Known and unknown unknowns

Uncertainty estimation in satellite remote sensing data

A.C. Povey and R.G. Grainger

National Centre for Earth Observation, University of Oxford (✉adam.povey@physics.ox.ac.uk)

The importance of uncertainty

All measurements are subject to error, the difference between the value obtained and the theoretical true value (a.k.a. the measurand). The uncertainty on a measurement describes the expected magnitude of the error by characterising the distribution of error that would be found if the measurement was infinitely repeated. These concepts are sketched in Fig. 1. Uncertainty is a vital component of data as it provides

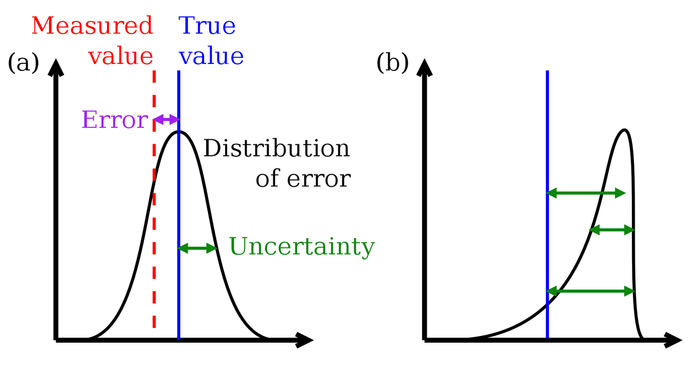

Fig. 1: The difference between uncertainty and error.

- a means to efficiently and consistently communicate the strengths and limitations of data to users, and
- a metric with which to compare and consolidate different estimates of a measurand.

Standardised methods for uncertainty estimation can be insufficient for satellite remote sensing data as they assume a well-constrained measurement where the sources of error are established — *known, quantifiable unknowns*. The dominance of systematic errors in satellite remote sensing data introduce *known, unquantifiable unknowns* (such as the impact of cloud filtering) and *unknown unknowns* (such as variability on scales smaller than that observed).

It is important that uncertainty is handled in a manner appropriate to the information available. This poster briefly summarises the discussions of a paper of the same name currently under discussion in AMTD (doi:10.5194/amtd-8-8509-2015).

Formal definition

As defined by [1], uncertainty is a "parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand." A technique for calculating uncertainty should demonstrate,

universality all manners of observation can apply the techniques to calculate their uncertainty;

internal consistency the calculation of uncertainty requires no additional information;

transferability it must be of use to a data user.

Complications with satellite data

Conventional estimations of uncertainty (with a standard deviation and error propagation) are useful with satellite data but poorly represent systematic errors. It is unclear if such errors are distributed symmetrically, such that the emphasis on traditional techniques may contribute to many analysts neglecting important systematic errors as they cannot be quantified with confidence.

The basis chosen to describe a system impacts the expression of uncertainty. Consider retrieving cloud top pressure with an infrared radiometer. Transforming the observed radiance into the cloud top's radiating temperature is non-linear, such that a symmetric distribution of random error in radiance is not symmetric when considering temperature, as sketched in Fig. 2(b). Converting from temperature to pressure further distorts the distribution of error (and introduces additional uncertainty).

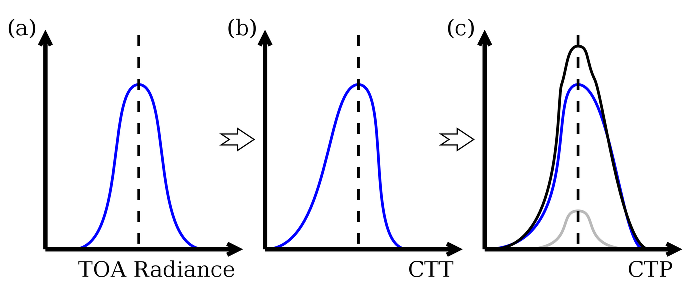

Fig. 2: Distorting the distribution of error when measuring cloud top radiance, temperature or pressure.

Evaluating errors

In addition to describing errors as 'random' or 'systematic', it can be more useful to classify them according to the source of the error, such as:

measurement statistical variation in the measurand and fluctuations in the detector;

parameter errors propagated from auxiliary data used in the retrieval;

approximation explicit simplifications in the formulation of the forward model;

system differences between various sensible descriptions of the environment and reality;

resolution variability at unobserved scales.

