Caveat Emptor

The precise line work, complex content, multiple colors, and professional finish of published maps may convey to users an unwarranted degree of perfection concerning the information presented. The finished product may, without closer examination, appear to stand isolated from the process that produced it. Unless the confidence in both the underlying information and its transformation to the final map are made explicit, users are left with the implication that factual error is either absent or insignificant when neither is necessarily true.

In spite of necessary imprecision, maps serve many purposes well. As is the case with other types of information, what matters with maps is the relationship between the decisions people are trying to make and the quality of the information available to them. In daylight a pilot requires that a map be accurate only to within several miles, while at night in fog, its accuracy must be much higher. In this section, maps from the WRB atlas are used to illustrate types and amounts of error, how map accuracy is defined, how errors accumulate in spatial analyses, and how to use knowledge of map accuracy limitations in making decisions.

Sources of Uncertainty

Uncertainty in maps arises in identifying features and in positioning them to scale on a two-dimensional representation of a portion of the three-dimensional earth. Disagreement may exist in how to define categories of features and in which categories to place features, a process called **classification**. Even if no disagreements exist, errors inevitably arise in doing the work. **Scale** is the ratio of a feature's size on the map to its real world size - map size divided by actual size. **Accuracy** refers to the proximity of a reported value to a value accepted as actual for the phenomenon. For positioning mapped features, accuracy *error* is measured as the distance between a feature's map coordinates and its actual location on earth. **Precision** refers to the amount of detail used in reporting a measurement. A value of 8.315 is more precise than 8.3, but may or may not be more truthful. The more that classification and measurement errors increase, the larger is the "cloud" of potential values within which the true value lies, hence the more poorly we know where it is; this is the meaning of **uncertainty**.[161]

Map errors often arise as artifacts of the methods and tools used in recording, transforming, and representing features. In a map of soil types, for example, crisp lines divide one type from another when in fact different soil types merge within transitional interface zones. The conversion of data and maps to digital form is a transformation step that may introduce errors of **generalization**—the loss of detail, as well as decreased positional accuracy and systematic errors such as "terracing" in the representation of surface elevation arising from the characteristics of automated processes.

Some data such as satellite images originate in digital form. Their accuracy is limited by their **resolution**—the size of the smallest object the satellite instrument can detect, by the sensitivity of the instrument's sensors to specific wavelengths of light reflected from the earth, by atmospheric conditions at the time the scene was recorded, and by other factors such as topographic slope, the steepness of the earth's surface at each sampled location.

The phenomena with which maps are concerned are in constant change over periods ranging from hours to millennia. This means that a map may no longer represent accurately some of its reported facts by the time it is published, a characteristic called **temporal** accuracy. A completed map may comprise multiple **themes**—discrete phenomena or topics of interest, each of which may have come from a separate source map. When combined, the error rates of the individual maps can interact in surprising ways. If, according to some standard, each of two maps is 90% accurate, the result of their combination is 81% accurate, and if a third at the same accuracy is added, the result is 72% (0.9 x 0.9 x 0.9).

Measuring and Managing Error

If the accuracy of each of the source maps we are combining is known, then a guide exists for reducing the uncertainty created when they disagree - the more accurate map can be used to correct the less accurate. This requires, however, that methods of objectively characterizing the accuracy of maps must exist. Geographers have developed measures of the **planimetric**—two dimensional—position error (inversely the accuracy) of maps. Based on these measures, standards of accuracy have been established, and constantly modified, to be used in the production of maps. These standards are used to measure and report the accuracy of maps.

In 1941, the U.S. Office of the Budget defined positional accuracy standards for use by the U.S. Geological Survey (USGS) in terms of distances on a map for three specific map scales: 1:62,500 (0.51 mm), 1:24,000 (0.63 mm), and 1:12,000 (0.85 mm). In 1947 the standard was changed to 0.5 mm for all map scales coarser than 1:24,000.[162] For positional accuracy the USGS standard requires that 90% of mapped feature locations be less than or equal to these distances away from their actual locations. Translated to the ground, the 0.5 mm error distance at 1:24,000 scale is 39.4 ft (12 meters), often rounded to 40 feet in practice. Figure 214 depicts the meaning of the 90%-of-points-within-40 ft standard. Most mapped locations will be closer than this limit to actual position, while 10% will be farther away and in any direction with equal likelihood.

