Lecture 27. Uncertainty Analysis (UA) for Models

"Uncertainty is not a weakness. Understanding uncertainty is a strength, and a key part of using any model . . ."
— Gettelman and Rood [5]

1   What is uncertainty analysis?

Let's start with why we might look to analyze uncertainty. Gettelman and Rood have a thorough definition of uncertainty itself, and Coleman and Steele have a very simple definition for uncertainty analysis (UA).

uncertainty "the noncorrectable or unknown part of the inaccuracy of an instrument, system, or model. Uncertainty represents the limit of measurement (or forecast) precision."[5]

uncertainty analysis "the analysis of the uncertainties in experimental measurements and in experimental and simulation results."[3]
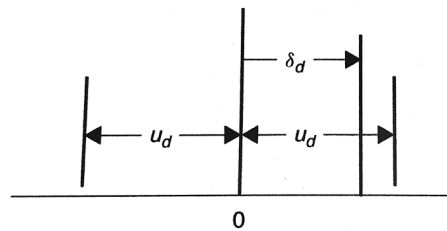

Figure 1: An uncertainty u defines an interval that is estimated to contain the actual value of an error of unknown sign and magnitude. [3]

1.1   Why perform UA?

When performing an uncertainty analysis, we are asking these kinds of questions:

- How much of a *big deal* is it if I'm not too confident in something going into my model?
- What is the effect of *stochastic variations* in some input(s) on the output(s) of the model?
- What is the effect of having a *range of potential values* or a particular *distribution of potential values* in some input(s) on the output(s) of the model?

So uncertainty analysis is a deeper look at what we don't know and how it impacts us. These questions go hand-in-hand with sensitivity analysis.

2   Sources of uncertainty

Recognizing that uncertainty describes things we can't correct or know completely, it is easy to identify many sources of uncertainty when you apply a critical eye toward modeling problems and measurements taken from energy systems. Imagine that in Fig. 2 you have a thermocouple taking a temperature

measurement. What are five things that could be causing your measurement to stray from the 'true' value?

Because computational models depend on some form of experimental measurements or empirical data (or at least they should!), the modeler has to deal with uncertainty from experimental data as well as uncertainty associated with the computer simulation itself. Considering how uncertainty compounds throughout the data gathering-modeling-simulation process is called **uncertainty propagation**, and is illustrated in Fig. 3.
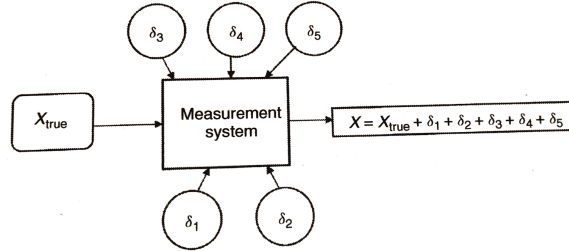

Figure 2: Measurement of a variable influenced by five error sources. [3]

We may be concerned with uncertainty in:

- Empirical data used within the model, or to validate the model
- Calculated data
- The structure of the model itself
- Model parameters

3   Uncertainty considerations in the modeling process

The following list is adapted from Coleman and Steele [3], who provide these questions to be considered as part of designing your experimental approach. However, considering the computer simulation to be a kind of numerical 'experiment,' they apply quite well.

1. What question are we trying to answer?
2. How accurately do we need to know the answer? (How will we use the answer?)
3. What physical principles are involved? (What are the governing equations?)
4. What model or set of simulation techniques might provide the answer?
   . . .
5. Can the requirements be satisfied within the budget and time constraints (including computational time)?
6. What techniques should be used to analyze the data?
7. What is the most effective way to present the data?
8. What unanticipated questions are raised by the data? (Should we revisit prior steps or redesign our modeling process?)
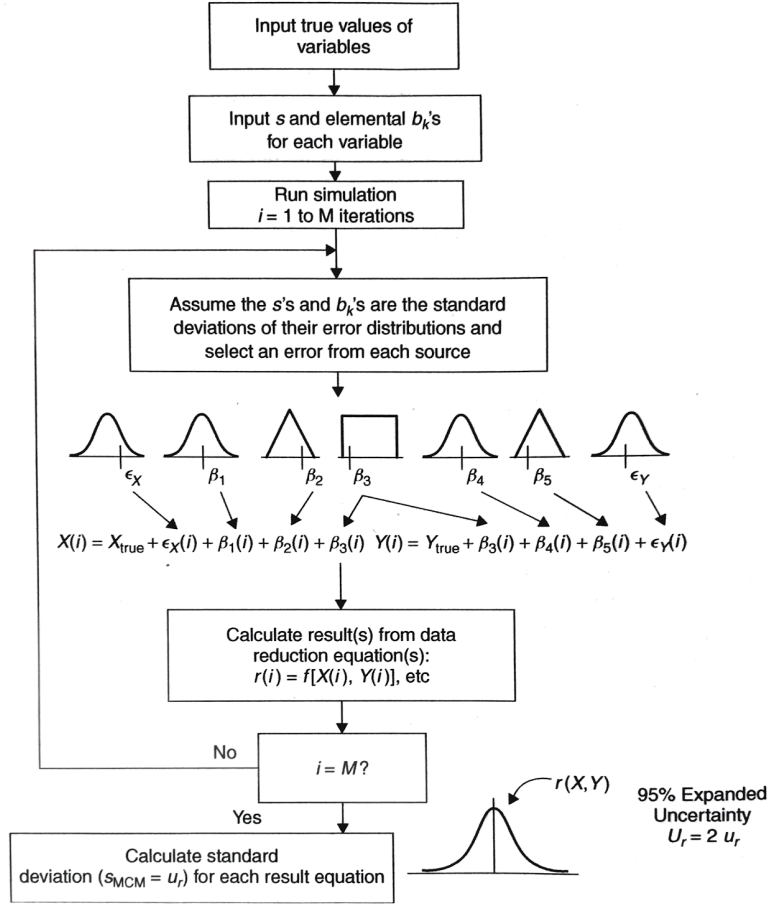9. How can we best report the data and results?

