Environmental Data Laboratory
Professor Kirchner

Laboratory 3: Propagation of uncertainty in hydrological mass balances

Note: you'll need a calculator for this lab

This lab is designed to expose you to several key concepts in error propagation and uncertainty analysis. In this lab, you will use mass balance methods to determine the evapotranspiration rate of a forest in Wales.

Mass balances are often used to help unravel how ecosystems function. Where ecosystem processes are not amenable to direct measurement, one can still measure their net effect by mass balance. In a mass balance, one sums all of the fluxes of a given chemical into an ecosystem, then subtracts all of the fluxes leaving the ecosystem; the difference between inputs and outputs is the net uptake or release by the ecosystem. Note that this does not tell us *where* in the ecosystem the uptake or release is occurring, or *how*, or *why*. This kind of "black box" analysis does not, by definition, tell you much about the details of what's going on inside the box.

Here you will be working with data from the Plynlimon watershed in Wales, which has been the site of research on timber management, water quality, and forest biogeochemistry for nearly 15 years. The Institute of Hydrology in the U.K. has provided the data you will be working with. You will be working with only a tiny fraction of the whole Plynlimon data set. I have edited the data, in order to re-create a simple problem in biogeochemical research: assessing what kind of sampling program would be sufficient to determine, with acceptable precision, the rate at which a forest is evaporating water to the atmosphere. While this may not be exactly the same as the experimental design problems you may face in your future work, it is a simple example that illustrates some important points. In this lab, you will perform uncertainty analysis by using the simple formulas of the Error Propagation Toolkit, and you will test your results with a "Monte Carlo" simulation on JMP. Anything you need to hand in (questions you need to answer, and so forth) is indicated by the symbol "•" and a number, so you can just hand in a sheet of paper with your answers, comments, explanations, and so forth indicated by the numbers used in this handout.

Evapotranspiration is the difference between rainfall and runoff, assuming there is no net change in groundwater storage over annual timescales. The uncertainty in the annual rainfall and the annual runoff will together contribute to uncertainty in the evapotranspiration you calculate. We will start from the bottom and work up.

Two memorable rules for propagation of uncertainty are (see Error Propagation Toolkit):

"Simple rule for sums and differences":

<u>if</u>	z=x plus or minus y, plus or minus q, plus or minus w, etc.
<u>and</u>	*x*, *y*, *q*, *w*, etc. are uncorrelated with one another
then	the standard errors add *in quadrature* (that is, squared, added, and then square rooted).

$$s_{\bar{z}} = \sqrt{\left(s_{\bar{x}}\right)^2 + \left(s_{\bar{y}}\right)^2 + \left(s_{\bar{q}}\right)^2 + \left(s_{\bar{w}}\right)^2}$$

"Simple rule for products and ratios":

<u>if</u>	*z*=*x* times/divided by *y* times/divided by *q* times/divided by *w* times/divided by ...
<u>and</u>	*x*, *y*, *q*, *w*, etc. are uncorrelated with one another
then	the percent (or fraction) standard error of *z* can be found by adding the percent (or fraction) standard error in each of its components, in quadrature:

$$\frac{s_{\bar{z}}}{\bar{z}} = \sqrt{\left(\frac{s_{\bar{x}}}{\bar{x}}\right)^2 + \left(\frac{s_{\bar{y}}}{\bar{y}}\right)^2 + \left(\frac{s_{\bar{q}}}{\bar{q}}\right)^2 + \left(\frac{s_{\bar{w}}}{\bar{w}}\right)^2}$$

A: Uncertainty in rainfall and runoff

In the JMP file "Rainfall&runoff.jmp" you will find annual totals of rainfall (in units of water depth--mm per year) from rainfall totals caught each week in rain buckets at Plynlimon. You will also find annual runoff totals (in the same units). These are the sums of once-a-week readings of stream gauges. Note that there is an important difference between the methods used to measure rainfall and runoff. The rainfall measurements are *cumulative*; anything that falls into the buckets becomes part of the sample. By contrast, the runoff measurements are *instantaneous*; streamflow at one instant during each week is assumed to be representative of the whole week's flow. Thus it is likely that the runoff totals are less reliable than the rainfall totals.

Open the JMP file, "Rainfall&runoff". Using the "distribution of Y's" analysis, plot histograms of the annual rainfall and runoff totals. •1: What are the 10-year means of rainfall and runoff, and what are the standard errors of the means? The clearest way to write this is, "1234±56 mm/yr (mean±std. error)". That's the format you should use; you can abbreviate standard error as "s.e." if you want to.

