Analysis of Uncertainty in Calculations

R. Russell Rhinehart
russ@r3eda.com, www.r3eda.com

INTRODUCTION

Applications

If an experiment is replicated (exactly duplicated as far as humanly possible) the results should be exactly the same. But, they are not. There is uncontrolled variation in the processing, testing, source materials, environmental conditions, test machines, operator techniques, sensor discrimination, etc.; and these lead to variation in the supposed-to-be exactly-the-same outcomes.

This propagates to uncertainty in the models that are fit to the data. Then, model-calculated uncertainty becomes uncertainty in the forecast values in the model. Further, model "givens" are often estimates of the values of properties or future events, which have uncertainty that affects predictions. The questions are related to determining uncertainty in experimental data, model coefficients, and model predicted values.

There are diverse applications for analysis of uncertainty in models. These include determining:

1. Uncertainty of calculated values in laboratory analysis – taking test data and models to calculate analysis or properties, then determining the range of uncertainty in analysis or properties.
2. Uncertainty of calculated values in modeling or prediction – using a model to see what the range of outcomes might be due to uncertainty in the basis or givens. Or in inverse of models when sensor data is used to determine the process variable (for example dP is used for Q-dot, I.R. or T or conductivity is used for composition).
3. Uncertainty of models that were fit to empirical data – Whether empirical models or phenomenological models, determining uncertainty in
    a. regression coefficients and
    b. model prediction due to vagaries in experimental data.
4. Validation of models by comparing residuals to expected uncertainty of the data – if residual σ matches data replicate σ (or propagated σ on data) then the model might be good.
5. The number of replicates or number of experimental data sets – to obtain an average measurement or a model with adequate uncertainty.
6. When a time series has achieved steady state – such as measuring sample moisture content and claiming it is dry when three measurements in a row are "equal" within uncertainty.

This presentation summarizes techniques for each.

Objectives/Rationale

There are diverse objectives or reasons for applying uncertainty analysis:

1. One objective in analysis of uncertainty, in either laboratory or process results, is process quality improvement. Analysis of uncertainty will let you see how uncertainty (often termed random error, but it could also be a systematic error) is propagated through calculations to quantify the uncertainty of the calculated value. Uncertainty could be quantified as the range of the values or the standard deviation of values (when ideally the replicate values are identical). If the calculated uncertainty matches the actual, then you understand your process. Then, this analysis can point you to process or procedural changes that reduce variability, improve quality.
2. As a second objective, uncertainty analysis provides legitimacy to a claim about uncertainty that you report to others.

3. Alternately, if the expected and actual variation do not match, then this would be a trigger to find out why, and learn more about your process. A third objective is to truly understand your process. Then you can be better at trouble shooting and improving quality.
4. Analysis of uncertainty also relates to developing models. Models generated from empirical data reflect the perturbations in the particular data sample. A replicate set of experiments will be subject to different perturbations. In modeling from empirical data, the question is, "How does experimental variability impact the certainty limits of the model?" A fourth objective is to answer that question.
5. Further, in developing phenomenological models (mechanistic approach, from first principles), the question relates to the residuals (the deviations between model and data). If residuals are greater than propagation of uncertainty expects, then the model should be rejected. Additionally, if residuals are either unexpectedly large or small, the process understanding should be questioned. A fifth application is relative to phenomenological model validation.
6. We use models to predict, forecast, design, analyze, etc. But "givens" in the models have uncertainty. These include empirically derived coefficient values, but also the estimates of future costs, compositions, fees, tax rates, production/sales/demand/load/duty and such. A sixth application is to propagate uncertainty of the "givens" in a problem statement to the calculated value. This will provide an uncertainty range on designs and engineering calculations that are essential to assign safety factors or assess risk.
7. Seventh, in optimization we seek the best value: lowest cost, least risk, most reliable, etc. However, if experimental results guide the evolution of the decision variables (DV) in optimization, the uncertainty on the data will lead to uncertainty on the optimum DV values. Further, if the optimization is using models, uncertainty on the coefficients and other "givens" will create uncertainty on the optimum DV values. Analysis of uncertainty is needed to reveal how the model or experimental uncertainty affects the range of DV values.
8. Eight is in experimental design, for example determining the number of data samples or measurements to reduce uncertainty to an acceptable level, or to assess when sequential measurements indicate that a process has come to steady state or equilibrium.

