Uncertainty Topic 1: Introduction to Uncertainty

What is uncertainty?

We are faced with uncertainty every day. '*Will it rain today?*' or *'How long will it take to drive to work*?' are examples of uncertainties that we commonly encounter. With uncertainty, we cannot perfectly know the exact outcome of something. However, we use information, such as weather forecasts and our observations of the day's weather, to help infer the most likely outcome given our knowledge and to make decisions. Typically, the more information we have, the less uncertainty we face. For example, an hourly forecast may provide more information than a daily forecast. Reports on current road conditions can help inform how long our commute will take.

Uncertainty and methane emissions

Uncertainty is also present in our understanding of methane emissions. In a scientific context, uncertainty is an expression of doubt or incomplete knowledge about an outcome or quantity, such as total methane emissions, given the information that we have. Analytically, uncertainty is usually expressed as an interval of plausible values within which the true, but unknown, quantity is likely to exist. For example, a scientific study may conclude that total emissions are 6.5 million kg/year with a 95% confidence interval of (5, 8). This could also be written as 6.5 ± 1.5 million kg/year and indicates that our best estimate of total emissions is 6.5, but the uncertainty of 1.5 indicates that emission values between 5 and 8 are also plausible given the information (data) we have observed. This type of interval is called an uncertainty interval.

Like our work commute example, we expect uncertainty will decrease as we obtain more information and/or information of higher quality. For example, if more measurements had been collected or an instrument with improved measurement capabilities was used, the reported uncertainty would likely have been smaller resulting in a narrower uncertainty interval.

Why is uncertainty quantification important?

Uncertainty quantification is important for conveying the amount of information used to develop estimates of total emissions. Uncertainty quantification conveys information about the underlying quantity of, and variation among, measurements used to develop emissions estimates. Without uncertainty quantification, there is no indication of information quality and quantity and therefore no way to compare different estimates. For example, if one Amazon product is rated at 4.8 stars and another is rated at 4.6 stars, is the quality of these products different? Additional knowledge that the 4.8-star product has 12 ratings, and the 4.6-star product has 1200 ratings can help contextualize those ratings. Similarly, if one study estimates emissions at 6.5 million kg/year and another study estimates emissions at 6.9 million kg/year, do these studies agree or disagree with one another? Uncertainty quantification enables us to formally make those comparisons.

There are several sources of variation and error that give rise to uncertainty. Uncertainty quantification attempts to account for this variation and error in a mathematically rigorous manner. There are several ways to name and group the sources that contribute to uncertainty. We describe three broad categories that give rise to uncertainty in methane emission estimates.

1. Limited sampling and sampling variability are the first sources of uncertainty. Limited sampling occurs because measurements are often time consuming, expensive, and/or logistically challenging to obtain. Thus, instead of a census, we aim to collect a representative sample with as many measurements as our resources will allow. Sampling variability refers to the idea that different samples will yield different estimates of emissions due to random selection. One study could happen to randomly select and measure sources with relatively small emissions while another study could happen to select relatively large emissions. This sample-to-sample variation creates uncertainty in our final estimates.
2. Measurement error is the second source of uncertainty. When you measure ½ cup of flour for your favorite recipe, is the amount of flour always exactly ½ cup? Your measurements have variability and will have some measurement errors. Measurement error is the difference between the actual quantity being measured and the value of the measurement produced by an instrument. In most cases, instruments are calibrated to have an *average* measurement error of 0. However, individual measurements could still have large errors even if they are 0 on average. In general, larger measurement errors mean greater uncertainty.
3. Modeling error is a third source of uncertainty. Just like a topographic map is an abstract model for the terrain a hiker might experience, mathematical models are abstract descriptions of physical phenomena and used as approximations. For example, a Gaussian plume model is often used to describe how methane disperses from an emission point. A linear regression model can be used in an analysis to explain the relationship between measured CH_4 concentrations and emission rates. Modeling error reflects the idea that the way we describe how the data are generated is an *approximation,* and this approximation introduces error under complex systems or when important factors are not included in the model. While models provide a framework for improving our understanding of the physical or data generating process, they are approximations that are subject to their own errors and assumptions.

An important aspect of uncertainty is that it is *probabilistic*. An uncertainty interval provides a range of values within which the true, but unknown, quantity is *likely* to exist. This probabilistic nature arises from underlying probability and statistical theory and is quantified using the *level of confidence*. The greater the level of confidence, the wider the uncertainty interval. This relationship is analogous to fishing with a net – to be more confident that you'll catch the fish you're after, you need to cast a wider net.

Even though uncertainty is unavoidable, there are several ways to control or reduce the sources of uncertainty. Most political polls are designed and conducted to have a margin of error of 3 points, a measure of uncertainty in the polls results. This chosen level of uncertainty is achieved through sample planning. Sample planning is used to account for sampling variability by identifying an acceptable level of uncertainty and computing the number of samples that are needed to achieve that level of uncertainty. Measurement errors can be reduced by improving the measurement system, either through improved instrumentation, more testing, and/or better instrument calibration. Similarly, modeling errors could be reduced by improving the models with more data or developing models that better approximate the data generating process.

Unfortunately, not all sources of variability and error are reducible. Measurement systems have limitations, and as statistician George Box famously said, "All models are wrong, but some are useful." Furthermore, there is a cost to achieving reductions in uncertainty. Samples can be expensive and time consuming to collect. Managing and analyzing large data streams requires data storage and expertise in modeling and data management. Measurement systems can be expensive to test, deploy, and operate. Complex models can demand large quantities of data or computing power. Finally, the sources of error don't operate independently. For example, uncertainty arising from low quality measurements cannot always be offset by more samples or improving the model.

The information presented here provides the basic concepts to start understanding uncertainty and is the first white paper in a series on uncertainty and methane emissions quantification. Readers that would like to learn more about uncertainty quantification and methods for uncertainty quantification can read the Veritas Uncertainty Guidance document. Additionally, readers that want a deeper dive can read the International Organization for Standardization's "Guide to the Expression of Uncertainty in Measurements"[1]

References

(1) Mathew, K. J. *Guide to the expression of uncertainty in measurements*; LA-UR-17-29643; United States, 2017. https://www.osti.gov/biblio/1402579 https://www.osti.gov/servlets/purl/1402579DOI: 10.2172/1402579.