As we have said in class, there should be a 95 percent chance that a sample mean falls within two standard errors (on either side) of the true mean. Equivalently, we can have about 95 percent confidence that the true mean lies within about 2 standard errors on either side of our sample mean. The exact t value for a 95-percent, two-tailed confidence interval is $t_{0.025,9}=2.262$. •2: What is the 95-percent confidence interval for the 10-year mean rainfall? What's the same confidence interval for the 10-year mean runoff? (If you aren't sure how to calculate confidence intervals, consult the Toolkit, "Confidence Intervals".) Probably the clearest way to write these confidence intervals is, "2000 (1600-2400) mm/yr (mean and 95% CI), so use that format.

B: Monte Carlo simulation of uncertainty in rainfall and runoff

Okay, the standard error is an estimate of the expected uncertainty in the 10-year averages of rainfall and runoff. That is, if we had a large sample of these 10-year averages, we would expect them to differ from the "true" average by about the standard error that you calculated. You have also calculated a 95-percent confidence interval; in our hypothetical large sample of 10-year averages, we would expect about 95 percent of them to lie within that interval. Do your standard error and confidence interval perform as advertised? One way to find out would be to wait until we had several centuries of data, so we could calculate many 10-year averages, and see how they are distributed. I don't know about you, but to me that seems like a long time to wait, particularly for a three-hour lab. So here we will instead use the data we already have to *simulate* the results of many years of sampling, using randomization methods. These techniques are called Monte Carlo methods; I leave it to you to guess where the name comes from...

What we will do is randomly sample, with replacement, from the data we already have. Imagine that we write the annual rainfall totals on 10 slips of paper and put them in a hat. Then we take a slip of paper out of the hat at random, note the rainfall, and put the slip back in the hat. We do this for a total of 10 simulated years, then calculate the 10-year average for our random sample of rainfall totals. In a similar way, we could randomly generate 10 years of runoff "data". Then we could repeat this process over and over, and thus compile, by brute-force simulation, distributions of evapotranspiration estimates. The technical Monte Carlo term for this is a "bootstrap procedure"; you are "bootstrapping" from a limited sample to simulate a larger number of possibilities. These sorts of brute-force simulations are ideal for computers, which can draw numbers out of a hat all day and never get bored.

The concept outlined above is straightforward enough, but actually doing it in JMP is a bit squirrely, since JMP wasn't exactly designed with Monte Carlo simulation in mind. But follow along, and you'll get the hang of it (you may actually already *have* the hang of it, since this is similar to Lab #2).

Now we first need to tell JMP to draw 10 rainfall totals at random, and average them. Here's how it's done. First, create a new formula column called, say, "rainfall 10-yr random avg." Then, do the following to build the formula. First, from the function window, select "Statistical" (you may need to scroll the list to do this), then select "Summation" from the sub-list that appears. Then insert "rainfall(mm/yr)" from the left-hand variable list into the big empty box marked "body". Now, change the "NRow()" at the top of the summation to "10" (click or double-click on "NRow()", type "10", and press return or enter). What you'll see at this stage is:

$$\sum_{i=1}^{10} rainfall(mm/yr)$$

This still isn't quite what you want, for two reasons. First, it's not an average, it's just a sum. Second, it's simply the sum of the first ten values, not ten randomly chosen values. To get randomly chosen values, here's what you need to do. Select the box enclosing 'rainfall(mm/yr)' in the formula. Then from the list of functions (the right-hand list), select "Row" and then "Subscript" from the sub-list that appears. Now, highlight the subscript box by clicking on it (if it is not highlighted already), and from the list of functions select "Random" (you will probably need to scroll down to see this), and then select "Random Integer" from the sub-list that appears. Next, highlight the box inside the parentheses of the "RandomInteger()" function, and enter the number 10 there.

Finally, you need to divide by 10. So, select the whole function (by clicking on the outline box around the whole thing), click "÷", and enter the value "10" below the line. Now you should see:

$$\frac{\sum_{i=1}^{10} rainfall(mm/yr)_{RandomInteger(10)}}{10}$$

Now, this is just what you want. What it says is, "Pick a random integer between 1 and 10, then go get the rainfall value that is stored on that row. Do this ten times, sum the results, and divide by ten."

