CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

We currently live in what is often termed *the information age*. Aided by new and emerging technologies, data are being collected at unprecedented rates in all walks of life. For example, in the field of surveying, total station instruments, global navigation satellite systems (GNSSs) equipment, digital metric cameras, laser-scanning systems, LiDAR, mobile mapping systems, and satellite imaging systems are only some of the new instruments that are now available for rapid generation of vast quantities of observational data.

Geographic information systems (GISs) have evolved concurrently with the development of these new data acquisition instruments. GISs are now used extensively for management, planning, and design. They are being applied worldwide at all levels of government, in business and industry, by public utilities, and in private engineering and surveying offices. Implementation of a GIS depends on large quantities of data from a variety of sources, many of them consisting of observations made with the new instruments such as those noted above and others collected by instruments no longer used in practice.

However, before data can be utilized whether for surveying and mapping projects, for engineering design, or for use in a geographic information system, they must be processed. One of the most important aspects of this is to account for the fact that *no measurements are exact. That is, they always contain errors*.

The steps involved in accounting for the existence of errors in observations consist of (1) performing statistical analyses of the observations to assess the magnitudes of their errors, and study their distributions to determine whether they are within acceptable tolerances, and if the observations are acceptable, (2) adjusting them so they conform to exact geometric conditions or other required constraints. Procedures for performing these two steps in processing measured data are principal subjects of this text.

1.2 DIRECT AND INDIRECT MEASUREMENTS

Measurements are defined as observations made to determine unknown quantities. They may be classified as either direct or indirect. *Direct measurements* are made by applying an instrument directly to the unknown quantity and observing its value, usually by reading it directly from graduated scales on the device. Determining the distance between two points by making a direct measurement using a graduated tape, or measuring an angle by making a direct observation from the graduated circle of a total station instrument are examples of direct measurements.

Indirect measurements are obtained when it is not possible or practical to make direct measurements. In such cases the quantity desired is determined from its mathematical relationship to direct measurements. For example, surveyors may observe angles and lengths of lines between points directly and use these observations to compute station coordinates. From these coordinate values, other distances and angles that were not observed directly may be derived indirectly by computation. During this procedure, the errors that were present in the original direct observations are *propagated* (distributed) by the computational process into the indirect values. Thus, the indirect measurements (computed station coordinates, distances, directions, and angles) contain errors that are functions of the original errors. This distribution of errors is known as *error propagation*. The analysis of how errors propagate is also a principal topic of this text.

1.3 MEASUREMENT ERROR SOURCES

It can be stated unconditionally that (1) no measurement is exact, (2) every measurement contains errors, (3) the true value of a measurement is never known, and thus (4) the exact size of the error present is always unknown. These facts can be illustrated by the following. If an angle is measured with a scale divided into degrees, its value can be read only to perhaps the nearest tenth of a degree. However if a better scale graduated in minutes were available and read under magnification, the same angle might be estimated to tenths of a minute. With a scale graduated in seconds, a reading to the nearest tenth of a second might be possible. From the foregoing, it should be clear that no matter how well the observation is taken, a better one may be possible. Obviously in this example, observational accuracy depends on the division size of the scale. But accuracy depends on many other factors, including the overall reliability and refinement of the equipment used, environmental conditions that exist when the observations are taken, and human limitations (e.g., the ability to estimate fractions of a scale division). As better equipment is developed, environmental conditions improve, and observer ability increases, observations will approach their true values more closely, but they can never be exact.

By definition, an *error* is the difference between a measured value for any quantity and its true value, or

$$\varepsilon = y - \mu \tag{1.1}$$

where ε is the error in an observation, y the measured value, and μ its true value.

As discussed above, errors stem from three sources, which are classified as instrumental, natural, and personal. These are described as follows:

1. *Instrumental errors*. These errors are caused by imperfections in instrument construction or adjustment. For example, the divisions on a theodolite or total station instrument may not be spaced uniformly. These error sources are present whether the equipment is read manually or digitally.
2. *Natural errors*. These errors are caused by changing conditions in the surrounding environment. These include variations in atmospheric pressure, temperature, wind, gravitational fields, and magnetic fields.
3. *Personal errors*. These errors arise due to limitations in human senses, such as the ability to read a micrometer or to center a level bubble. The sizes of these errors are affected by personal ability to see and by manual dexterity. These factors may be influenced further by temperature, insects, and other physical conditions that cause humans to behave in a less precise manner than they would under ideal conditions.

