1

Introduction and Definitions

This book provides an introduction to basic aspects of uncertainty analysis as used in the physical sciences and engineering. We focus on one objective: enabling readers to determine appropriate uncertainty limits for measured and calculated quantities. Better conclusions can be drawn from results when they are reported with the appropriate uncertainty limits.

The topic of data uncertainty is often called *error analysis*, a phrase that has the unfortunate connotation that some error has been made in the measurement or calculation. When something has been done incorrectly, we call that a blunder or a mistake. Known mistakes should always be corrected. Due to limitations in both the precision and accuracy of experiments, however, uncertainty is present even when a measurement has been carried out correctly. In addition, there may be hard-to-eliminate random effects that prevent a measurement from being exactly reproducible. Our goal in this book is to present a system for recognizing, quantifying, and reporting uncertainty in measurements and calculations.

It is straightforward to use replication and statistics to assess uncertainty when only *random errors* are present in a measurement – errors that are equally likely to increase or decrease the observed value. We discuss random error and the use of statistics in error analysis in Chapter 2. While replication techniques are powerful, they cannot directly quantify many *nonrandom* errors, which are very common and which often dominate the uncertainty. It is much harder to account for nonrandom or *systematic* errors because they originate from a wide variety of sources and their effects often must be estimated. In dealing with systematic errors it helps if we, too, are systematic, so that we can keep track of both our assumptions and how well our assumptions compare with reality. By following an orderly system, we zero in on good error estimates and build confidence in our understanding of the quality and meaning of our measurements.

In this book we focus on three categories of error: random (Chapter 2), reading (Chapter 3), and calibration (Chapter 4). For all three types we provide worksheets (Appendix A) that guide our choices about the likely error magnitudes. If, as is often the case, an experimental result is to be used in a follow-on calculation, we must propagate the errors from each variable to obtain the uncertainty in the calculated result. This process is explained in Chapter 5, and again a worksheet is provided. The worksheet format for error propagation is easily translated to a computer for repeated calculations. In Chapter 6, we provide an introduction to determining error measures when using software to produce curve fits (Microsoft Excel's LINEST or MATLAB).

Through the use of the error-analysis methods discussed in this text, one can determine plausible error limits for experimental results. The techniques are transparent, with the worksheets keeping track of assumptions made as well as the relative impacts of the various assumptions. Keeping track of assumptions facilitates the ever-so-important process of revisiting error estimates, when, for example, seeking a more precise answer, seeking to improve a measurement process, or trying to improve the error estimates previously made.

Having stated our goals, we now devote the rest of this chapter to discussing how four important concepts relate to our error-analysis system: precision and accuracy; significant figures; error limits; and types of uncertainty or error. In Chapter 2 we turn to the statistics of random errors – the statistics of random processes form the basis of all error analyses.

1.1 Precision and Accuracy

In common speech the words "precision" and "accuracy" are nearly synonyms, but in terms of experimentally determined quantities, we use these words to describe two different types of uncertainty. *Accuracy* describes how close a measurement is to its true value. If your weight is 65 kg and your bathroom scale says that you weigh 65 kg, the scale is accurate. If, however, your scale reports 75 kg as your weight, it fails to reflect your true weight and the scale is inaccurate. To assess accuracy, we must know the true value of a measured quantity.

> To assess accuracy, we must know the true value of a measured quantity.

Precision is also a measure of the quality of a measurement, but precision makes no reference to the true value. A measurement is precise if it may be distinguished from another, similar measurement of a quantity. For example, a bathroom scale that reports weight to one decimal place as 75.3 kg is more precise than a scale that reports no decimal places, 75 kg. Precise numbers have more digits associated with them. With a precise bathroom scale, we can distinguish between something that weighs 75.2 kg and something that weighs 75.4 kg. These two weights would be indistinguishable on a scale that reads only to the nearest kilogram.