Figure 3: Schematic of Monte Carlo Method for uncertainty propagation when random standard uncertainties for individual variables are used. [3]

4 Bayesian model averaging for uncertainty estimation

In classical statistics, models are created similarly to the flowchart in Fig. 3. The analyst knows or assumes some distributions for the variables and proceeds from there. You should recognize β_1 as a normal distribution and β_3 as a uniform distribution. β_2 is another simple distribution called a triangular distribution, which is pretty self-explanatory.

Bayesian statistical approaches turn this usual pattern on its head. Pfeiffer describes it this way: "Since the population mean is a quantity about which there is uncertainty, it is modeled as a random variable whose value is to be determined by experiment. In this view, the population distribution is conceived as randomly selected from a class of such distributions." [6]. Here, the population mean represents the 'true' value and the experiment can be a computational simulation.

If you've taken a statistics course, you may remember the 'chain rule' provided by Bayes that describes how likely it is for something to happen *given that you know the likelihood of something else that is related to the event you are interested in.* Christian and Griffiths [2] point out that Bayes' Rule "makes it possible to *reverse-engineer* all kinds of prior distributions, even ones about which it's harder to get authoritative real-world data." It follows, then, that "the richer the prior information we bring to Bayes' rule, the more useful predictions we can get out of it" [2]. This is true for practically any modeling technique you will use: If 'rich' data means it is more representative of the system/process you are trying to model, then additional data and higher quality data will improve your predictive accuracy. You may remember that the representativeness of the data set was a key concern for creating machine learning-based models.

Bayesian model averaging is simply a way to implement this statistical knowledge for the benefit of our modeling efforts. Monte Carlo (randomized sampling) methods are frequently used, and Saltelli et al. [7] describes a case study using the Generalized Likelihood Uncertainty Estimation (GLUE) approach, which is a simplified way to implement Bayesian inference using a weighting function. They also caution that "It is necessary that the sample generated for the Bayesian analysis is also designed for the computation of variance-based sensitivity indices," [7] which is important when performing a combined GSA-UA analysis of a model.

Key ideas on Bayesian Model Averaging [7]:

- Attempts to take into account all possible sources of uncertainty
- Bayesian statistical methods account for the likelihood of events given other known distributions or likelihoods
- GLUE [7, 1] is just one method of implementing this.

Possibilities for using GSA in Bayesian uncertainty estimation [7]:

- Model output itself
- **Likelihood**   "a function that quantifies how well that particular parameter combination (or model) simulates the system" [1]

Here is a free book (web version) from Green Tea Press that is a great resource for learning more about Bayesian analysis (particularly if you already use Python):

Think Bayes [4]

We may use GSA and Bayesian uncertainty estimation together to gain better information about which factors are most important, how complex the model and factor interactions is/are, and which factors are involved in those interactions and in the model outcome.

References

[1] Roberta-Serena Blasone et al. "Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling". In: *Adv. Water Resour.* 31.4 (2008), pp. 630–648. ISSN: 0309-1708. URL: http://www.sciencedirect.com/science/article/pii/S0309170807001856.

[2] Brian Christian and Tom Griffiths. *Algorithms to Live By: The Computer Science of Human Decisions.* en. Macmillan, 19 4 2016. ISBN: 9781627790369.

[3] H W Coleman and W G Steele. *Experimentation, Validation, and Uncertainty Analysis for Engineers.* Cambridge monographs on mechanics. Wiley, 2009. ISBN: 9780470168882. URL: https://books.google.com/books?id=6VJ_7cCW3VkC.

[4] Allen B Downey. *Think Bayes: Bayesian Statistics in Python.* en. 1 edition. O'Reilly Media, Apr. 2013. ISBN: 9781449370787. URL: http://greenteapress.com/wp/think-bayes/.

[5] Andrew Gettelman and Richard B Rood. *Demystifying Climate Models: A Users Guide to Earth System Models.* en. Springer Berlin Heidelberg, 19 4 2016. ISBN: 9783662489574.

[6] Paul E Pfeiffer. *Applied probability.* cnx.org, 2009. URL: http://link.springer.com/content/pdf/10.1007/978-1-4612-6335-7.pdf.

[7] A Saltelli et al. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models.* en. John Wiley & Sons, 16 7 2004. ISBN: 9780470870945.