1.4 DEFINITIONS

From the discussion thus far it can be stated with absolute certainty that all measured values contain errors, whether due to lack of refinement in readings, instabilities in environmental conditions, instrumental imperfections, or human limitations. Some of these errors result from physical conditions that cause them to occur in a systematic way, whereas others occur with apparent randomness. Accordingly, errors are classified as either systematic or random. But before defining systematic and random errors, it is helpful to define mistakes. These three terms are defined as follows:

1. *Mistakes*. These are caused by confusion or by an observer's carelessness. They are not classified as errors and must be removed from any set of observations. Examples of mistakes include (*a*) forgetting to set

the proper parts-per-million (ppm) correction on an EDM instrument, or failure to read the correct air temperature, (*b*) mistakes in reading graduated scales, and (*c*) mistakes in recording (i.e., writing down 27.55 for 25.75). Mistakes are also known as *blunders* or *gross errors*.

2. *Systematic errors*. These errors follow some physical law, and thus, these errors can be predicted. Some systematic errors are removed by following correct observational procedures (e.g., balancing backsight and foresight distances in differential leveling to compensate for earth curvature and refraction). Others are removed by deriving corrections based on the physical conditions that were responsible for their creation (e.g., applying a computed correction for earth curvature and refraction on a trigonometric leveling observation). Additional examples of systematic errors are (*a*) temperature not being standard while taping, (*b*) an indexing error of the vertical circle of a total station instrument, and (*c*) use of a level rod that is not standard length. Corrections for systematic errors can be computed and applied to observations to eliminate their effects.
3. *Random errors*. These are the errors that remain after all mistakes and systematic errors have been removed from the observed values. In general, they are the result of human and instrument imperfections. They are generally small and are as likely to be negative as positive. They usually do not follow any physical law and therefore must be dealt with according to the mathematical laws of probability. Examples of random errors are (*a*) imperfect centering over a point during distance measurement with an EDM instrument, (*b*) bubble not centered at the instant a level rod is read, and (*c*) small errors in reading graduated scales. It is impossible to avoid random errors in measurements entirely. Although they are often called accidental errors, their occurrence should not be considered an accident.

1.5 PRECISION VERSUS ACCURACY

Due to errors, repeated measurement of the same quantity will often yield different values. A *discrepancy* is defined as the algebraic difference between two observations of the same quantity. When small discrepancies exist between repeated observations, it is generally believed that only small errors exist. Thus, the tendency is to give higher credibility to such data and to call the observations *precise*. However, precise values are not necessarily accurate values. To help understand the difference between precision and accuracy, the following definitions are given:

1. *Precision:* Precision is the degree of consistency between observations and is based on the sizes of the discrepancies in a data set. The degree of precision attainable is dependent on the stability of the environment

during the time of measurement, the quality of the equipment used to make the observations, and the observer's skill with the equipment and observational procedures.

2. *Accuracy:* Accuracy is the measure of the absolute nearness of an observed quantity to its true value. Since the true value of a quantity can never be determined, accuracy is always an unknown.

The difference between precision and accuracy can be demonstrated using distance observations. Assume that the distance between two points is paced, taped, and measured electronically and that each procedure is repeated five times. The resulting observations are:

Observation	Pacing (p)	Taping (t)	EDM (e)
1	571	567.17	567.133
2	563	567.08	567.124
3	566	567.12	567.129
4	588	567.38	567.165
5	557	567.01	567.114

The arithmetic means for these sets of data are 569, 567.15, and 567.133, respectively. A line plot illustrating relative values of the electronically measured distances denoted by e, and the taped distances, denoted by t, is shown in Figure 1.1. Notice that although the means of the EDM data and of the taped observations are relatively close, the EDM set has smaller discrepancies. This indicates that the EDM instrument produced a higher precision. However, this higher precision does not necessarily prove that the mean of the electronically observed data is implicitly more accurate than the mean of the taped values. In fact, the opposite may be true if, for example, the reflector constant was entered incorrectly causing a large systematic error to be present in all the electronically observed distances. Because of the larger discrepancies, it is unlikely that the mean of the paced distances is as accurate as either of the other two values. But its mean could be more accurate if large systematic errors were present in both the taped and electronically measured distances.