From this discussion we see that accuracy and precision refer to very different things: a three-digit weight is precise, but knowing three digits does not ensure accuracy, since the highly precise scale may be functioning poorly and thus reporting an incorrect weight. Measurement precision and accuracy both affect how certain we are in a quantity. In this book, precision is associated with random and reading error (discussed in Chapters 2 and 3, respectively), while accuracy is addressed in calibration error (discussed in Chapter 4).

1.2 Significant Figures

An important convention used to communicate the quality of a result is the number of *significant figures* reported. The significant-figures convention (sig figs) is a short-hand adopted by the scientific and engineering communities to communicate an estimate of the quality of a measured or calculated value. Under the sig-figs convention, we do not report digits beyond the known precision of a number. The rules of the sig-figs convention are provided in Appendix B.

The rules of significant figures allow the reporting of all certain digits and, in addition, one uncertain digit. The idea behind this is that we likely know a number no better than give-or-take "1" in the last digit. For example, if we say we weigh 65 kg, that probably means we know we weigh in the 60s of kilograms, but our best estimate of the exact value is no better than the range 64–66 kg (65 ± 1 kg). As we discuss in this book, determining the true uncertainty in a measurement is more complicated than just making the "± 1 in the last digit" assumption (it must involve knowing how the quantity was measured, for example). The sig-figs convention is therefore only a rough, optimistic indication of certainty. The limited goal of the sig-figs convention is to prevent ourselves reporting results to more precision than is justified.

> The significant-figures convention is
> a rough, optimistic indication of certainty.

A person making a calculation often faces the difficulty of assigning the correct number of significant figures to a result. This occurs, for example, when extra digits are generated by arithmetic. For example, if we use a precise bathroom scale to weigh 12 apples as 2.3 kg (two significant figures in both numbers), we can use a calculator to determine that the average mass of an apple is

$$\text{average apple mass} = \frac{2.3\ \text{kg}}{12} = 0.19166666666667\ \text{kg}$$

The calculator shows a large number of digits to accurately convey the arithmetic result. It would be inappropriate, however, to report the average mass of these apples to 14 digits. The rules of significant figures guide us to the most optimistic number of digits to report. As discussed in Appendix B, there are rules for how significant figures propagate through calculations, depending on whether we are adding/subtracting or multiplying/dividing numbers in the calculation. The sig-fig rules are based on the error-propagation methods discussed in Chapter 5. In the case of the average mass of an apple, we assign the answer as 0.19 kg/apple (two sig figs). This is because the sig-figs rule when multiplying/dividing is that the result may have no more sig figs than the least number of sig figs in the numbers in the calculation – in this case, two (see Appendix B).[1]

Learning and using the significant-figures rules allows us to write results more correctly. The principal advantage of following the sig-figs convention is that it is easy, and it prevents us from inadvertently reporting significantly more precision than we should. However, because the sig-figs convention is approximate and may still greatly overestimate precision and accuracy (see Example 1.1), we cannot rely solely on sig-figs rules for scientific and engineering results. When important decisions depend on our calculations, we must perform rigorous error analysis to determine their reliability – this is the topic of this book.

[1] In intermediate calculations we retain all digits to avoid round-off error. It is only the final reported result that is subjected to the sig-figs truncation. For greater arithmetic accuracy, intermediate calculations made by calculators and computers automatically employ all available digits. If the number obtained is itself an intermediate calculation, all digits should be employed in downstream calculations.

Example 1.1: Temperature displays and significant figures. *A thermocouple is used to register temperature in an experiment to determine the melting point of a compound. If the compound is pure, the literature indicates it will melt at* 46.2 ± 0.1^oC. *A sample of the chemical is melted on a hot plate and a thermocouple placed in the sample reads* 45.2^oC *when the compound melts. Is the compound pure?*

Solution: The tenths place of temperature is displayed by the electronics connected to the thermocouple. Assuming the display follows the rules of significant figures, this optimistically implies that the device is "good" to $\pm 0.1^\circ$C. Since the highest observed melting temperature is therefore $45.2 + 0.1 = 45.3^\circ$C, we might conclude that the compound is not pure, since the observed melting point is less than the lowest literature value obtained for the pure compound, $46.2 - 0.1 = 46.1^\circ$C.