A similar rule applies to classification accuracy: in 90% of instances tested, the correct category *from among those defined* must be chosen. In order to determine if a map meets such standards another source in which greater confidence exists must also be available. A statistically significant number of randomly chosen features is selected. Each is checked for positional and classificatory accuracy against either the original sources or field notes, all of which have their own errors, hopefully smaller in magnitude and fewer in number. For recently produced digital maps, the Federal Geographic Data Committee (FGDC) has defined standards of information describing the production processes, sources, and technical features of maps including their accuracy levels. Called **metadata**, these descriptive data, when available, are attached to and accompany each digital map.

Ultimately, all positional references depend on knowing where on the ground invisible lines of **latitude**, position North-South, and **longitude**, position East-West lie. The U.S. Department of Commerce National Geodetic Survey (NGS), an agency that continues work initiated by President Thomas Jefferson, maintains the National Spatial Reference System (NSRS) for this purpose. For centuries, determining position on earth has been based on reference to celestial objects and accurate time measurement. In addition to other sources, NGS now uses the Global Positioning System (GPS) of satellites managed by the US Navy, the organization historically responsible for national time keeping. Figure 213 is a diagram of a set of GPS-derived **control points**, measured at the intersections of stream and transportation networks, that were developed as part of the PNW-ERC project to improve the accuracy of the spatial data in this atlas. The 90% circular error distance for these points is 9.8 ft (3 m).

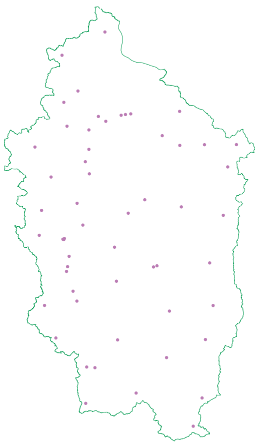


Figure 213. *GPS points*

Figure 214a is an example of uncertainty arising from combining four source maps, derived from data used in this atlas. It shows a stream in blue being crossed by two different converging lines indicating the same road. The small black star marks the location of the bridge associated with the stream crossing. The distance between the star and the intersection of the stream and the road line is 131 ft (40 m). But the distance that matters is that between each feature's location on the map, and where it is on the ground. Although metadata were available for the streams map, none were provided for the roads or the bridge. While we may have higher confidence in the positional accuracy of the streams, we still do not know where the road actually crosses the stream and therefore where the bridge should be.

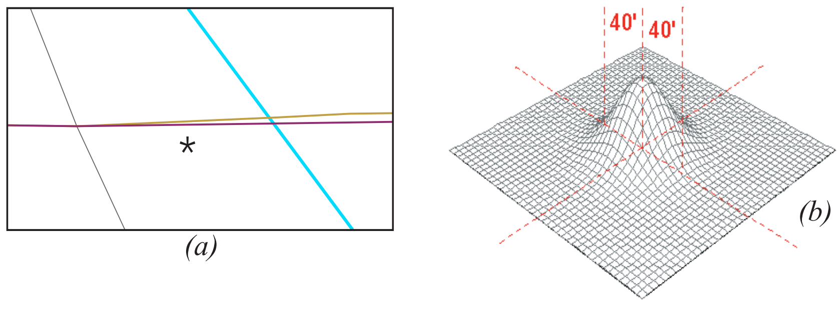


Figure 214. *Illustration of uncertainty caused by combination of data from different sources. a) stream crossing; (b) circular error probability.*

Figure 215 shows how convergence can help to reduce uncertainty in maps. Location A is a GPS control point measured at the west edge of the Willamette River bank just south of Harrisburg underneath the northern railroad bridge (black lines). Point B shows the location on the map which should correspond to the GPS control point. The distance between point A and point B is 525 ft (160 m). The control point (A) does, however, correspond to the map of the bank edge of the river, shown in light blue. Since the bank edge also agrees in position with the surface elevation map, Figure 215b, and the center line river map agrees with both of those, we gain confidence in the accuracy of these sources at this location. Further examination of the railroad map confirmed its low positional accuracy. The spatial convergence of the other themes tell us which map should be adjusted, the railroad, and gives us the location to which to adjust it.