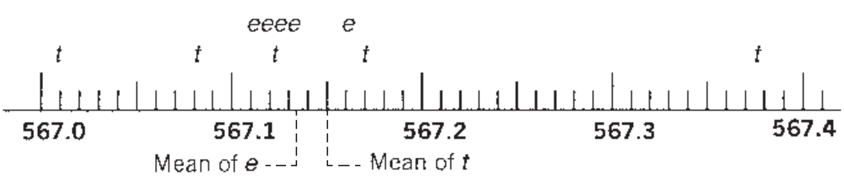

FIGURE 1.1 Line plot of distance quantities.

Another illustration explaining differences between precision and accuracy involves target shooting, depicted in Figure 1.2. As shown, four situations can occur. If accuracy is considered as closeness of shots to the center of a target at which a marksman shoots and precision as the closeness of the shots to each other then (1) the data may be both precise and accurate, as shown in Figure 1.2(*a*); (2) the data may produce an accurate mean but not be precise, as shown in Figure 1.2(*b*); (3) the data may be precise but not accurate, as shown in Figure 1.2(*c*); or (4) the data may be neither precise nor accurate as shown in Figure 1.2(*d*).

Figure 1.2(*a*) is the desired result when observing quantities. The other cases can be attributed to the following situations. The results shown in Figure 1.2(*b*) occur when there is little refinement in the observational process. Someone skilled at pacing may achieve these results. Figure 1.2(*c*) generally occurs when systematic errors are present in the observational process. This can occur, for example, in taping if corrections are not made for tape length and temperature, or with electronic distance measurements when using the wrong combined instrument-reflector constant. Figure 1.2(*d*) shows results obtained when the observations are not corrected for systematic errors and are taken carelessly by the observer (or the observer is unskilled at the particular measurement procedure).

In general, when making measurements, data such as those shown in Figure 1.2(*b*) and 1.2(*d*) are undesirable. Rather, results similar to those shown in Figure 1.2(*a*) are preferred. However, in making measurements the results of Figure 1.2(*c*) can be just as acceptable if proper steps are taken to correct

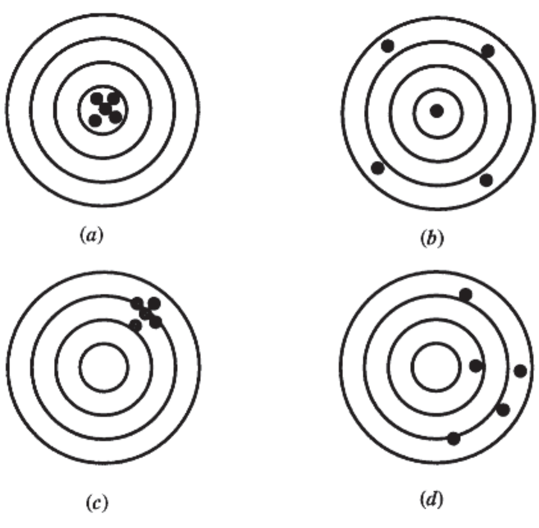

FIGURE 1.2 Examples of precision versus accuracy.

for the presence of the systematic errors. (This correction would be equivalent to the marksman realigning the sights after taking the shots.) To make these corrections, (1) the specific types of systematic errors that have occurred in the observations must be known, and (2) the procedures used in correcting them must be understood.

1.6 REDUNDANT OBSERVATIONS IN SURVEYING AND THEIR ADJUSTMENT

As noted earlier, errors exist in all observations. In surveying, the presence of errors is obvious in many situations where the observations must meet certain conditions. In level loops that begin and close on the same bench mark, for example, the elevation difference for the loop must equal zero. However, in practice this is hardly ever the case due to the presence of random errors. (For this discussion it is assumed that all mistakes have been eliminated from the observations and appropriate corrections have been applied to remove all systematic errors.) Other conditions that disclose errors in surveying observations are that (1) the three measured angles in a plane triangle must total 180°, (2) the sum of the angles measured around the horizon at any point must equal 360°, and (3) the algebraic sum of the latitudes (and departures) must equal zero for closed traverses that begin and end on the same station. Many other conditions could be cited; however, in any of them, the observations rarely, if ever, meet the required conditions, due to the presence of random errors.