Unfortunately, this would not be a justified conclusion. As we discuss in Chapter 4, the most accurate thermocouples have a calibration error that reduces their reliability to no better than $\pm 1.1^\circ$C. The manufacturers of thermocouple-based temperature indicators provide extra digits on the display to allow users to avoid round-off error in follow-on calculations; it is the responsibility of the user to assign the appropriate uncertainty to a reading, taking the device's calibration into account.

In the current example and using calibration error limits of $\pm 1.1^\circ$C (see also Example 4.6), the highest possible observed melting temperature is $45.2 + 1.1 = 46.3^\circ$C, which is within the expected range from the literature. It does not mean that the substance is pure, however; rather, we can say that our observations are consistent with the compound being pure or that we cannot rule out that the compound is pure. To actually test for purity with melting point, we would need to use a temperature-measuring device known to be accurate to within the accuracy of the literature value, $\pm 0.1^\circ$C.

1.3 Error Limits

This book explains accepted techniques for performing rigorous error analysis. We use *error limits* to report the results of our error analyses. Error limits are a specified range within which we expect to find the true value of a quantity. For example, if we measure the length of 100 sticks and find an average stick length of 55 cm, and if we further find that most of the time (with a 95% level of confidence) the stick lengths are in the range 45–65 cm, then we might report our estimate of the length of a stick as 55 ± 10 cm. The number 55 cm

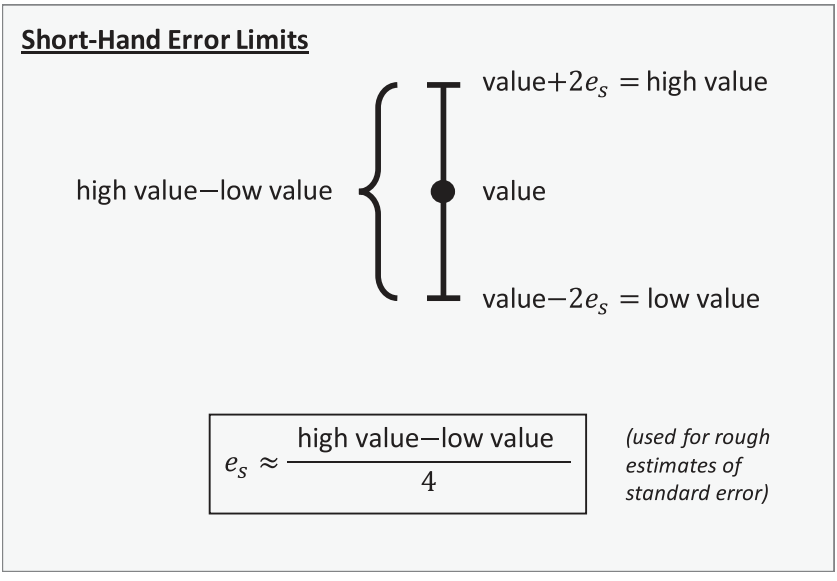

Figure 1.1 When errors are distributed according to the *normal* distribution (see Chapter 2 and Appendix E), then at a 95% level of confidence the true value will be within a range of about plus or minus two standard errors, e_s; thus, the span from the smallest probable value to the highest probable value is four standard errors. If we determine a highest and lowest reasonable values for a quantity, we can estimate a plausible standard error as one-fourth of that range.

is our best estimate of the length of a stick, but the uncertainty in the length is such that the length of any chosen stick might be as low as 45 cm and as high as 65 cm.