Propagation of Uncertainty

In all of those diverse applications, the concept is to use a model to propagate uncertainty. For instance, the calculation might be to compute stress, S, from tensile load and cross sectional area, $S = L/A$. If there is uncertainty, possible error, on load and area, ε_L and ε_A, then propagation of maximum error on S using the analytical approach is $\varepsilon_S = \left|\frac{\partial S}{\partial L}\varepsilon_L\right| + \left|\frac{\partial S}{\partial A}\varepsilon_A\right|$, which reduces to $\varepsilon_S = \left|\frac{1}{A}\varepsilon_L\right| + \left|\frac{L}{A^2}\varepsilon_A\right|$, and still more simply $\varepsilon_S/S = \varepsilon_L/L + \varepsilon_A/A$. Here "L" and "A" would be the called the independent or input variables, the "givens", and "S", the calculated consequence, termed the dependent or response variable.

Whether analytical or numerical propagation, whether reporting probable or maximum error, and whether uncertainty on the independent variables is evaluated by range or standard deviation, to analyze uncertainty on a calculated value (dependent variable) requires 1) a model, and 2) some measure of uncertainty on the elements (independent variables).

In the diverse applications listed above, there are common techniques. The first section of this monograph, Fundamentals, provides definitions, classification, and basic formula. The second section,

Propagation in Calculations, reveals the basic approaches. The following sections provide nuances for the diverse applications.

Nomenclature

R = range high minus low values.
ε = error, half range, $\pm\varepsilon$ should include 97% or so of the values.
x = variable. It is variously used to represent the calculated value, or the independent values. It is also used to represent the givens or coefficient values.
σ = the true population standard deviation, which is unknowable. To obtain an experimental value for σ, you need to sample the entire population, infinite samples. If you have a theoretical basis for σ, then you have an ideal value, which can only be confirmed with infinite number of samples from the entire population. Either way, the value of σ is unknowable, it is either a theoretical idealization of reality, or an empirical approximation.
s = the estimate of the population standard deviation.
y or $\tilde{y}$ = model calculated response variable, the model output.
x^* = optimum value of a model input, an independent variable, the decision variable.
$a, b, c,$ = model coefficients.

Resources

This presentation is based on several publications and postings on my web site:

- ASTM Standard E29-88, Standard Practice for Using Significant Digits in Test Data to Determine Conformance with Specifications, American Society for Testing and Materials Annual Book of Standards, Sect. 14, Vol. 2, 1989.
- ASTM Standard E177-86, Standard Practice for Use of the Terms Precision and Bias in ASTM Test Methods, American Society for Testing and Materials Annual Book of Standards, Sect. 14, Vol. 2, 1989.
- Rhinehart, R. R., Instrument and Automation Engineers' Handbook, Vol I, Process Measurement and Analysis, 5th Edition, B. Liptak & K. Venczel Editors, Section 1.14 (maybe 1.10), "Uncertainty – Estimation, Propagation, & Reporting" Taylor and Francis, CRC Press, Boca Raton, FL, September, 2016.
- Bethea, R. M. and R. R. Rhinehart, Applied Engineering Statistics, Chapman & Hall/CRC, Boca Raton, FL (originally Marcel Dekker, Inc., New York, NY), July 1991. ISBN 0-8247-8503-7.
- Rhinehart, R. R., Nonlinear Regression Modeling for Engineering Applications: Modeling, Model Validation, and Enabling Design of Experiments, Wiley, New York, NY, September, 2016. ISBN 9781118597965.
- Rhinehart, R. R., Engineering Optimization: Applications, Methods, Analysis, Anticipated to be released in 1st Quarter 2018, John Wiley & Sons
- Law and Kelton, Simulation Modeling and Analysis, 2nd Edition, McGraw Hill, 1991.
- www.r3eda.com

FUNDAMENTALS

What Variation Is and Is Not

To experimentally estimate uncertainty, only use replicate trials, independent trials that should provide the exact same result, because they nominally were run at exactly the same conditions. Do not use an s-value that is calculated from all of the data, use the s-value from replicates only.