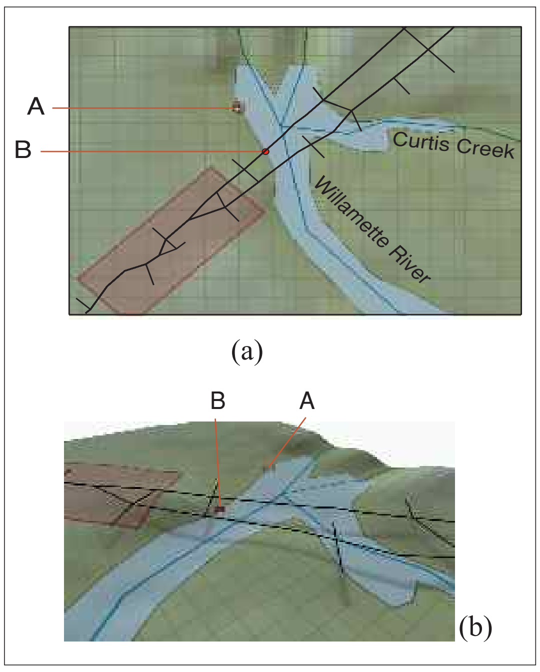


Figure 215. *Feature alignment*

Loss of detail through generalization appears in spatial data in several forms. Figure 216a is an aerial photograph of the confluence of the Santiam and Willamette Rivers, while Figure 216b shows how the same area might appear to one of the most commonly used satellite instruments, the Landsat Thematic Mapper. It "sees" the ground in cells 82 ft (25 m) on a side. Ultimately a single classification is chosen for each cell; no subdivision within is possible. If the actual area on the ground is 45% bare and 55% grass, the entire cell is classified as grass; this is one type of generalization. Figure 217 shows another type, in which information is lost through the act of tracing a complex shape and simplifying some of its geometric detail in the process of converting a finer grain, higher resolution depiction of vegetation boundaries to a coarser grain, lower resolution depiction.

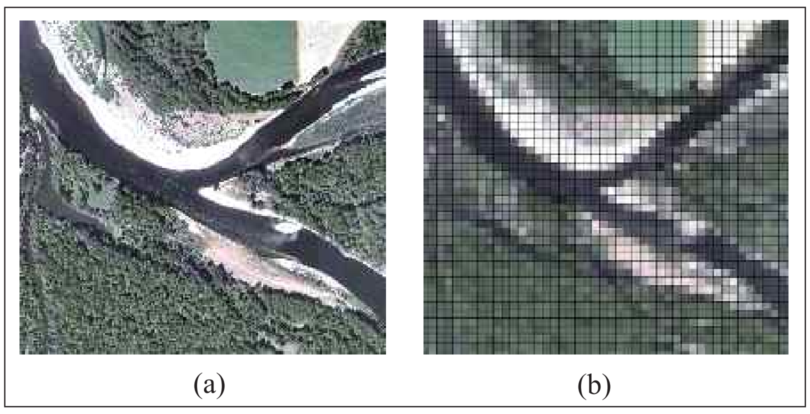
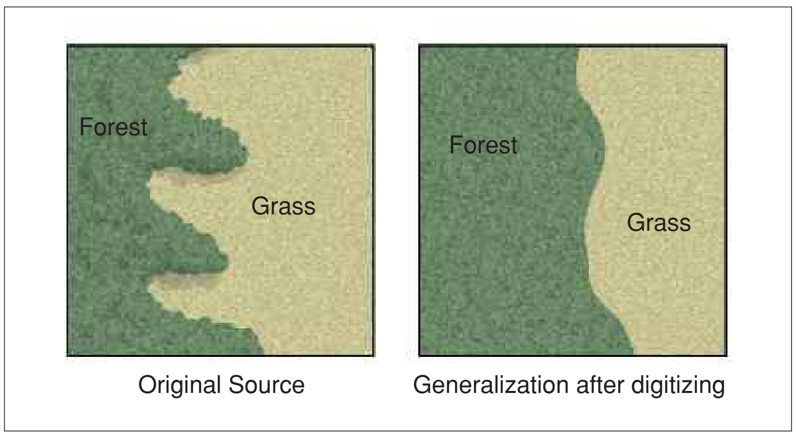

(a) (b)

Figure 216. *Illustration of loss of detail through generalization of spatial data. (a) photograph, (b) simulated satellite imagery.*