For the error limits in this example, we explained that the stated range represented the values observed with a 95% level of confidence. This type of error range is a form of 95% confidence interval (discussed in depth in Section 2.3.2), which is associated with an error range of plus or minus two *standard errors*. The standard error e_s is a statistic that we discuss more in Chapter 2. For now we can think of it as a kind of regularized error associated with a quantity. We need to regularize or standardize because uncertainty comes from many sources, and to combine errors from different sources the errors must be expressed in the same way. We discuss the accepted way to standardize errors in Chapter 2. Note that if with 95% confidence the true value is within plus or minus two standard errors of the best estimate of the number, then the total range between the lowest and highest probable values of a number (at a 95% level of confidence) is four standard errors (Figure 1.1). If we estimate a worst-case high value of a quantity and a worst-case low value, subtract the

two and divide by four, we obtain a rough estimate of the standard error e_s. We use this rule of thumb in situations where we are obliged to estimate error amounts from little information (see the example that follows and Chapter 4).

Example 1.2: Reader's weight, with uncertainty limits. *What do you weigh? Please give your answer with appropriate uncertainty limits. What is the standard error of your estimate?*

Solution: For a quantity, we are often able to estimate an optimistic (high) number and a pessimistic (low) number. As we have seen, it is conventional to use a 95% level of confidence on stochastic quantities, which corresponds to error limits that are about $\pm 2e_s$. For such a range, there is a span of $4e_s$ between the high value and the low value. This provides a method for estimating e_s.

Short-hand estimate for standard error

$$e_s \approx \frac{\text{high value} - \text{low value}}{4} \tag{1.1}$$

A person might answer the question about weight by saying he weighs at most 185 lb_f and at least 180 lb_f. Then we would write his weight as

$$e_s = \frac{185 - 180}{4} = 1.25 \text{ lb}_\text{f}$$
$$\text{average weight} = \frac{185 + 180}{2} \text{ lb}_\text{f}$$
$$\text{weight} = \text{average weight} \pm 2e_s$$
$$= 182.5 \pm 2.5 \text{ lb}_\text{f}$$

The standard error for this estimate is $e_s = 1.25$ lb_f.

Some scientists and engineers report error ranges using other than a 95% level of confidence: 68% or $\pm$ one standard error and 99% or $\pm$ three standard errors are sometimes encountered (These percentages assume that the errors are distributed via the normal distribution; see Appendix E). It is essential to make clear in your communications which type of error limit you are using. The 95% level of confidence is the most common convention; in this text we use exclusively a 95% level of confidence ($\approx \pm$ two standard errors).

1.4 Types of Uncertainty or Error

Measurement uncertainty has many sources. We find it convenient to divide measurement uncertainty into three categories – random error, reading error, and calibration error – and each of these errors has its own chapter in this book. We briefly introduce these categories here.

Figure 1.2 An electronic display (left) gives the reading from the scale underneath the water tank on the right. Displays from which we read data limit the precision of the measurement. This limitation contributes to reading error.

Random or *stochastic errors* are generated by (usually) unidentified sources that randomly affect a measurement. In measuring an outside air temperature, for example, a small breeze or variations in sunlight intensity could cause random fluctuations in temperature. By definition, random error is equally likely to increase or decrease an observed value. The mathematical definition of a *random process* allows us to develop techniques to quantify random effects (Chapter 2).

A second type of error we commonly encounter is *reading error*. Reading error is a type of uncertainty that is related to the precision of a device's display (Figure 1.2). A weighing scale that reads only to the nearest gram, for example, systematically misrepresents the true mass of an object by ignoring small mass differences of less than 0.5 g and by over-reporting by a small amount when it rounds up a signal. Another component of reading error is needle or display-digit fluctuation: we can estimate a reading from a fluctuating signal, but there is a loss of precision in the process. In Chapter 3 we discuss the statistics of reading error and explain how to estimate this effect.

