# Uncertainty and Variability in Geology

Another technical note in the Art of Science series by Dan Cornford of IGI Ltd.

Uncertainty and variability are often confused. The former represents a state of our minds, the later a state of nature. This short note explains the key concepts to understand the distinction between the two concepts, and how they relate to each other. Begg et al. (2014) provides a complete overview which introduces the key concepts in some detail. This note provides a focussed view on the relation between uncertainty and variability.
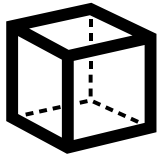
To start, imagine we have a cube of rock, let's say $1m^3$ for somewhere in Earth's crust. For now, let's call this our sample, the population being all the rock in the Earth's crust. Imagine we are interested in the Total Organic Carbon (by mass) in the rock volume (as a proportion of the total rock mass). This is the classic TOC measurement, reported as a %wt.

Uncertainty

We can reasonably ask the question before we take any measurements, what is the TOC of the $1m^3$ of rock?

There are several ways we could answer this. We know TOC must be between 0 – 100%, so we could simply say we don't know, but it is somewhere between zero and one hundred percent.
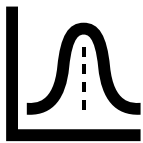
Another reasonable answer would be more refined. We have some idea of the distribution of TOC in rocks of the Earth. Most sedimentary rocks will have TOC in the range 0.1 to 12 %wt, and even coals rarely exceed 85 %wt TOC. Igneous rocks will typically have TOC of 0 %wt. We could define our distribution of TOC using a probability distribution of some form, or a non-parametric approach such as percentiles, e.g. our p10, p50 and p90 estimates.

I'd put p10 = 0 %wt, p50 = 0.8 %wt, p90 = 2 %wt based on what I know. Your judgements are likely to be different.

Now I would note there is no randomness in the TOC of the $1m^3$. There is a value. We do not know it. This is uncertainty due to lack of knowledge – **epistemic uncertainty**.

Your epistemic uncertainty can be different to mine. It depends on what you know, or rather believe you know. Only nature 'knows' the true value of TOC in that $1m^3$ of rock.

I suppose at this point most of you will be wanting to ask more questions, where is the rock from, what is the lithology, the stratigraphy, colour, the mass, etc.

If I now told you that the rock came from the Kimmeridge Clay Formation in the UK sector of the Northern North Sea, most people would revise their judgements on the TOC of the $1m^3$.

The rock has not changed, the TOC is no different, but our beliefs about it will change.

There is a subjective uncertainty, but this is a property of our minds, not of nature.

Variability

Now let's change frame of reference. Let's assume the 1m³ block is our population. Imagine we slice this very carefully into one million 1cm³ samples. It seems very unlikely that the TOC will be spatially homogenous within this block. Vertical variation of TOC is typical, but it is likely that there is also lateral variation.

If we perfectly[1] analyse all one million 1cm³ samples we can calculate the statistics of these, let's say these have a mean of 1.5 %wt, and a sample (= population) standard deviation of 0.5 %wt. At this point we can state the mean TOC of the 1m³ is 1.5 %wt, and the standard error on that estimate is zero. That is, we do not have uncertainty, simply variability of TOC within the block.

There is a 'true' statistical variability (distribution), and it is a property of nature.

When considering real world properties, variability is typically associated with spatial or temporal variation.

Sampling, variability, and uncertainty

Imagine we pick ten 1cm³ blocks at random (or we could use an experimental design strategy to select the blocks – that will be the topic of a future note). Assume we can again measure the TOC of each 1cm³ block perfectly.

Histogram of TOC.Leco
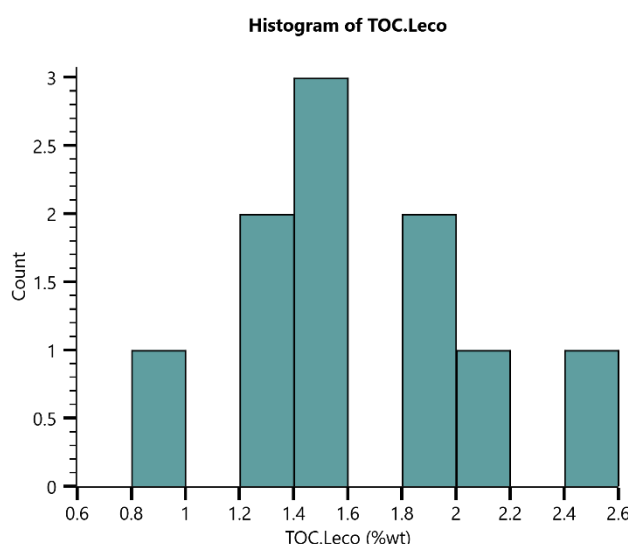
The 10 values we measure are: 1.2, 1.3, 1.5, 1.9, 2.4, 1.9, 1.4, 1.5, 0.9, 2.1 %wt. The results vary, due to the variability of TOC within the 1m³ block.

We can calculate the mean (1.61 %wt) and the sample standard deviation (0.46 %wt) which characterise the **variability** of TOC within the block, *based on the finite sample*.