Now, you're going to want to perform this experiment more than 10 times. So, using "add rows" from the "rows" menu, add 490 rows to the bottom of your data table, yielding a total of 500 rows. You now have 500 randomly-generated simulations of 10-year averages of the rainfall data. Now you need to create another formula column that averages 10 randomly-chosen runoff figures (the formula is the same as above, but with runoff rather than rainfall). Note that by cleverly copying and pasting in the formula windows, you can do this very easily. So now you have 500 decades of simulated rainfall and runoff averages.

Tedious? Perhaps...but not nearly as tedious as waiting five millennia for the real measurements to come in. Now we can get back to the original questions we asked above: do the standard error and confidence interval that you calculated above accurately reflect the likely range of variability in the decadal averages of rainfall and runoff? The answer lies in the distribution of the Monte Carlo 10-year averages. Generate histograms of these using the "distribution of Y's" option.

****Semantic Confusion Alert**** It's important to remember that you are looking at a distribution of *means*, not a distribution of individual data points. So, the standard *error* you calculated above should roughly equal the standard *deviation* of these means around their own central value. The standard error in *this* window has no real meaning, for two reasons. First, the distribution itself is made up of means rather than individual measurements; we want to know about the variability of the means, not the mean of the means. Second, the standard error is a function of the number of input data and our Monte Carlo procedure generates an artificially large *n*.

•3: Are the means of the rainfall and runoff 10-year averages close to the averages that you got in #1, above? (They should be; if not, you've done something wrong).

•4: How variable (as expressed by the standard deviation) are the rainfall and runoff 10-year averages? The clearest way to write this is "10-year avg. rainfall 1234±567 (mean±std. dev.)". Is that variability close to what you estimated as the standard error of the means in #1, above? (Again, it should be; if it isn't, you've done something wrong).

•5: What are the quantiles that enclose the middle 95 percent of your Monte Carlo samples (that is, what are the 97.5th and the 2.5th percentiles of the distribution)? Were your estimates in #2, above, approximately correct? (Note that they won't be exactly correct, because the thin tails of any distribution are more vulnerable to random variation than the middle is.)

<u>C: Predicting uncertainty in evapotranspiration:</u>

Now, let's try a simple error propagation exercise. From the difference between rainfall and runoff, we can estimate net evapotranspiration by the forest (evapotranspiration rate=rainfall-runoff). •6: Using the simple rule for sums and differences (given on the first page), calculate the 10-year average evapotranspiration rate and its standard error. *Calculate* this (and show your work), don't *simulate* it, at least not yet.

Equivalently, we can have about 95 percent confidence that the true mean lies within 2 standard errors (more precisely 2.262 standard errors) on either side of the mean we might get from 10 years of data. •7: What are the upper and lower bounds of the 95-percent confidence interval for the 10-year average evapotranspiration?

You will note that this confidence interval includes negative values, and you might wonder what negative evapotranspiration means. It could mean one of two things. First, if there is direct condensation that is not counted as rainfall, it could be considered a negative evapotranspiration flux. But remember that your estimate of uncertainty doesn't just reflect the variability in what evapotranspiration (ET) *actually is*, it also reflects the accuracy (or inaccuracy) of your methods for *measuring* or *inferring* ET. So even if negative ET were physically impossible, you might arrive at a negative *measurement* of ET.

It is worth considering how much of the uncertainty in ET comes from uncertainty in rainfall, vs. uncertainty in runoff. (For example, you might want to know where you could best spend your money to improve the accuracy of your ET measurements). •8: If you knew rainfall *exactly*, what would be the uncertainty in ET? So how much does the uncertainty in rainfall *increase* the uncertainty in ET? •9: Conversely, if you knew runoff *exactly*, what would be the uncertainty in ET? How much does the uncertainty in runoff *increase* the uncertainty in ET? •10: The uncertainty in mean rainfall is about 1/3 of the uncertainty in runoff. Why, then, is rainfall such a *trivial* part of the uncertainty in ET? (This should become clear if you inspect the equation you used to calculate the uncertainty in ET.)

Another way to express uncertainty is as the *percentage* uncertainty in the mean (that is, the standard error of the mean, expressed as a percentage of the mean). •11: What is the percent standard error of the mean of rainfall? Of runoff? Of ET? •12: *Why* is the percent standard error of ET so much bigger than the percent standard error of rainfall or runoff? (Hint: to see this, you need to look at what happens to both the numerator and the denominator of the ratio, s.e.(mean)/mean.)

