

Computational Geology 1

Significant Figures!

H.L. Vacher, Department of Geology, University of South Florida, 4202 E. Fowler Ave., Tampa FL 33620

Introduction

It is not unusual for undergraduate geology majors to be required to take a year of calculus and a year of physics. It is also not unusual for those students to leave that information in "that other building" when they go to their geology classes. Let me give you an example: the issue of significant figures.

On day one, or close to it, your physics professor no doubt told you about significant figures. You dutifully learned that if you multiply, say, 153.671 by 4.7, the result is 720, not the 722.249 that is displayed by your calculator; the operative reason, of course, is that 4.7 has only two significant figures, and so the result of the multiplication cannot have more than two. Similarly, you learned that if you needed to add those same two numbers, the result would be 158.4; in the case of this addition, the determinative fact is that the 4.7 is known to only the tenths place, and so the result of the addition cannot be known to hundredths or thousandths. Anyway, that is what you learned in physics. Let's go on to geology.

After the calculus and physics you may have a course in hydrogeology in your senior year. This is one of the courses in which mathematics is essential to one's ability to understand the material and where solving of problems is a crucial ingredient of learning it. Here is a sample problem (modified slightly from problem 9 of chapter 5 in Fetter's *Applied Hydrogeology*, 3rd edition, MacMillan Publishing Company., the standard undergraduate text on the subject):

A confined aquifer is 12 ft thick. The potentiometric surface is at 277.86 ft above sea level (a.s.l.) in one well and 277.32 ft a.s.l in a second well. The two wells are 792 ft apart and (fortuitously) aligned in the direction of groundwater flow. If the hydraulic conductivity is 21 ft/day, and the effective porosity is 0.17, what is the average linear velocity (also known as the "tracer velocity", the average velocity at which a tracer, moving with the groundwater, would travel from the first well to the second)?

This problem is of the variety that students come to call, fondly, "plug and chug". The equation, into which numbers are to be plugged, is the one-dimensional form of Darcy's Law written for velocity:

$$v = -\frac{K \Delta h}{n_e \Delta s}, \quad (1)$$

where v is the average tracer velocity, n_e is the effective porosity, Δh is the difference in elevation of the potentiometric surface between the two wells, and Δs is the distance between the two wells in the direction of flow; the thickness of the aquifer does not enter into this problem. Plugging the values into the equation you get:

$$v = -(21/0.17)*(277.32 - 277.86)/792 . \quad (2)$$

Chugging out the value, your calculator displays

$$v = 0.0842246. \quad (3)$$

What number do you write down as the answer? Regrettably, many students would write 0.0842246 ft/day on their homework paper or exam. Many would recognize that six significant figures would be excessive but then, amazingly, write down 0.0842 ft/day. Both groups would be wrong. The correct answer is 0.084 ft/day, and it is clearcut. K , n_e , and Δh , which are obviously multiplied and divided together – each has only two significant figures.

Why It Matters

The issue is one of "truth in advertising": implied uncertainty. When you give the result of a calculation, you communicate two things: (1) the number itself, and (2) *how well that number is known*.

If you say from the given data that $v = 0.0842$ ft/day, then you are misleading your reader, your client, yourself – whomever you do the calculation for – about the uncertainty attached to the calculated result. The amount that you are misleading your audience is considerable. If, for example, you give three significant figures when only two are appropriate, you are saying that the uncertainty is, in general, ten times less than it actually is. If you give four significant figures instead of the correct two, you understate the uncertainty, generally, by 100 times. In other words, you are significantly misrepresenting your knowledge; you are implying that the precision of your result is much better than it actually is.

The key to the issue is that the number of significant figures is related to the implied uncertainty. A rough rule of thumb connecting the two is:

<u>number of sig figs</u>	<u>implied uncertainty</u>
1	10%
2	1%
3	0.1%
4	0.01%

or, in general, if a is the number of significant figures, then the rule of thumb says,

$$\text{implied uncertainty} = 100 * 10^{-a}$$

where the implied uncertainty is given as a percent. With this rule of thumb, the students who give $v = 0.0842246$ ft/day for the preceding problem are implying that this velocity is known with a precision on the order of a *ten-thousandth of a percent*, whereas the data going into the calculation have a precision of only about one percent.

