Error Propagation in GIS

By M. Sawada

Uncertainty in GIS

In GIS, we often work with numerous data sources. All data has some inherent error because of sampling limitations, limited measurement precision, instrumental errors, et cetera…

An error may denote the difference between a measured value and the "true" value or the estimated uncertainty with respect to a given observation or set of observations. For example, if we know that Ottawa is located at –76 W 45 N, and we look at a map and observe that Ottawa is located at –76 W 44N then we can see that the map is in error of 1^{o}N latitude as to the "true" position of Ottawa.

Error vs. Uncertainty

Errors lead to uncertainty about the value being measured. It is always possible that the value we measure is the "true" value, however, we are usually only relatively certain that a particular value lies within some interval. For example, suppose you measure the length of your desk using a meter stick, which has millimeter divisions. The precision of the instrument is 1 mm, which means that you can only report the length of the desk to the nearest millimeter based on the precision of the measuring instrument. If you measure the desk as being 93 mm, it is possible that the edge of the desk may fall between the 93 mm and 94 mm marks. The ruler you are using presents no gradations finer than 1 mm. This means that nothing can be measured at the sub 1 mm level. Thus, the instrument has an error of ½ x 1 mm or 0.5 mm. So you can say with certainty that the desk's length lies somewhere between

Lower bound = measured value – error = 93 mm – 0.5 mm = 92.5 mm
Upper bound = measured value + error = 93 mm + 0.5 mm = 93.5 mm

So you would say that the desk's length is 93 +/- 0.5 mm in length. You don't know whether 93 is correct, it might be but the measuring instrument cannot tell you this. The ruler can only give you a value, which, with the error, provides you a range within which the true length of the desk will fall. This is what is meant by uncertainty vs. error: if we knew the true error in the measurement we would not be uncertain of the true value of the observation, however, we really don't know the true error of the desk's length, only that its true length is within a given interval, and as such, the +/- .5 mm represents how uncertain we are of that true length.

Types of Errors

Errors can be of two main types, either random or systematic. Random errors occur when repeated measurements, do not agree and the deviations are distributed normally, that is the majority of deviations unbiased. For example, if you have a series of measured deviations, $d_1 \ldots d_n$, they will be distributed evenly about their mean value. Systematic errors, on the other hand, would have a distribution about the mean value that is strongly positive or negative indicating that the individual deviations tend to be errors of similar sign and value. An example of a systematic error would be an instrumental calibration error, or in GIS the differences in measured coordinate values under NAD83 vs. NAD27. The latter is actually a discrepancy rather than an error.

According to Beers (1962), random and systematic errors can be either determinate or indeterminate. Determinate errors are ones that can be evaluated through numerical or logical means, whereas indeterminate errors cannot be evaluated except qualitatively.

Determining the Error

The first step in error analysis is to determine the errors that are present in your spatial and aspatial data. We will not look in any detail at what you have already learnt in 1st year statistics courses, and it is assumed that you are familiar with those statistics and can review them on your own. What will be new to you are the spatial statistics for determining error and statistics for determining error in nominal level data.

Numeric Data (Interval/Ratio)

Errors in data collected on interval and ratio scales are most often determined via standard descriptive statistics, in particular the mean:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

where x_i are measured values ranging from *i…n*. and the Standard Deviation:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} \tag{2}$$

where sigma σ is the standard deviation and $\bar{x}$ is the mean of all observations. The units of σ are in the units of $\bar{x}$, so that if we have $\bar{x}$ meters then the associated uncertainty of $\bar{x} \pm \sigma$ meters.

This is important because when we look at the propagation of uncertainty (errors), the output uncertainty, is a function of the uncertainties associated with the individual uncertainties in the inputs, that is

$$p = f(\delta x_1 .. \delta x_n) \tag{3}$$

which says that *p*, the output uncertainty is a function of the uncertainty of the individual inputs. The δx is read delta-x and represents the small uncertainty in some variable x. Most often, the input uncertainties are going to be given by the standard deviation or equivalent property, and as such, we can say that the standard deviation is approximately equal to the estimated uncertainty in a given quantity, or $\sigma \approx \delta x$.