Original Source Generalization after digitizing

Figure 217. *Loss of line detail*

The use of reporting units introduces another kind of generalization, as shown in Figure 218. Figure 218a shows a single census tract in gray for which the 1970 federal census reported a single numeric value for the human population total. Of course, people are not evenly distributed within this area, as depicted in Figure 218b, which shows peaks of higher density as high points on a shaded surface within the tract as of the 1990 census. In addition to generalization error, these **choropleth** (Gk: choros = place, pleth = value) reporting units introduce another kind of error when phenomena unrelated to the reason the units were defined are reported via these spatial units. Rates of disease incidence, for example, may actually relate to moisture or vegetation differences, but be mapped by voting precincts because the data were gathered at those locations, thereby obscuring important, potentially causal, correlations. Where available, spatial data uncertainty estimates are included in PNW-ERC map metadata. These data are available on the PNW-ERC website at http://www.oregonstate.edu/dept/pnw-erc.

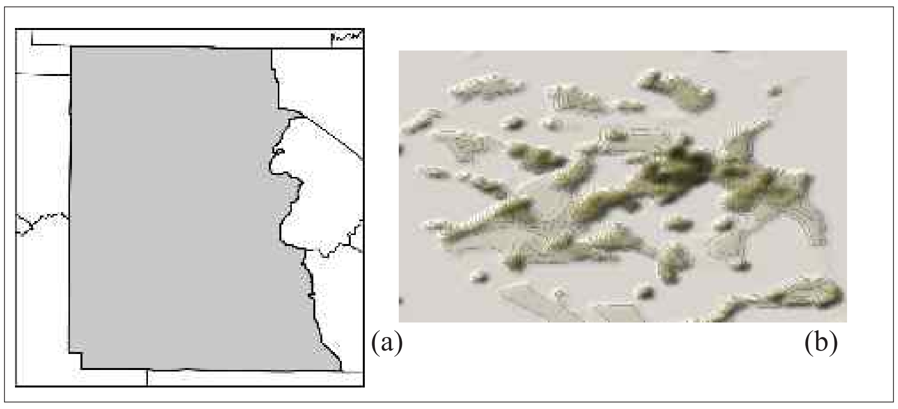


Figure 218. *Reporting units showing: a) single 1970 census tract becoming b) multiple census blocks in 1990.*

Making Decisions with Imperfect Maps

What something is and where it is can turn out to be tightly entwined. Figure 219 shows federal census reporting units in Marion County. Called Minor Civil Divisions (MCD) in 1930 (red), Census County Divisions (CCD) in 1970 (green), and tracts in 1990 (blue), their differing boundaries reveal both positional error in map making and changing definitions of the spatial reporting units themselves. In 1930, the MCDs were county subdivisions based on voting precinct boundaries and were the smallest spatial unit for which human population totals were reported. In 1970, CCD boundaries were defined by greater reliance on natural features, but also included many of the 1930 precinct boundaries as well, in this case combining many 1930 precincts into a unit whose outer boundary intends in many areas to denote the same place on earth. The 1970 census reported data for these much larger CCDs but also separately for towns and cities within them. The 1990 tracts intended to map the same entities as the 1970 CCDs, but the 1990 census also reported population at a much smaller subdivision of tracts called blocks. All three boundaries show the fluctuating degree of misalignment typical of positional errors, as well as definitional changes.

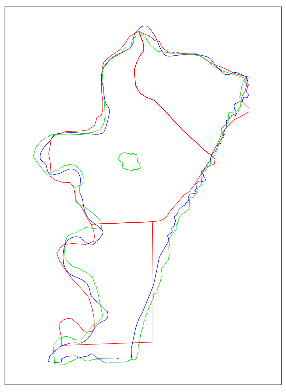


Figure 219. *Census units*

In spite of the complexity of these differences, the core area for which all three agree amounts to 83% of the total. The distinction between core and margin can be useful in establishing bounds on the uncertainty inevitably created by errors. As illustrated by Figure 214b, confidence is justifiably higher near the central value.

Together, the availability of accuracy information about maps, the use of more accurate sources to improve less accurate ones, and using the convergence of multiple sources to establish a consensus, increase the quality of information available to decision makers. With awareness of these factors, the questions asked of maps can then be adjusted to match the available information. It is this two step process that ultimately defines what is meant by acceptable accuracy. Using the federal census data, we cannot determine how human occupancy of the flood zone of the Willamette River has changed between 1930 and 1990, an unfortunate loss in the present context. We can, however, quantify and map trends in human population density in the basin over time, information of vital importance in the development and application of land use policy. The actual usefulness of maps increases when their limitations are known.