To see where the rule of thumb comes from, consider the number 5, halfway through the range of one-digit numbers. If you read that a continuous variable (such as elevation, or distance, or hydraulic conductivity) has a value of 5, you have to assume that the author means 5 ± 0.5 (i.e., 4.5 to 5.49, with the convention of "rounding up" to even numbers). For if the author would have meant 5.3, say, then that would be the number that was used. In the same way, if you see that the variable has a value of 5.3, then, absent any information to the contrary, you have to assume that the number can be anywhere in the range, 5.3 ± 0.05 . The reason is that values within that range would be rounded off to 5.3; similarly values within the range 5 ± 0.5 would be rounded off to 5. These plus-minus terms illustrate the implied uncertainties associated with any stated number.

The implied uncertainties of 0.5, 0.05, and the like, are implied *absolute* uncertainties. In contrast, there are implied *relative*, or fractional, uncertainties. In the case of 5 ± 0.5 , the relative uncertainty is $0.5/5$ or 10%; in the case of 5.0 ± 0.05 , the relative uncertainty is $0.05/5.0$ or 1%. Thus the relative uncertainty is simply the absolute uncertainty divided by the stated value of the number. It should be clear that the implied relative uncertainty is independent of the order of magnitude of the stated number; that is, the implied relative uncertainty of 5 is the same as the implied relative uncertainty of 0.5, 0.005, 5×10^6 , or 5×10^{-6} . All that matters is that the value 5 is known to only one significant figure. In contrast, 5.0×10^6 has an implied relative uncertainty of 1%. The rule of thumb outlined in the preceding table is simply a listing of the implied relative uncertainties associated with the numbers 5, 5.0, 5.00 and so on.

The reason that the rule of thumb is only a rule of thumb is that the implied relative uncertainties depend on the number itself. For example, the number 1.0 (two significant figures) has an absolute implied uncertainty of 0.05 and a relative implied uncertainty of 5%; similarly the number 9.9 (also, two significant figures) has an absolute implied uncertainty of 0.05 and a relative implied uncertainty of 0.5%. So, the implied relative uncertainty of a number with two significant figures actually spans a range from 0.5% to 5%; the rule of thumb says 1%. Similarly, the implied relative uncertainty is 6% to 50% for a number with one significant figure and 0.05% to 0.5% for a number with three significant figures; the rule of thumb for these gives 10% and 0.1 percent respectively.

The point to remember: When you say that you have determined the value of something to a number with three significant figures, then you are saying that you know that number to a plus-minus of the order of 0.1 percent. If the rules of significant figures say that you should use only two significant figures for that number, then you should be communicating a plus-minus term of 1%. In other words, you are effectively overstating how well you know the number by a factor of 10 for each figure that you incorrectly add to the string of digits.

How Good Are the Rules of Significant Figures?

The point of all the preceding is that the rules of significant figures provide a way of handling uncertainties: by following these rules, you guard against understating the uncertainty attached to the numbers you communicate. It is fair, then, to ask: How successfully do these rules handle uncertainties?

To get an idea about the efficacy of the rules, we can compare the result that we get using

significant figures to the result we get by dealing with the implied uncertainties explicitly. Recall, the answer to the Darcy's Law problem posed in the Introduction was 0.084 ft/day; with the two significant figures, and using the rule of thumb for implied relative uncertainties, the implication is that we know the calculated value to a precision of about 1%. What do you get if you bring all the implied plus-minus terms out into the open and carry them through the arithmetic of Darcy's Law?

The question that has just been asked comes under the heading of *error propagation*, where, in the lingo of the subject, "error" simply means "uncertainty". There are definite rules on this matter, but they are beyond the scope of the present column. If you are interested in the subject, I recommend the book, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, by John R. Taylor (2nd ed., University Science Books, Miller Valley CA, 352 pp., 1997; ISBN: 093570275. \$29.50). This book is intended for an introductory (freshman- or sophomore-level) college course in experimental physics. Incidentally, this book has one of the all-time best covers for college textbooks: a photograph of a locomotive hanging out of an upper-story bank of windows in a train station – how is that for an image appropriate to the issue of ignoring uncertainties?