Propagation of Uncertainty in GIS

Propagation is the process of additive uncertainty, that is to say, that equation (3) says that the output uncertainty, p, is a function of the input uncertainties, or $p = f(\delta x_1 .. \delta x_n)$. In other words, if we have more than one measured quantity in a GIS operation such as additive, subtractive, multiplicative or divisive relations, then there will be error in the output that is a function of the input errors. **Assuming** that errors are ***independent and random*** in one or more variables the following equation can be used to derive general formulae for error propagation:

$$p = \sqrt{\sum_{i=1}^{i=n} \left(\frac{\partial u}{\partial x_i} \right)^2 \delta x_i^2} \tag{4}$$

where, p is the propagated uncertainty and x is some variable in the equation $u = f(x_1 .. x_n)$, and the term $\partial u / \partial x_i$ is the partial derivative of the function $f(x_1 .. x_n)$ or $\frac{\partial}{\partial x_i} f(x_1 .. x_n)$. Consider the following examples to see how error propagation formulae are derived. First we will consider a single mapped variable, $u = f(x)$, whose output error will be p. In other words, the function u is a function of one map variable layer called x. So this is not a multivariable function but we can use (4) to show that this formula still holds:

$$
\begin{aligned}
p &= \sqrt{\left(\frac{\partial u}{\partial x}\right)^2 \delta x^2} \\
&given, \\
u &= f(x) = x \\
\frac{\partial u}{\partial x} &= 1x^{1-1} = 1x^0 = 1 \\
\therefore &= \sqrt{(1)^2 \delta x^2} \\
&= \sqrt{1\delta x^2} \\
&= \sqrt{\delta x^2} \\
&= \delta x
\end{aligned} \tag{5}
$$

Consequently, we can see that for a map, the error at any point will be equal to itself, that is, if the function is of one variable the above equation reduces to the input error. This is only of interest to show that the equation (3) is consistent for a single variable. Now we will derive a formula for error propagation when we have an equation of the form $u = f(Cx)$, where C is a constant rather than a variable:

$$
\begin{aligned}
p &= \sqrt{\left(\frac{\partial u}{\partial x}\right)^2 \delta x^2} \\
&given, \\
u &= f(Cx) = Cx \\
\frac{\partial u}{\partial x} &= 1Cx^{1-1} = Cx^0 = C \\
\therefore &= \sqrt{(C)^2 \delta x^2} \\
&= \sqrt{C^2 \delta x^2}
\end{aligned} \tag{6}
$$

Thus, we can see that if a map with a known uncertainty δx is multiplied by a constant, the error propagates as the root of the squared products of the constant, C multiplied the uncertainty δx.

Rule 1
If your mapped values, x, are multiplied by some constant then the uncertainty in the output is given by: $$p = \sqrt{C^2 \delta x^2} \tag{7}$$

The power of equation (4) is really useful when we have a GIS process that is a function of more than one variable. Next, we can derive some general formulae for cases where our mapping function involves more than one variable, for example, if we add a series of maps together to get some output map, like is often done for calculations of total annual precipitation from monthly total precipitations in a region. Thus, we may have a function of the form $u = f(a,b) = a + b$, and in such a case:

$$p = \sqrt{\sum_{i=1}^{n}\left(\frac{\partial u}{\partial x_i}\right)^2 \delta x_i^2}$$

$$given,$$

$$u = f(x,y) = a + b$$

$$let \;\; x_1 = a$$

$$x_2 = b$$

$$\frac{\partial u}{\partial x_1} = x_1 + x_2 = 1x_1^{1-1} + 0 = 1 + 0 = 1 \qquad (8)$$

$$\frac{\partial u}{\partial x_2} = x_1 + x_2 = 0 + 1x_2^{1-1} = 0 + 1 = 1$$

$$\therefore = \sqrt{(1)^2 \delta x_1^2 + (1)^2 \delta x_2^2}$$

$$= \sqrt{1\delta x_1^2 + 1\delta x_2^2}$$

$$= \sqrt{\delta x_1^2 + \delta x_2^2}$$

Thus, we can see that for addition of a number of variables, their associated error is the square root of the squared sum of the individual errors. This can be easily shown to generalize to any number of additions, such as x + y + z + u + k … + n, so our next error propagation formula is:

Rule 2

If you are adding/subtracting a number of mapped variables together, the uncertainty propagates in the following way for independent random errors:

$$p = \sqrt{\sum_{i=1}^{i=n} \delta x_i^2} \qquad (9)$$

Which is simply to say that the root of the sum of the squared errors of a number of maps is the total error, or $p = \sqrt{\delta a_1^2 + \delta b_2^2 + \delta c_2^2 + \ldots + \delta z_n^2}$, where a, b, c and z are the errors at a point in maps a, b, c, z..