The examples above not only demonstrate that errors are present in surveying observations but also illustrate the importance of *redundant observations*; those measurements made that are in excess of the minimum number needed to determine the unknowns. Two measurements of the length of a line, for example, yield one redundant observation. The first observation would be sufficient to determine the unknown length, and the second is redundant. However, this second observation is very valuable. First, by examining the discrepancy between the two values, an assessment of the size of the error in the observations can be made. If a large discrepancy exists, a blunder or large error is likely to have occurred. In that case, observations of the line would be repeated until two values having an acceptably small discrepancy were obtained. Second, the redundant observation permits an adjustment to be made to obtain a final value for the unknown line length, and that final adjusted value will be more precise statistically than either of the individual observations. In this case, if the two observations were of equal precision, the adjusted value would be the simple mean.

Each of the specific conditions cited in the first paragraph of this section involve one redundant observation. For example, there is one redundant observation when the three angles of a plane triangle are observed. This is true because with two observed angles, say A and B, the third could be computed as $C = 180° - A - B$, and thus, observation of C is unnecessary. However measuring angle C enables an assessment of the errors in the angles to be made, and it also makes an adjustment possible to obtain final angles with statistically improved precision. Assuming the angles were of equal precision, the adjustment would enforce the 180° sum for the three angles by distributing the total discrepancy in equal parts to each angle.

Although the examples cited here are indeed simple, they help to define redundant observations, and to illustrate their importance. In large surveying networks, the number of redundant observations can become extremely large, and the adjustment process is somewhat more involved than it was for the simple examples given here.

Prudent surveyors always make redundant observations in their work, for the two important reasons indicated above: (1) to enable assessing errors and making decisions regarding acceptance or rejection of the observations; and (2) to make possible an adjustment whereby final values with higher precisions are determined for the unknowns.

1.7 ADVANTAGES OF LEAST SQUARES ADJUSTMENT

As indicated previously, in surveying it is recommended that redundant observations always be made and that adjustments of the observations always be performed. These adjustments account for the presence of errors in the observations and increase the precision of the final values computed for the unknowns. When an adjustment is completed, all observations are corrected so that they are consistent throughout the survey network (i.e., the same values for the unknowns are determined no matter which corrected observation(s) are used to compute them).

Many methods have been derived for making adjustments in surveying; however, the method of least squares should be used because it has significant advantages over all other rule-of-thumb procedures. The advantages of least squares over other methods can be summarized with the following four general statements: (1) it is the most rigorous of adjustments; (2) it can be applied with greater ease than other adjustments; (3) it enables rigorous post-adjustment analyses to be made; and (4) it can be used to perform presurvey planning. These advantages are discussed further below.

The least squares method is rigorously based on the theory of mathematical probability, whereas in general, the other methods do not have this rigorous base. As described later in the book, in a least squares adjustment, the following condition of mathematical probability is enforced: *The sum of the squares of the errors times their respective weights are minimized*. By enforcing this condition in any adjustment, the set of errors that is computed has the highest probability of occurrence. Another aspect of least squares adjustment that adds to its rigor is that it permits all observations, regardless of their number or type, to be entered into the adjustment and used simultaneously in the computations. Thus, an adjustment can combine distances, horizontal angles, azimuths, zenith or vertical angles, height differences, coordinates, and even GNSS observations. One important additional asset of least squares method is that it enables *relative weights* to be applied to the observations in accordance with their expected relative reliabilities. These reliabilities are based on expected precisions. Thus, if distances were observed in the same survey by pacing, taping, and using an EDM instrument, they could all be combined in an adjustment by assigning appropriate relative weights.

Years ago, because of the comparatively heavy computational effort involved in least squares, nonrigorous or so-called rule-of-thumb adjustments were most often applied. However, now because computers have eliminated the computational expense, the reverse is true and least squares adjustments are performed more easily than these rule of thumb techniques. Least squares adjustments are less complicated because the same fundamental principles are followed regardless of the type of survey or the kind of observations. Also, the same basic procedures are used regardless of the geometric figures involved (e.g., triangles, closed polygons, quadrilaterals, or other more complicated networks). Rules of thumb, on the other hand, are not the same for all types of surveys (e.g., level nets use one rule and traverses use another), and the rules vary for different geometric shapes of each. Furthermore, the rule of thumb applied for a particular survey by one surveyor may be different from that applied by another surveyor. A favorable characteristic of least squares adjustments is that there is only one rigorous approach to the procedure, and thus no matter who performs the adjustment for any particular survey, the same results are obtained.