Without the rules of error propagation, we can still get an idea of the uncertainty in the calculated velocity that would result from all the implied uncertainties of the variables plugged into the Darcy's Law problem. Specifically, the hydraulic conductivity is implied to be in the range, 21 ± 0.5 ft/day; the effective porosity is 0.17 ± 0.005 ; the potentiometric surface at the first well is 277.86 ± 0.005 ft; the potentiometric surface at the second well is 277.32 ± 0.005 ft; and the distance is 792 ± 0.5 ft. Then, to get the maximum velocity, v , take the largest K (21.5 ft/day), the smallest n_e (0.165), the largest elevation for the potentiometric surface at the first well (277.865 ft), the smallest elevation for the second well (277.315), and the smallest distance (791.5 ft). Plugging these values into the equation gives 0.0905 ft/day. For the minimum velocity, use the smallest K (20.5 ft/day), the largest n_e (0.175), the smallest elevation for the first well (277.855 ft), the largest elevation for the second well (277.325), and the largest distance (792.5). The result from these values is 0.0783 ft/day.

The finding from this inelegant calculation of propagated uncertainties is that the velocity range generated by all the uncertainties in the original variables is 0.0783 to 0.0905 ft/day, or 0.0842 ± 0.006 ; the 0.0842 comes from using the straight values for the variables. This absolute uncertainty (0.006) translates to a relative uncertainty of 7%. This is the uncertainty in v that results from the implied uncertainties in K , n_e , h_1 , h_2 and Δs , if they all interact in the worst possible way. Clearly, this propagated uncertainty is much larger than the 1% you are implying if you use the rules of significant figures *correctly* and report 0.084 ft/day as your velocity. (If you use the techniques discussed in Taylor's book and allow for likely partial cancellation of the effects of the uncertainties of the independent variables, you get 0.0842 ± 0.003 , or a relative uncertainty of 4%. This is still larger than the 1% implied by using two significant figures.)

The result of this example is not at all unusual. Commonly, the correct number of significant figures understates the amount of uncertainty propagated from the implied uncertainties of the original variables. The message to take from this is: Although the rules of significant figures *try* to deal with the propagation of implied uncertainties, the rules are NOT conservative. That means that even if you obey the rules, you may well be misleading your audience to some extent about how well you know the number you are reporting. That fact makes the overstatement caused by *ignoring* the rules of significant figures even more problematic.

The Rules, Again

Assuming that you are convinced now that the rules of significant figures are not to be trifled with, you might like to see them spelled out again in one place. Here they are:

1. When you multiply or divide, round off to the same number of figures as in the factor with the least number of figures.
2. When you add or subtract, it is the position of the digits that is important, not how many there are. Imagine that the various terms that are added or subtracted together are in a vertical column with the decimal points all lined up. Spot the term that goes least far to the right. Round off to that position.
3. If you have a calculation that involves multiple steps, keep additional digits through the intermediate results and round off at the end. If you need to report one of the intermediate results, round off to the appropriate number of significant figures when you report it, but go back and pick up the discarded figures when you proceed further with the calculation.

As you probably know, there are some things to watch out for:

First, be careful about zeros to the left of the decimal point. Unless you take steps to be explicit, there are ambiguities about whether a number such as 720, which came up in the multiplication at the very beginning of this discussion, has two or three significant figures. One way to explicitly state the number of significant figures is to write 7.2×10^2 , which indicates two significant figures; three significant figures would be written 7.20×10^2 . Another way is to underline the last significant figure: $7\underline{2}0$. And, there's nothing more straightforward than telling your reader exactly what you mean: "720 to two significant figures."

Second, be aware that some numbers in equations are immune to considerations of significant figures. For example, in the equation giving the circumference of a circle as $2\pi r$, the "2" has an infinite number of zeros behind the decimal; the same is true for the "2" in πr^2 for the area of the circle. Coefficients and exponents in empirical equations, however, are not infinitely precise. For example, if you are using a linear equation

$$y = ax + b$$

or an exponential function

$$y = a \exp(-bt)$$

or a power function

$$y = ax^b$$

where the values of a and b were determined by some form of curve fitting (regression), then there are uncertainties associated with these numbers and you cannot consider them to be like the "2" in the circle formulas. Similarly, you need to be aware of the number of significant figures you are using for irrational numbers (e.g., π) that appear in equations.

Finally, keep track of how many significant figures you use for conversion factors. Use enough so that the precision of the final result is limited by the precision of the measured variables rather than the number of significant figures that you use in converting units.

Concluding Comment

The purpose of attending to significant figures, remember, is to communicate an appropriate uncertainty for your calculated results. There is a better way, of course: explicitly state the uncertainties of the measured variables -- 21.2 ± 0.7 , say -- and work directly with them. You have two choices: attend to significant figures, or learn and apply the rules of error propagation. You cannot choose to do neither.

Acknowledgment

I thank Prof. P. Mukherjee of USF Physics Department for helpful comments on an early draft of the manuscript.