Suffice it to say, that the same can be derived from (4) for difference functions because of the squared partial derivatives in (4). Thus, for functions of the form $u = f(a,b) = a - b$ you may also use equation (9).

Next, we can derive some error propagation rules for cases when we have exponentiation of a single mapped variable. For example, you may have the case where $u = f(x) = Cx^t$ and C is a constant and t is a positive exponent, eg., x^2, x^3, x^4 etc..

$$
\begin{aligned}
p &= \sqrt{\sum_{i=1}^{n}\left(\frac{\partial u}{\partial x_i}\right)^2 \delta x_i^2} \\
given, i &= 1, n = 1, C = 1, and \\
u &= Cx_1^t \\
\frac{\partial u}{\partial x} &= \frac{du}{dx} = tCx^{t-1} \\
\therefore &= \sqrt{(tCx^{t-1})^2 \delta x^2} = \sqrt{t^2 C^2 x^{2(t-1)} \delta x^2} \\
&= \sqrt{t^2 (1)^2 x^{2(t-1)} \delta x^2} \\
&= \sqrt{t^2 x^{2(t-1)} \delta x^2}
\end{aligned}
\tag{10}
$$

Here it is obvious that if our function was $u = f(x) = Cx^t$ and $C = 1$ and $t = 1$ that the result of (10) would work out to $p = \sqrt{(1)^2 x^{2(1-1)} \delta x^2} = \sqrt{1x^{2(0)} \delta x^2} = \sqrt{1x^0 \delta x^2} = \sqrt{1 \times 1 \delta x^2} = \sqrt{\delta x^2}$ which is equivalent to equation (5), that is, if the constant and exponent are equal to the number one, then the input error is equal to the output error. Thus, (10) is only useful if $t > 1$, and if $t = 1$ and $C > 1$ then (10) reduces to (7). Thus, we can say that the following rule can be used for calculation of uncertainty in the exponentiation of a mapped variable:

Rule 3

If a mapped variable with uncertainty, δx, is raised to the power of $t > 1$ then we can use the following equation to determine the propagation of uncertainty:

$$\sqrt{t^2 x^{2(t-1)} \delta x^2} \tag{11}$$

If a constant is involved in the calculation and that constant is $C > 1$ then we can use the following modification of (10), from (9).

$$\sqrt{t^2 C^2 x^{2(t-1)} \delta x^2} \tag{12}$$

Our next example, we can derive an error propagation rule for cases when we have multiplication of a number of independently derived mapped variables such as a map of a variable a and a map of variable b being multiplied. For example, you may have the case where $u = f(a,b) = ab$ and you need to calculate how errors propagate through this process.

$$
\begin{aligned}
p &= \sqrt{\sum_{i=1}^{n}\left(\frac{\partial u}{\partial x_i}\right)^2 \delta x_i^2} \\
given, x_1 &= a, x_2 = b \\
u &= a \times b = ab \\
\frac{\partial u}{\partial x_1} &= \frac{\partial u}{\partial a} = 1a^{1-1}b = 1a^0 b = 1b = b \\
\frac{\partial u}{\partial x_2} &= \frac{\partial u}{\partial b} = a1b^{1-1} = a1b^0 = 1a = a \\
\therefore &= \sqrt{b^2 \delta a^2 + a^2 \delta b^2}
\end{aligned}
\tag{13}
$$

Now, look at an example of three variables, where $u = f(a,b,c) = abc$

$$p = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial u}{\partial x_i} \right)^2 \delta x_i^2}$$

$$given, x_1 = a, x_2 = b, x_3 = c, \quad and \quad u = a \times b = abc$$

$$\frac{\partial u}{\partial x_1} = \frac{\partial u}{\partial a} = 1a^{1-1}bc = 1a^0bc = 1bc = bc, and$$

$$\frac{\partial u}{\partial x_2} = \frac{\partial u}{\partial b} = a1b^{1-1}c = a1b^0c = 1ac = ac \tag{14}$$

$$\frac{\partial u}{\partial x_3} = \frac{\partial u}{\partial c} = ab1c^{1-1} = ab1c^0 = ab1 = ab$$

$$p = \sqrt{\left(\frac{\partial u}{\partial a} \right)^2 \delta a^2 + \left(\frac{\partial u}{\partial b} \right)^2 \delta b^2 + \left(\frac{\partial u}{\partial c} \right)^2 \delta c^2} = \sqrt{(bc)^2 \delta a^2 + (ac)^2 \delta b^2 + (ab)^2 \delta c^2}$$