Measurement and parameter errors are generally well represented by traditional techniques. These are useful but only describe one aspect of the uncertainty — the 'unknowns' that are known and quantifiable. Approximation and system errors represent the inability of the analysis to describe the environment observed and are the dominant source of error in most passive satellite remote sensing data. Analysts are aware of these 'unknowns', such as the representation of the surface's bi-directional reflectance, but cannot quantify them with a standard deviation. Even well-constrained analyses will be affected by system errors resulting from quality control, cloud filtering being the most common. Resolution errors describe the disconnect between what occurs in nature and the means by which it is observed, primarily resulting from the instrument's sampling.

The difficulty with the last three categories is that they can be highly non-linear (they depend on the state observed and the accuracy with which it can be described). Propagation of errors assumes that the equations used are accurate and that errors affect them linearly. Uncertainties currently reported with satellite remote sensing data neither represent the actual (non-linear) distribution of errors nor the full range of information known about the errors.

Comparing retrievals

Retrievals can only be compared over some subset of the possible state vectors (e.g. a SST product compared to ship-based measurements will only encapsulate the variation in SST over major shipping lanes rather than globally). As systematic errors are circumstantial, the validation only samples the complete distribution — just as the definition of a measurand frames how it can be understood, the scope of a validation frames the understanding of systematic errors.

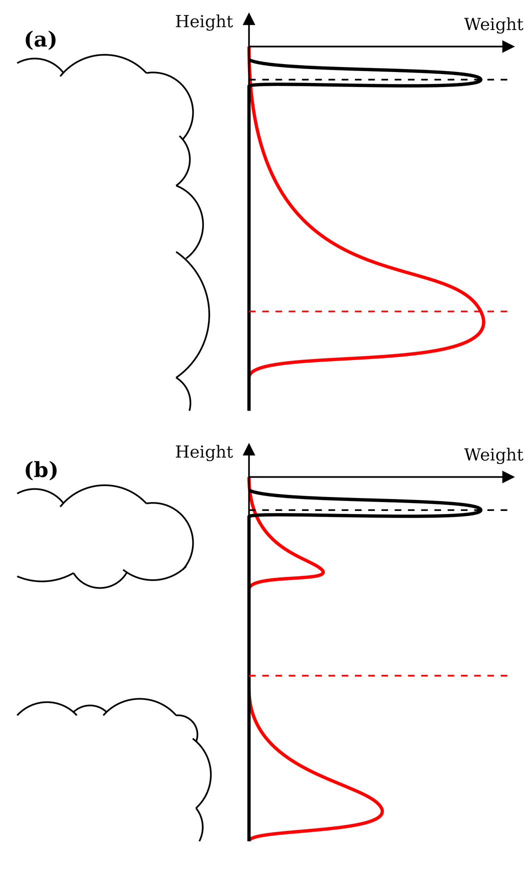

Fig. 3: Weighting functions of CTH for an infrared radiometer (red) and lidar (black), with dashed lines denoting the value retrieved.

Towards the aim of repeatability, validation should be performed in a manner that, if an additional data source was introduced, the conclusions should not substantially change. [2] noted that this does not apply to retrievals with differing averaging kernels and developed a formalism to compensate. Consider cloud top height (CTH). A radiometer measures an average of the cloud's temperature profile weighted by the probability that a photon from that level is observed. That weight is known as the weighting function (Fig. 3).

A lidar measures the number of particles in its beam, resulting in sharper weighting function. A comparison of these products will find radiometer CTH are consistently lower than those from the lidar. To properly validate the satellite against the lidar, it is necessary to use the satellite's weighting function to calculate an 'effective cloud radiating height' from the lidar profile. If the averaging kernel is not calculated, it is not possible to rigorously compare the data from different sensors, even from the same algorithm.

