base Cartographic Data of the Federal Geographic Data Committee (FGDC) produced the *U.S. National Cartographic Standards for Spatial Accuracy* (NCSSA) (FGDC, 1998) to create positional accuracy standards for medium- and small-scale maps.

Following public review, the NCSSA was significantly modified so as to adopt positional accuracy assessment procedures in lieu of accuracy assessment standards. The result was the 1998 publication of FDGC *National Standard for Spatial Data Accuracy* (NSSDA) (FGDC, 1998), which relies heavily on the ASPRS standards and "implements a statistical and testing methodology for estimating the positional accuracy of points on maps and in digital geospatial data, with respect to georeferenced ground positions of higher accuracy." The standard explicitly does not establish threshold standards (as did the NMAS and ASPRS), but encourages map users to establish and publish their standards, which it was recognized would vary depending on the user's requirements.

Also relying on Greenwalt and Schultz (1962, 1968), the NSSDA specifies that positional accuracy be characterized using RMSE, requires that accuracy be reported in ground distance units at the "95% confidence level,"† and provides guidance on how samples are to be selected. NSSDA continues to be the accepted standard on positional accuracy assessment. It is often used in conjunction with the ASPRS large-scale map standards, with NSSDA providing standardized processes for assessing positional accuracy and the ASPRS (1990) standards setting the maximum errors allowable for different map scales.

More recently, three new guidelines have been established for assessing digital elevation data. All three call for the stratification of positional accuracy assessment samples into land cover types. Two of the guidelines also mandate that accuracy be reported at the "95th percentile error" in addition to the NSSDA statistic.

† Confusion exists in the mapping field between the terms "95% precision level" and "95% confidence level." Chapter 3 examines the difference in detail.

Thematic Accuracy Assessment

Unlike positional accuracy, there is no government standard for assessing and reporting thematic accuracy. This omission is partially due to the inherent complexity of thematic accuracy, but primarily to the fact that when maps were made from aerial photographs, thematic accuracy was generally assumed to be at acceptable levels. It was the development and use of digital remote sensing devices that had the most profound impact on thematic accuracy assessment of maps created from all remotely sensed data.

Spurr, in his excellent book *Aerial Photographs in Forestry* (1948), presents the early prevailing opinion about assessing the accuracy of photo interpretation. He states, "Once the map has been prepared from the photographs, it must be checked on the ground. If preliminary reconnaissance has been carried out, and a map prepared carefully from good quality photographs, ground checking may be confined to those stands whose classification could not be agreed upon in the office, and to those stands passed through en route to these doubtful stands." In other words, a qualitative visual check to see if the map looks right has traditionally been the recommended course of action for assessing photo interpretation.

However, in the 1950s some researchers saw the need for quantitative assessment of photo interpretation in order to promote their discipline as a science (Sammi, 1950; Katz 1952; Young, 1955; Colwell, 1955). In a panel discussion entitled "Reliability of Measured Values" held at the 18th Annual Meeting of the American Society of Photogrammetry, Mr. Amrom Katz (1952), the panel chair, made a very compelling plea for the use of statistics in photogrammetry. Other panel discussions were held, and talks were presented that culminated with a paper by Young and Stoeckler (1956). In this paper, these authors actually propose techniques for a quantitative evaluation of photo interpretation, including the use of an error matrix to compare field and photo classifications, and a discussion of the boundary error problem.

Unfortunately, these techniques never received widespread attention or acceptance. The *Manual of Photo Interpretation* published by the American Society of Photogrammetry (1960) does mention the need to train and test photo interpreters. However, it contains no description of the quantitative techniques proposed by those brave few in the 1950s.

There is no doubt that photo interpretation has become a time-honored skill, and the prevailing opinion for decades was that a quantitative thematic accuracy assessment was unnecessary. In speaking with some of the old-time photo interpreters, they remember those times when quantitative assessment was an issue. In fact, they mostly agree with the need to perform such an assessment and are usually the first to point out the limitations of photo interpretation. However, it was mostly agreed that the results of any photo interpretation grouped areas that were similar and that there was more variation between these polygons or vegetation types or forest stands than between them. Hence, with this goal achieved, no quantitative assessment was necessary. Therefore, the quantitative assessment of photo interpretation is typically not a requirement of any project. Rather the assumption that the map was correct or at least good enough prevailed. Then along came digital remote sensing, and some of these fundamental assumptions about photo interpretation needed to be further scrutinized and adapted.

As in the early days of aerial photography, the launch of Landsat 1 in 1972 resulted in a great burst of exuberant effort as researchers and scientists charged ahead trying to develop the field of digital remote sensing. In those early days, much progress was made and there was not much time to sit back and evaluate how they were doing. This “can do” mentality is common in many developing technologies. The GIS (geographic information system) community has experienced a similar development pattern. However, as a technology matures, more effort is dedicated to data quality and error/accuracy issues. By the early 1980s, some researchers began to consider and realistically evaluate where they were going and, to some extent, how they were doing with respect to the quality of maps derived from digital remotely sensed data.