Now let's think about the TOC of the whole 1m³ of rock. The expected value of the mean would be 1.61 %wt, and the estimator of the standard error (of the mean) would be 0.14 %wt. Recall the standard error is the sample standard deviation divided by the square root of the sample size. Note we do not need to make strong distributional assumptions here thanks to the Central Limit Theorem, which tells us that the distribution of the mean of any variable, sampled from any distribution will tend to a Gaussian (or Normal) distribution as the sample size increases.

So, is this variability, or uncertainty in the TOC of the 1m³? As framed currently it is uncertainty, induced by having an incomplete sample of the population, which itself is variable. The variability established from the sample informs our uncertainty judgements about the population.

The relation between variability and uncertainty

One source of confusion between variability and uncertainty is that we use the same language to describe both. We talk about mean and standard deviation to describe both the statistical variability

---

[1] I will assume a perfect measurement in one with sufficient accuracy that measurement bias and uncertainty is not practically relevant

of the samples, and our uncertainty about a quantity[2]. It is important to note we are using the same language, indeed concepts, to describe quite different things.

The confusion is exacerbated by the fact that we can use information about variability to inform uncertainty judgements. And this can be done rigorously if we take care to define things carefully.

When considering our uncertainty about a value, we can often use measurements, and the variability of these measurements to help guide our judgements, or beliefs. In the previous section we considered how a small sample of ten measurements of TOC could be used to estimate the variability of TOC within the block. We showed then how this could be used (with assumptions) to estimate our uncertainty about the mean TOC in the whole $1m^3$ volume. Depending on the assumptions we make (our judgements), different people might still obtain different uncertainties. As more samples are taken, it is likely the weight of the evidence from the measurements will reduce the impact of our prior assumptions and our judgements will converge.

Variability can equal uncertainty?

Now let's consider another case where knowledge of variability can inform our uncertainty judgements. Let's assume that we know the *population* mean and standard deviation of TOC in the $1cm^3$ samples to sufficient accuracy that we have no uncertainty about these: mean = 1.5 %wt, standard deviation = 0.5 %wt. This is an idealised case, but now image we pick a $1cm^3$ block at random from the $1m^3$ block. What is the TOC of this $1cm^3$? Since we know the population variability and we know the sample was drawn at random, the uncertainty in the TOC value will be equal to the population variability.

We should not, however, confuse the numeric equality of the uncertainty to the variability in this specific case to infer they are the same thing. Variability can inform uncertainty judgements.

Randomness
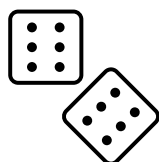
There are very few natural examples of randomness. True randomness induced uncertainty that is not epistemic, but rather aleatory. The key distinction is that aleatory uncertainty cannot be reduced by obtaining more knowledge. It is a state of nature, not of our minds. The most famous, canonical, example of randomness is the timing of radioactive decay of isotopes. This process is believed to be truly random, but outside of radioactive decay, above the quantum scale, I would claim there are no truly random processes in geoscience. There are chaotic processes, which make prediction practically impossible. For example, turbulent flow is in essence unpredictable beyond a very short timescale and can only be described in terms of its statistics. But that is another story …

For most practical applications randomness is not an important factor. Variability is. Uncertainty (of an epistemic form) is. Neither require randomness.

The importance of precision

You have probably all been told to beware of false precision. The TOC of the sample was 1.631307 %wt is a precise answer, but that precision has two practical issues. First, measurement error is likely to be significantly larger than the implied accuracy. Secondly, it is not practically relevant to worry about TOC to 6 decimal places of precision.

---

[2] Note summary statistics / moments are just one way to represent variability or uncertainty. We can use a range of other methods including probability distributions, non-parametric methods such as percentiles / histogram, or discrete samples / realisations.

But not all precision is redundant. When considering problems of communication, and the confusion between variability and uncertainty, one of the most significant causes is a lack of precision in describing what we are interested in learning about or making judgements about.

It is important to precisely, and clearly, define what quantity we are considering. This will help us understand whether we should be considering variability or uncertainty, and whether we can use variability to inform our uncertainty judgements. Be semantically precise, and numerically accurate.

Summary

Epistemic uncertainty is a state of our minds. Variability is a state of nature.

Uncertainty can be reduced by measurements. Variability can be estimated by measurements but does not change (for a stationary process).

Uncertainty can be informed by variability.

Managing uncertainty is essential to the scientific method and plays a key role in the geosciences. Working with uncertainty can be challenging, and has not been widely, or well, taught in the geosciences. However, once you embrace the subjective nature of uncertainty it can open doors to a richer and more honest description of the problems we tackle and in particular facilitate rational decision making and risk management.

References

Begg, S. H., Welsh, M. B., and Bratvold, R. B., 2014. "Uncertainty vs. Variability: What's the Difference and Why is it Important?." Paper presented at the SPE Hydrocarbon Economics and Evaluation Symposium, Houston, Texas, May 2014. doi: https://doi.org/10.2118/169850-MS