Ensemble techniques

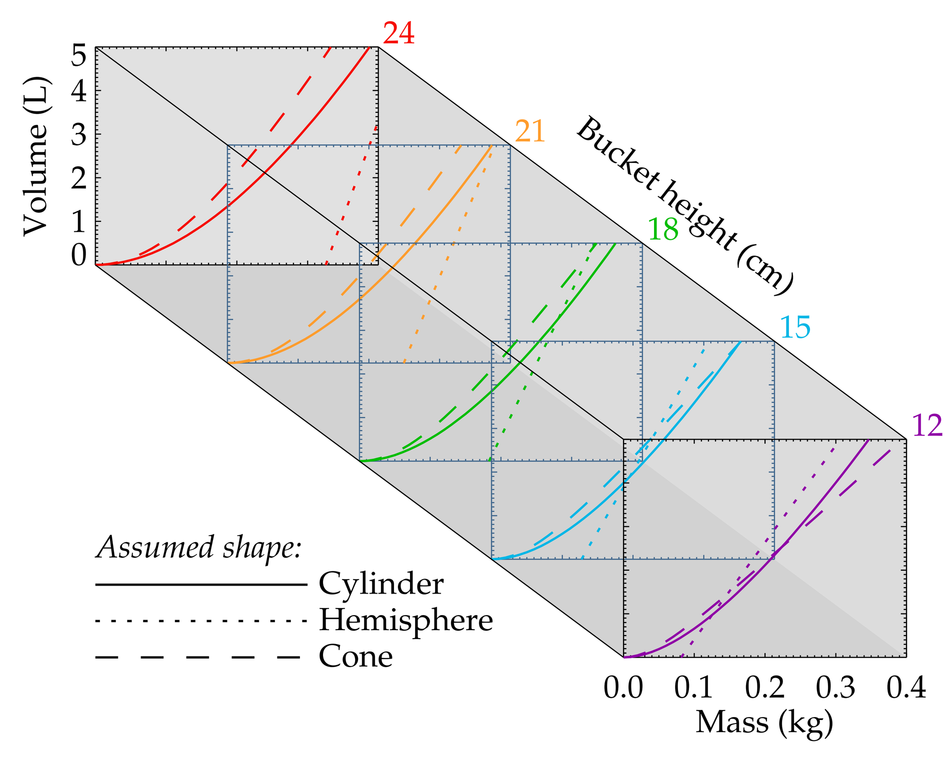

Fig. 4: An ensemble of forward models for the volume of a bucket as a function of its mass. The impact of the bucket's height, which is assumed, is shown over five slices of the z-axis.

Standard error propagation techniques do not properly represent the distribution of non-linear errors. The uncertainty can be represented by the variation in an ensemble of individually self-consistent predictions, as used in NWP or climate modelling. To illustrate, consider estimating the volume of an aluminium bucket knowing only its mass. Knowing the bucket's density and thickness, the volume can be calculated from the mass by assuming the shape and height of the bucket. Those choices (i.e. the forward model) will greatly affect how the retrieval interprets the mass measurement. In Fig. 4 each line represents a different forward model for converting mass into volume. A slice (lines of the same colour) shows the impact of shape on the form of the forward model. Looking through the slices (different colours of the same line style) shows the impact of the assumed height.

When the bucket is assumed to have a height of 12 cm (purple), the three models produce consistent results between 0.15 and 0.3 kg. The error due to using an inappropriate model there will be small, but increases for masses > 0.3 kg. Hence, the error is a function of the true state. For a height of 24 cm (red) the models diverge; a 0.32 kg bucket could have a volume between 0.10 and 11 L. Thus, the error is also a function of the choice of forward model. In this example the actual shape of the bucket is unknown, so it is not possible to rigorously quantify the error resulting from the choice of forward model. Without additional information, it is impossible to identify the appropriate forward model despite their different interpretations of the data.

The form of the ensemble will depend on its intended use and a priori knowledge. In this example, the ensemble would be three estimates of the volume (one for each shape). The uncertainty resulting from errors in the weight, density, and thickness would be given separately for each ensemble member. If genuinely nothing was known about the height, the ensemble could be extended to represent a range of heights. In reality, some auxiliary information will exist that should constrain the values.

References

[1] Joint Committee for Guides in Metrology (2008), Tech. Rep. JCGM 100:2008, `http://www.iso.org/sites/JCGM/GUM-introduction.htm`.
[2] C.D. Rodgers and B.J. Connor (2003), doi:10.1029/2002JD002299.