Least squares has the advantage that after an adjustment has been finished, a complete statistical analysis can be made of the results. Based on the sizes and distribution of the errors, various tests can be conducted to determine if a survey meets acceptable tolerances or whether the observations must be repeated. If blunders exist in the data, these can be detected and eliminated. Least squares enables precisions for the adjusted quantities to be determined easily, and these precisions can be expressed in terms of error ellipses for clear and lucid depiction. Procedures for accomplishing these tasks are described in subsequent chapters.

Besides its advantages in adjusting survey data, least squares can be used to plan surveys. In this application, prior to conducting a needed survey, simulated surveys can be run in a trial-and-error procedure. For any project, an initial trial geometric figure for the survey is selected. Based on the figure, trial observations are either computed or scaled. Relative weights are assigned to the observations in accordance with the precisions that are estimated by considering the different combinations of equipment and field procedures. A least squares adjustment of this initial network is then performed, and the results analyzed. If goals have not been met, the geometry of the figure and the observation precisions are varied and the adjustment performed again. In this process, different types of observations can be used and observations can be added or deleted. These different combinations of geometric figures and observations are varied until one is achieved that produces either optimum or satisfactory results. The survey crew can then proceed to the field, confident that if the project is conducted according to the optimum design, satisfactory results will be obtained. This technique of applying least squares in survey planning is discussed further in later chapters.

1.8 OVERVIEW OF THE BOOK

In the remainder of the book, the interrelationship between observational errors and their adjustment is explored. In Chapters 2 through 5, methods used to determine the reliability of observations are described. In these chapters, the ways that errors of multiple observations tend to be distributed are illustrated, and techniques used to compare the quality of different sets of observed values are examined. In Chapters 6 through 9 and in Chapter 13, methods used to model error propagation in observed and computed quantities are discussed. In particular, error sources present in traditional surveying techniques are examined, and the ways these errors propagate throughout the observational and computational processes are explained. In Chapters 21 and 25, methods used to locate blunders in sets of observations after an adjustment is explored. In the remainder of the book, the *principles of least squares* are applied to adjust observations in accordance with random error theory, and techniques used to locate mistakes in observations are examined.

PROBLEMS

Note: Partial answers to problems marked with an asterisk are given in Appendix H.

1.1 Describe the steps involved in accounting for the existence of errors in observations.

1.2 List two examples of direct measurements.

1.3 List two examples of indirect measurements.

1.4 List four unconditional statements about measurements.

1.5 Define the sources of errors and provide an example of each.

1.6 State whether the following are mistakes, systematic errors, or random errors.

(a) Earth curvature and refraction.
(b) Imperfect centering over a point.
***(c)** Contraction of a tape due to a temperature other than standard.
(d) Recording a length of 135.48 as 135.44.
(e) Reading of the horizontal circle on a total station.

1.7 Discuss the difference between a mistake and an error.

1.8 Give an example of

(a) a random instrumental error.
(b) a random natural error.
(c) a random personal error.

1.9 Define the term discrepancy.

1.10 Identify each of the following errors as either systematic or random.

(a) Bubble not centered at the instance of reading a rod.
(b) A vertical circle indexing error on a total station.
***(c)** Use of a tape that is not of standard length.
(d) Reading the graduated horizontal circle on a total station.

1.11 Define precision and accuracy.

1.12 Identify each of the following errors according to its source (natural, instrumental, personal):

***(a)** Level rod length.
(b) EDM–reflector constant.
(c) Air pressure in an EDM observation.
(d) Reading a graduation on a level rod.
(e) Earth curvature in leveling observations.
(f) An inclined line of sight in an automatic level.

***1.13** The calibrated length of a particular line is 199.998 m. A length of 200.001 m is obtained using an EDM. What is the error in the observation?

1.14 In Problem 1.13, if the observed length is 199.996 m, what is the error in the observation?

1.15 Why do surveyors measure angles using both faces of a total station (i.e., direct and reversed)?

1.16 List two examples of redundant observations in surveying.

1.17 Discuss the importance of making redundant observations in surveying.

1.18 A line is observed as 156.58 ft, 156.60 ft, and 156.61 ft. How many redundant observations were observed?

1.19 Why do prudent surveyors always obtain redundant measurements?

1.20 What mathematical condition is enforced in a least squares adjustment?

1.21 List the advantages that the least squares method has over rule-of-thumb adjustment procedures.