You can now see the general pattern when we are dealing with a function that is a product. This may be unwieldy and tedious to figure out each time we need to determine the error propagation in a product if one is unfamiliar with differential calculus. However, using a basic algebraic rule that any square root of a squared number divided by the square root of another squared number equals the square root of the two squared number divided by one another, e.g.,

$$q = \frac{\sqrt{h^2}}{\sqrt{g^2}} = \sqrt{\frac{h^2}{g^2}} \tag{15}$$

we can show that a more simple rule for error propagation can be derived using (4), that will be easier to apply in most situations. For example,

$$\frac{p}{\sqrt{u^2}} = \frac{\sqrt{\sum_{i=1}^{n}\left(\frac{\partial u}{\partial x_i}\right)^2 \delta x_i^2}}{\sqrt{u^2}}$$

$$\frac{p}{\sqrt{u^2}} = \frac{\sqrt{\left(\frac{\partial u}{\partial a}\right)^2 \delta a^2 + \left(\frac{\partial u}{\partial b}\right)^2 \delta b^2 + \left(\frac{\partial u}{\partial c}\right)^2 \delta c^2}}{\sqrt{u^2}}$$

$$= \frac{\sqrt{(bc)^2 \delta a^2 + (ac)^2 \delta b^2 + (ab)^2 \delta c^2}}{\sqrt{u^2}}$$

$$where, u = abc$$

$$= \frac{\sqrt{(bc)^2 \delta a^2 + (ac)^2 \delta b^2 + (ab)^2 \delta c^2}}{\sqrt{abc^2}}$$

$$= \sqrt{\frac{(bc)^2 \delta a^2 + (ac)^2 \delta b^2 + (ab)^2 \delta c^2}{(abc)^2}}$$

$$= \sqrt{\frac{(bc)^2 \delta a^2}{(abc)^2} + \frac{(ac)^2 \delta b^2}{(abc)^2} + \frac{(ab)^2 \delta c^2}{(abc)^2}}$$

$$= \sqrt{\frac{b^2c^2\delta a^2}{a^2b^2c^2} + \frac{a^2c^2\delta b^2}{a^2b^2c^2} + \frac{a^2b^2\delta c^2}{a^2b^2c^2}} \quad (16)$$

$$= \sqrt{\frac{\cancel{b}^2\cancel{c}^2\delta a^2}{a^2\cancel{b}^2\cancel{c}^2} + \frac{\cancel{a}^2\cancel{c}^2\delta b^2}{\cancel{a}^2b^2\cancel{c}^2} + \frac{\cancel{a}^2\cancel{b}^2\delta c^2}{\cancel{a}^2\cancel{b}^2c^2}}$$

$$\therefore \frac{p}{\sqrt{u^2}} = \sqrt{\frac{\delta a^2}{a^2} + \frac{\delta b^2}{b^2} + \frac{\delta c^2}{c^2}}$$

Which shows that by dividing $\sqrt{u^2}$, strictly here defined as the absolute value of u or $|u|$, we have shown that we can derive a general rule for error propagation when a series of mapped values are multiplied together, however, because we have divided $p/\sqrt{u^2}$ we must still solve for p in the following manner, given some number k,

$$\frac{p}{\sqrt{u^2}} = k \quad (17)$$

$$k = p\sqrt{u^2}$$

Consequently, in a moment we will have a general rule, and this rule can be applied in cases of division and multiplication, however, only the product case is derived here since the division case is very similar, and consequently it would only be a waste of space to do it as well. Therefore, as a general rule:

Rule 4

If we multiply or divide two or more mapped variables, say variables a, b, c, that are a product of the function $u = f(a,b,c) = abc$, the error propagates in the following manner:

$$p = \sqrt{(abc)^2}\sqrt{\left(\frac{\delta a}{a}\right)^2 + \left(\frac{\delta b}{b}\right)^2 + \left(\frac{\delta c}{c}\right)^2 + \ldots + \left(\frac{\delta n}{n}\right)^2} \quad (18)$$

or in general, for a function $u = f(x_i..x_n) = \prod_{i=1}^{i=n} x_i$, of some set of mapped variables that are multiplied* by each other

$$p = \sqrt{u^2}\sqrt{\sum_{i=1}^{i=n}\left(\frac{\delta x_i}{x_i}\right)^2} \qquad (19)$$

Note: The $\prod_{i=1}^{i=n} x_i$ means the product of a set of numbers $x_i..x_n$ and the $\prod$ symbol means multiply, e.g., $x_1 * x_2 * x_3 * ... * x_n$, just like the sigma symbol $\sum$ means to add a series of numbers, $x_1 + x_2 + x_3 + ... + x_n$.

We will not look at any derivations for the case of possible correlation between errors in multivariate equations as this is slightly more complicated and even general forms become unwieldy after more than two variables. References, however, for this can be found in Taylor (1982, p.177) and Beers (1962, p.31). From these references, it can be shown that for a function of the form $u = f(a,b)$, where there is correlation between the errors δa and δb is $r_{\delta a \delta b}$, where r is the coefficient of correlation that the propagated error p, can be larger or equal to the errors that are obtained from equation (4) but never larger than the sum of the individual errors.

Rule 5

If you have a function of two variables, $u = f(a,b)$ where the correlation between the errors of δa and δb is $r_{\delta a \delta b}$, then we can determine the propagation of error using the following modification to equation (4):

$$p = \sqrt{\left(\frac{\partial u}{\partial a}\right)^2 \delta a^2 + \left(\frac{\partial u}{\partial b}\right)^2 \delta b^2 + 2\left(\frac{\partial u}{\partial a}\right)\left(\frac{\partial u}{\partial b}\right)\delta a \delta b r_{\delta a \delta b}} \qquad (20)$$

Note: This equation can be extended for more than two variables as shown in Taylor (1982) and Beers (1962) as well as in Gerard et al. (1989).

Taylor (1982) states that the role of covariance "in the discussion of error propagation is purely theoretical; and, in fact, the concept of the covariance does not usually play a practical role in error propagation". Thus, for most purposes, we are dealing with non-correlated error terms and can avoid using Rule 5, however, if you have evidence that errors are correlated you should use this rule. I will provide an example calculation in the examples section to facilitate your learning.

Examples

We have now covered the four main rules for error propagation that you are likely going to encounter in using data from various sources in GIS. Now we can look at a number of example applications in GIS.

Spatial Database Errors

⇒ ***Example 1***

You have a series of maps, one from a 1:30,000 map (map_1), one from a 1:50,000 map (map_2), and one from a 1:250,000 map (map_2). Assuming that the smallest discernable feature on each is .5 mm, determine the total positional error if these maps are used in a single project.

This is a common problem in GIS, spatial data are often combined from different source scales and therefore different positional uncertainties are mixed together in a final map or analysis. We will assume that these maps were digitized with complete accuracy for their reference scales within the inherent map resolution. We can determine the individual contributions to the total output map uncertainty by pretending that combining the maps is equivalent to undertaking an additive process. As such, our mapping function u , can be looked at as:

$$u = f(map_1, map_2, map_3) = map_1 + map_2 + map_3 \tag{21}$$

where the (+) signs represent more of a logical union than an arithmetic one. Consequently, our first step is to determine the errors on these maps,

$$\begin{aligned} &map_1 \pm \delta map_1 \\ &map_2 \pm \delta map_2 \\ &map_3 \pm \delta map_3 \end{aligned} \tag{22}$$

even though we have not been explicitly given error terms. First we assume that the person who digitized these maps was able to exactly follow the lines on each map with the digitizing puck. By following the lines, I mean that the digitizer cross-hair did not accidentally go outside of any of the lines that s/he was following with the puck. Consequently, we only need to concern ourselves with the error's inherent in the maps themselves based on their scale and the effective resolution at this scale. Because the smallest feature or thinnest line on the maps is .5 mm then, depending on the scale of the map, this 0.5 mm (e.g., a small circle of 0.5 mm) will on the ground surface actually represent a circle of a different size. Remember that map scale is given by:

$$\text{map scale} = \frac{\text{distance on ground}}{\text{distance on map}} \tag{23}$$

for example, if the distance from your house to the school is 20 km and on a map this measures 10 cm then the map scale would be from (23):

$$\text{map scale} = \frac{\text{distance on ground}}{\text{distance on map}} = \frac{20000\text{m}}{.1\text{m}} = 200{,}000 \tag{24}$$