The third category of error we consider is *calibration error*. While reading error addresses issues of precision, issues of accuracy are addressed through calibration. Instruments are calibrated by testing them against a known, accurate standard. For example, a particular torque transducer may be certified by its manufacturer to operate over the range from 0.01 to 200 mN·m. The manufacturer would typically certify the level of accuracy of the instrument by,

for example, specifying that over a chosen performance range the instrument is calibrated to be accurate to within ±2% of the value indicated by the device. In Chapter 4 we outline the issues to consider when assessing uncertainty due to calibration limitations.

As discussed further in Chapter 2, all three error types – random, reading, and calibration – may be present simultaneously, which leads to a complicated situation. If one of the three error sources dominates, then only that error matters, and the error limits on the measurement (at a 95% level of confidence) are plus or minus twice the dominant error in standard form. The task in that case is to determine the dominant error in standard form. If more than one error source matters, we must combine all effects when reporting the overall uncertainty. To combine independent errors we put them in their standard form and combine them according to the statistics of independent stochastic events (Chapter 2). It turns out that standard errors combine in quadrature [52].

Independent errors combine in quadrature

$$e_{s,cmbd}^2 = e_{s,random}^2 + e_{s,reading}^2 + e_{s,cal}^2 \tag{1.2}$$

where $e_{s,cmbd}$ is the combined standard error, and $e_{s,random}$, $e_{s,reading}$, and $e_{s,cal}$ are standard random, reading, and calibration error of a measurement, respectively.

Summarizing, uncertainty in individual measurements comes from three types of error:

1. **Random** errors due to a variety of influences (these are equally likely to increase or decrease the observed value)
2. **Reading** errors due to limitations of precision in measuring devices (systematic)
3. **Calibration** errors due to limitations in the accuracy of the calibration of measuring devices (systematic)

For a given measurement, each of these independent error sources should be evaluated (Chapters 2, 3, and 4) and the results combined in quadrature (Equation 1.2).

1.5 Summary

1.5.1 Organization of the Text

The focus of this chapter has been to establish the footings on which to build our error-analysis system. The project is organized into six chapters as outlined here.

- In this first chapter, we have defined terms and categorized the uncertainties inherent in measured quantities. We identify three sources of measurement uncertainty: random error, reading error, and calibration error. These three sources combine to yield a combined error associated with a measured quantity.

Independent errors combine in quadrature

$$e_{s,cmbd}^2 = e_{s,random}^2 + e_{s,reading}^2 + e_{s,cal}^2$$

- In Chapter 2 we discuss random errors, which may be analyzed through random statistics. From the statistics of stochastic processes we learn a method to standardize different types of errors: measurement error is standardized by making an analogy between making a measurement and taking a statistical sample. We also present a technique widely used for expressing uncertainty, the 95% confidence interval.
- In Chapter 3 we discuss reading error, which is a systematic error produced by the finite precision of the reading display of a measuring device or method. Sources of reading error include the limited number of digits in an electronic display, display fluctuations, and the limit to the fineness of subdivisions on a knob or analog display.
- In Chapter 4 we discuss calibration error, a systematic error attributable to the finite accuracy of a measuring device or method, as determined by its calibration. The accuracy of the calibration of a device is known by the investigator who performed the calibration, and thus the instrument manufacturer is the go-to source for calibration accuracy. If the manufacturer's calibration numbers are not known or are difficult to find, we suggest how to estimate the calibration error using rules of thumb or other short-cut techniques (at our own risk). Also in Chapter 4 we begin two sets of linked examples that follow a calibration process and the use of a calibration curve. These examples are discussed and advanced in Chapters 4–6, as mapped out in Section 1.5.2.
- In Chapter 5 we discuss how error propagates through calculations; this discussion is inspired by how stochastic variables combine.
- The final chapter of the book is dedicated to one family of error-propagation calculations, those associated with ordinary least squares curve fitting. In Chapter 6 we discuss the process of fitting models to experimental data and of determining uncertainty in quantities derived from model parameters. These are computer calculations – we use Microsoft Excel's LINEST and MATLAB in our discussion.