D: Simulating uncertainty in evapotranspiration:

Now, try using Monte Carlo simulation to test your predictions of the uncertainty in evapotranspiration. You have already created 500 10-year averages of rainfall and runoff by random sampling. A 10-year average of ET is just the difference between the 10-year averages of rainfall and runoff (that is, the *average* of the *difference* between rainfall and runoff is equal to the *difference* between the *averages* of rainfall and runoff, since, mathematically speaking, subtraction is a linear operator). So for each of your 500 simulated 10-year averages of rainfall and runoff, you can calculate a 10-year average ET by simple subtraction. So, as Nike says, Just Do It.

•13: Now, what are the mean and standard deviation of the Monte Carlo sample of average ET's? Are they close to what you calculated in #6, above?

•14: What are the upper and lower bounds of the middle 95 percent of your Monte Carlo ET estimates (that is, what are the 97.5th and the 2.5th percentiles of the distribution)? Were your estimates in #7, above, approximately correct? (As with rainfall and runoff, they probably won't be exactly right.)

Try plotting histograms of your Monte Carlo averages of rainfall, runoff, and ET. Try looking at them with the "uniform axes" option, to help illuminate why the percentage uncertainty in ET is so large (as in item #12, above).

E: Estimating the sample size needed to constrain the estimate of evapotranspiration

Now, the 64 thousand dollar question (not a bad way to put it, since the funds at stake would probably be at least that much): *roughly* how many more years of sampling would be needed to estimate evapotranspiration rates more precisely? Specifically, what would it take to constrain the 95% confidence interval for ET to about 50% of the average, that is, 394 mm/yr with an 95% CI of about 194-594 mm/yr? We won't hold your hand all the way through this one, but we'll offer the following hints. (a): You can invert the formula for confidence intervals, to determine what the standard error needs to be, to make the confidence interval as narrow as you want. (b): the value of "t" depends on the sample size, but for any of the sample sizes you might consider here, you can assume t is 2.00, and you can only be off by about three percent. (c): You should assume that you'll have the same number of years of data for both rainfall and runoff; that way the standard errors of rainfall, runoff, and ET will all follow the square-root-of-*n* scaling law predicted by the Central Limit Theorem.

•15: Write down an estimate of the number of years that you'll think you'll need, and explain how you got it.

F: Simulating the effect of more years of sampling

You can use Monte Carlo methods, to check whether the number of years you estimated above is approximately correct. To do this, you change the number of years being averaged in the Monte Carlo formulas that you created earlier. Here, for example, is the formula to average rainfall over 25 years instead of 10:

$$\frac{\sum_{i=1}^{25} rainfall(mm/yr)_{RandomInteger(10)}}{25}$$

Note that you should *not* change the "10" in the subscript, since you want to keep sampling the same 10 data points (you just want to sample them more often). Note also that you need to change the averaging formulas for both rainfall and runoff.

•16: What are the upper and lower bounds of the middle 95% percent of your new Monte Carlo sample? Do they agree with the "target" confidence interval of roughly 400 (200-600) mm/yr that you were aiming for?

That's it! We hope you've seen:

- how uncertainty can propagate through a simple biogeochemical mass balance
- how the small difference between two big numbers can be hard to estimate accurately
- how Monte Carlo methods can be used to simulate error propagation
- how error propagation helps you to pinpoint your most significant sources of uncertainty
- how you can get a rough handle on how extensive a sampling program needs to be, to achieve some desired level of precision

Three closing notes:

Note #1: Your analysis assumes that the means (or central estimates) will not shift as you collect more data. In reality, they won't remain fixed. There are techniques for taking account of the likely shifts in the means, and you'll learn about them in a few weeks.

Note #2: Remember that Monte Carlo methods assume that your sample data, which are replicated and randomly re-sampled over and over again, are completely representative of the full universe of possible data. To the extent that this is not true, your Monte Carlo results will be unrepresentative of the range of future possibilities. For example, if the real-world data are not "stationary"--that is, if their average is drifting through time--they will diverge over time from the Monte Carlo results.

Note #3: *Never* use Monte Carlo methods to make results look more precise than they really are, by inflating the number of data points and thus shrinking the standard error. The reason this is not be legitimate is that we don't actually have 5000 independent data points; instead, we have the same 10 data points, each sampled 500 times. What you've done here is completely cricket, because you weren't claiming that you had more data than you actually did. But the unwary (or the devious) can create some fantastic statistical fairy tales with synthetic data.