On such a map the map scale would be expressed as 1:200,000, meaning that 1 mm on the map represents 200,000 mm on the actual earth. If on such a 1:200,000 scale map, the smallest map feature we can see with the naked eye is 0.5 mm in size[1], then that feature of 0.5 mm in size is actually covering a 'real-world' length of 0.5 x 200,000 mm or 100,000 mm, or 100 meters. Thus, any feature taken from the 1:200,000 scale map is really within +/- 50 meters of it's actual location (remember the example with the ruler?). In other words, the gradation of this map, or its resolution, can only measure to the nearest 100 meters. In this sense you can think of the map as an instrument with an inherent positional error for any feature on the map. From this idea, we can attain an idea of the positional error in our example above, namely,

$$\begin{aligned} \delta map_1 &= 30000 \cdot .5mm = 15000/2 = 7500mm = 7.5km \\ \delta map_2 &= 50000 \cdot 0.5mm = 25000/2 = 12500mm = 12.5km \\ \delta map_3 &= 250000 \cdot 0.5mm = 125000/2 = 62500mm = 62.5km \end{aligned} \tag{25}$$

[1] Note that an alternative definition for map resolution is the minimum distance at which two map features can be distinguished as being separate entities.

In this step we divide by two since a given location or point can be $+ n\ km$ from the center of the line being digitized or $- n\ km$ from the same centerline. Now, we can apply our Rule 2 to determine the error in our output map based on the input maps,

$$\begin{aligned} p &= \sqrt{\delta map_1^2 + \delta map_2^2 + \delta map_3^2} \\ &= \sqrt{7.5^2 + 12.5^2 + 62.5^2} \\ &= \sqrt{56.25 + 156.25 + 3906.25} \\ &= \sqrt{4118.75} = 64.2km \end{aligned} \tag{26}$$

Therefore, our output project will have an error of at least 128 km. Also notice here that the largest contribution to the output error is due to the third map, which had the largest error. Without this map the error would be much smaller, $\sqrt{56.25+156.25} = \sqrt{212.5} = 14.6km$. This makes it clear that the project could be quickly improved by utilizing a map at a larger scale than that of map 3.

This assumes that by following the lines perfectly there was random variation within the line boundaries by the digitizer and that this variation is normally distributed. The actual error would probably be somewhat larger than this, as shown in the next example.

⇒ **Example 2**

Suppose you are given a digital data file from a mapping agency with the following information provided in the metadata:

Map 1: Current Data File	**Tested 10.4 meters horizontal accuracy at 95% confidence level. Scale: 1:50,000**
Map 2: Control dataset	**1:5000 planimetric map Accuracy of Control dataset: 2.5 m**

Determine the likely error in Map 1.

First the NSSDA statistic is computed by multiplying the RMS error by 1.7308, therefore, we first need to determine the RMS error by dividing the NSSDA reported accuracy by the NSSDA conversion factor, e.g.,

$$RMS = \frac{NSSDA}{1.7308} = \frac{10.4m}{1.7308} = 6m \tag{27}$$

Therefore, our RMS error for the digital map which we will call map 1 is $\delta map_1 = 6m$. However, we need to take into account that the map that it was tested against had an accuracy of $\delta map_2 = 2.5m$. Our digital map is going to have an accuracy less than that reported since the control points used to determine the map accuracy clearly have uncertainties themselves. Again, this can be seen as an additive problem, and as such we can use Rule 2:

$$\begin{aligned} p &= \sqrt{\delta map_1^2 + \delta map_2^2} \\ &= \sqrt{6^2 + 2.5^2} = \sqrt{36+5} = \sqrt{41} = 6.4m \end{aligned} \tag{28}$$

Therefore, a better estimate of the uncertainty in our current map is $\delta map_1 \pm 6.4m$

Attribute Errors: Uncorrelated Errors

⇒ **Example 3**

For example, for a function of two variables we can look at how correlated errors can be equal to or greater than those estimated by application of (4). Consider, two attributes measured within a given drainage basin, namely rainfall, $E = 200m^3$, and runoff, $I = 140m^3$. For sake of argument, we may say that rainfall and runoff[1] are related in some way and that the error in the rainfall estimate, $\delta E = 10m^3$, and the error in the runoff estimate which is $\delta I = 14m^3$. We wish to estimate the proportion of rainfall that has runoff, that is to say, the amount of rainfall that contributed to the runoff within the basin. Therefore we have a function of the form $u = f(I,E) = I/E$.