The history of assessing the thematic accuracy of maps derived from remotely sensed data is relatively brief, beginning around 1975. Researchers, notably Hord and Brooner (1976), van Genderen and Lock (1977), and Ginevan (1979), proposed criteria and basic techniques for testing overall map accuracy. In the early 1980s, more in-depth studies were conducted and new techniques proposed (Aronoff, 1982, 1985; Rosenfield et al., 1982; Congalton and Mead, 1983; Congalton et al. 1983). Finally, from the late 1980s up to the present time, a great deal of work has been conducted on thematic accuracy assessment. More and more researchers, scientists, and users are discovering the need to adequately assess the thematic accuracy of maps created from remotely sensed data.

The history of digital accuracy assessment can be effectively divided into four parts or epochs. Initially, no real accuracy assessment was performed but rather an “it looks good” mentality prevailed. This approach is typical of a new, emerging technology in which everything is changing so quickly that there is not time to sit back and assess how good you are doing. Despite the maturing of the technology over the last 25 years, some remote sensing analysts and map users are still stuck in this mentality.

The second epoch is called the age of non-site-specific assessment. During this period, total acreages by map class were compared between reference estimates and the map without regard for location. It did not matter if you knew where it was; rather, just the total amounts were compared. While total acreage is useful, it is far more important to know where a specific land cover or vegetation type exists. Therefore, this second epoch was relatively short-lived and quickly led to the age of site-specific assessments.

In a site-specific assessment, actual locations on the ground are compared to the same location on the map and a measure of overall accuracy (i.e., percentage correct) presented. This method far exceeded the non-site-specific assessment, but lacked information about individual land cover/vegetation categories. Only overall map accuracy was assessed. Site-specific assessment techniques were the dominant method until the late 1980s.

Finally, the fourth and current age of accuracy assessment could be called the age of the error matrix. An error matrix compares information from reference sites to information on the map for a number of sample areas. The matrix is a square array of numbers set out in rows and columns which express the labels of samples assigned to a particular category in one classification relative to the labels of samples assigned

Reference Data

		D	C	AG	SB	row total
Classified Data	D	65	4	22	24	115
	C	6	81	5	8	100
	AG	0	11	85	19	115
	SB	4	7	3	90	104
	column total	75	103	115	141	434

Land Cover Categories

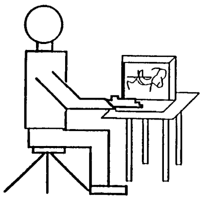
D = deciduous

C = conifer

AG = agriculture

SB = shrub

OVERALL ACCURACY = (65+81+85+90)/434 = 321/434 = 74%



PRODUCER'S ACCURACY

D = 65/75 = 87%
C = 81/103 = 79%
AG = 85/115 = 74%
SB = 90/141 = 64%

USER'S ACCURACY

D = 65/115 = 87%
C = 81/100 = 81%
AG = 85/115 = 74%
SB = 90/104 = 87%

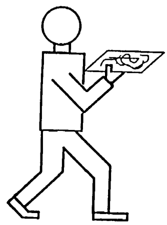


FIGURE 2.6 Example error matrix.

to a particular category in another classification (Figure 2.6). One of the classifications, usually the columns, is assumed to be correct and is termed the reference data. The rows are usually used to display the map labels or classified data generated from the remotely sensed image. Thus, two labels from each sample are compared to one another:

- Reference data labels: The class label or value of the accuracy assessment site, which is derived from data collected that is assumed to be correct; and
- Classified data or map labels: The class label or value of the accuracy assessment site derived from the map.

Error matrices are very effective representations of map accuracy because the individual accuracies of each map category are plainly described along with both the errors of inclusion (commission errors) and errors of exclusion (omission errors) present in the map. A commission error occurs when an area is included in an incorrect category. An omission error occurs when an area is excluded from the category to which it belongs. Every error on the map is an omission from the correct category and a commission to an incorrect category.

In addition to clearly showing errors of omission and commission, the error matrix can be used to compute not only overall accuracy, but also producer's accuracy, and user's accuracy, which were introduced to the remote sensing community by Story and Congalton (1986). Overall accuracy is simply the sum of the major diagonal (i.e., the correctly classified sample units) divided by the total number of sample units in the error matrix. This value is the most commonly reported accuracy assessment statistic and was part of the older, site-specific assessment. Producer's and user's accuracies are ways of representing individual category accuracies instead of just the overall classification accuracy (see Chapter 4 for more details on the error matrix).

Proper use of the error matrix includes correctly sampling the map and rigorously analyzing the matrix results. The techniques and considerations involved in the building and analyzing of an error matrix are the main themes of this book.