[1] You can easily imagine other variables, say of a socioeconomic importance like unemployment and income or the like and substitute these for rainfall and runoff.

We can see that the function $u = f(I,E) = I/E$ is a quotient of two numbers and therefore we can apply Rule 4 and Equation 18.

$$p = \sqrt{(abc)^2}\sqrt{\left(\frac{\delta a}{a}\right)^2 + \left(\frac{\delta b}{b}\right)^2 + \left(\frac{\delta c}{c}\right)^2 + \ldots + \left(\frac{\delta n}{n}\right)^2} \qquad (29)$$

Which becomes in our case:

$$\begin{aligned} p &= \sqrt{(I/E)^2}\sqrt{\left(\frac{\delta I}{I}\right)^2 + \left(\frac{\delta E}{E}\right)^2} \\ &= \sqrt{(140/200)^2}\sqrt{\left(\frac{14}{140}\right)^2 + \left(\frac{10}{200}\right)^2} \\ &= \sqrt{(0.7)^2}\sqrt{(.1)^2 + (.05)^2} \\ &= \sqrt{0.49}\sqrt{.01 + 0.0025} \\ &= \sqrt{0.49}\sqrt{.0125} \\ &= 0.7 \times .1112 = 0.0785 \end{aligned} \qquad (30)$$

Consequently, if the errors $\delta E = 10m^3$ and $\delta I = 14m^3$ are uncorrelated the propagated error $p \approx 0.07$, therefore our estimate is $u = f(I,E) = I/E = 0.7 \pm 0.0785$. Here the units' m^3 cancelled out and we have a proportion, which if multiplied by 100 would be 70% +/- 7.85%.

⇒ **Example 4**

In a certain part of Africa, there is an empirical relation between elevation Em and temperature t^oC that is based on physical laws called the dry adiabatic lapse rate. This takes the form of $t^oC = f(E) = 0.0098\left({}^oC/_m\right)Em$. In topoclimatic modeling, a person takes a digital elevation model (DEM), which shows the elevation at every point on a portion of the earth's surface, and multiplies this value by 0.001 to get an estimate of temperature. This can be very useful when there are no physical temperature observations in an area and temperature must be estimated for some research purposes. Consider a particular point in space where the elevation of a given polygon is 600 m with an elevation uncertainty $\delta E = 12m$. Determine the level of error associated with the temperature estimate from this equation.

First, we can see that only one variable is involved, namely, elevation, E. This variable is multiplied by a constant, when a variable is multiplied by a constant we can use Rule 1 and equation (7):

$$\begin{aligned} p &= \sqrt{C^2\delta x^2}, where \\ C &= 0.0098\,{}^oC/_m, and \\ \delta E &= 12m, thus \\ p &= \sqrt{\left(0.0098\,{}^oC/_{\cancel{m}}\right)^2 (12\cancel{m})^2} \\ &= \sqrt{0.00009604\,{}^oC^2/_{\cancel{m}^2} \cdot 144\cancel{m}^2} \\ &= \sqrt{0.01383^oC^2} \\ \therefore\ p &= 0.1176^oC \end{aligned} \qquad (31)$$

Consequently, we can see that the error in the temperature estimate for this area will be $t^oC = f(E) = 0.0098(600) = 5.88 \pm 0.12^oC$.

⇒ **Example 5** **As a brief example of correlated errors, for sake of argument, consider that we have a mapped function of two variables $u = f(I,E) = E + I$ such that their coefficient of correlation is $r_{\delta I \delta E} = 0.3$ and the error terms are given as $\delta E = 10 units$ and $\delta I = 14 units$. In this case $E = 700 units, and, I = 400 units$. Determine the error in the output for both the case when these variables are correlated and uncorrelated.**

First, in the case of uncorrelated errors we can see that this function is a sum of two variables and therefore can apply Rule 2 and equation (9):

$$\begin{aligned}
p &= \sqrt{\delta a_1^2 + \delta b_2^2 + \delta c_2^2 + \ldots + \delta z_n^2} \\
where, \delta I &= 14 \\
\delta E &= 10 \\
p &= \sqrt{\delta I^2 + \delta E^2} \\
&= \sqrt{14^2 + 10^2} \\
&= \sqrt{296} \\
\therefore\ p &= 17.2 units
\end{aligned} \tag{32}$$

However, if we have a case where the errors are correlated, with $r_{\delta I \delta E} = 0.3$, then we need to use Rule 5 and equation 20,

$$\begin{aligned}
& p = \sqrt{\left(\frac{\partial u}{\partial I}\right)^2 \delta I^2 + \left(\frac{\partial u}{\partial E}\right)^2 \delta E^2 + 2\left(\frac{\partial u}{\partial I}\right)\left(\frac{\partial u}{\partial E}\right)\delta I \delta E r_{\delta I \delta E}} \\
& where, u = I + E, r_{\delta I \delta E} = 0.3, and \\
& \frac{\partial u}{\partial I} = 1, and \\
& \frac{\partial u}{\partial E} = 1 \\
& p = \sqrt{(1)^2 \delta I^2 + (1)^2 \delta E^2 + 2(1)(1)\delta I \delta E r_{\delta I \delta E}} \\
& p = \sqrt{(1)^2 14^2 + (1)^2 10^2 + 2(1)(1)(14)(10)(0.3)} \\
& p = \sqrt{(1)(196) + (1)(100) + 2(1)(1)(14)(10)(0.3)} \\
& p = \sqrt{196 + 100 + 84} \\
& p = \sqrt{380} \\
& \therefore\ p = 19.49 units
\end{aligned} \tag{33}$$

Thus, we can see that when the errors are correlated with $r_{\delta I \delta E} = 0.3$, that the error is slightly larger, 19.49 units, than in the uncorrelated case which gave 17.2 units. It can be shown from this that if the correlation $r_{\delta I \delta E} = 1$, that the estimated error is equal to the sum of the input errors.

1. You have a polygon map whose values are being exponentiated, that is to say your map function is $u = f(a) = a^3$. Determine the output error in u of this function if the input error is $\delta a = \pm 2$.

2. You are adding a bunch of maps together, such that the function being used is $u = f(a,b,c,d,e) = a + b + c + d + e$. The errors involved for each variable are

$$\begin{aligned}\delta a &= \pm 3\\ \delta b &= \pm 4\\ \delta c &= \pm 2\\ \delta d &= \pm 6\\ \delta e &= \pm 3\end{aligned}\ .$$

Determine the value and output error of u.

3. For the same errors in question 1, suppose the function is $u = f(a,b,c,d,e) = a \cdot b \cdot c \cdot d \cdot e$. Determine the value and output of the error associated with u.

4. For the same errors in question 1, suppose the function is $u = f(a,b,c,d,e) = a - b - c - d - e$. Determine the value and output of the error associated with u.

5. From data used in Exercise 2a. Create a chloropleth map of population density in Ottawa-Carleton. You are told that the error in population is proportional to the population estimate, such that the relative error is 2%. The uncertainty in the polygon areas is also proportional to the area estimate and is 1%.

6. From exercise 2b, you calculated the RMS error for a set of control points and test points. If you are told that the control points had an error of +/- 2 m, then what is the propagated error for your test dataset?

References

Beers, Y. 1962. Introduction to the Theory of Error. Addison Wesley, Reading, Massachusetts.

Burrough, P.A. and Rachael A. McDonnell. 1998. Principles of Geographical Information Systems: Spatial information systems and geostatistics. Oxford University Press, New York.

Heuvelink, G.B., P.A. Burrough, and A. Stein. Propagation of errors in spatial modelling with GIS. International Journal of Geographical Information Systems 3(4):303-322.

Taylor, J.R. 1982. An Introduction to Error Analysis: The study of uncertainties in physical measurements. University Science Books, Mill Valley, California.

Thapa, K., and J.Bossler. 1992. Accuracy of spatial data used in geographic information systems. Photogrammetric Engineering and Remote Sensing 58(6):835-841.

Veregin, H., and D.P.Lanter. 1992. A research paradigm for propagating error in layer-based GIS. Photogrammetric Engineering and Remote Sensing 58(6):825-833.