Here is an example to clarify the difference: I measure the height of each grandchild when they come to visit. Stand them next to the wall, eye-ball level the pencil from the top of their head to the wall, and mark the wall. If the pencil is not perfectly horizontal, or their socks are thick, or I don't start from top-dead-center on their head, then the mark on the wall has some error to it. Perhaps each mark could be ± 1/8th inch from true. Each mark for each grandchild for each visit is off by a maximum of about 1/8 inch; but, not every mark is off that much, some are closer to the exact value. 0.125 inch maximum error on any one, means about 0.125/2.5 = .05 inches standard deviation. If I were to replicate one kid's measurement 100 times, and look at the distribution of marks, I expect the sigma of the variation for any one measurement would be about 0.05 inches.

At one point in time, Kennedy was 63", Jamaeka 62", Parker 61", Conor 45", Ashton 40", and Kain 33". Landon was not old enough to stand up yet. The average of 63, 61, 62, 45, 40, and 33 is 50.67". But, none of them are 50.67" tall. The standard deviation of the same set of height data is 13.003". But, each of my measurements are not in error by 13 inches.

What you need in propagation of uncertainty is the measure of uncertainty in a particular measurement (the 0.05"), not the standard deviation of all the data in your experimental conditions (the 13.003"). To experimentally estimate uncertainty, use only replicates, independent trials that should provide the exact same result.

As another example, roll a standard cubical game die. The average value of many rolls should be 3.5. But, you can only see values of 1, 2, 3, 4, 5, or 6. In N rolls the discrete uniform distribution indicates $\sigma_{on\ the\ average} \cong 1.71/\sqrt{N}$. Alternately, on a 10-sided die, the average is 5.5, and with N rolls $\sigma_{on\ the\ average} \cong 2.87/\sqrt{N}$. Replicating would only be rolling one type of dice. Only analyze data that is expected to have the same value. Do not collectively include values that are not expected to be identical.

So, when analyzing uncertainty, don't include Product A data with that of product B. Don't include test data at one temperature with that of another. Likely, you should not include data from one lab with data from another, or perhaps data prior to and after a device recalibration.

Measures of Variation

There are two key measures of deviation from the average (the expected or nominal value): error (ε) and standard deviation of the variable (s). These are related. Simplistically, $s \cong \varepsilon/2.5$.

There is a range of values that might be expected when something is replicated. Range (R) is the difference between the highest of replicates and the lowest. Ask someone what the range might be, and

they will not tell you the highest and lowest all-time possible extreme values. Similarly they will not tell you the last two measurements. They will tell you the normally expected range, perhaps the range that includes 97% of all possible values. This would be approximately the 5σ range. If the data is symmetric, then $R = 2\varepsilon \cong 5s$. So, $s \cong \varepsilon/2.5$. You need many samples, perhaps 100 or more, to see the range. Two measurements, for instance, will provide a range, but probably not a range that is representative of the population.

Standard deviation is calculated from all of the available data, $s = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$, which is a portion of the infinite number in the possible population. Range is only calculated from two values, the high and low in the data set, $R = x_{high} - x_{low}$. So, s, calculated from all N data, is a better, estimate of variability. However, when sigma is estimated from the range or half-range error, $s \cong \varepsilon/2.5$, it is a reasonable, but remains a coarse two-point approximation.

But, neither are perfect measures of variability. Roll a standard 6-sided die 2 times, theory indicates that you expect $\sigma_{of\ the\ values} \cong 1.71$, but you will probably not get that. For instance, if you roll a 3 and a 5 then $s = 1.41$, and if a 1 and a 4 then $s = 2.12$. These are 1.7-ish, which would indicate that the process is understood. Rolls with a greater number of dice would lead to a sigma that is closer to the expected value of about 1.71, but it may take a million rolls to experimentally see the 1.71 value.

Since you won't have millions of replicates, since you will only have a small sample of the population, any method that you use to estimate the sigma or range will be imperfect.

The Chi-square distribution can be used to indicate the expected range on s, which can indicate whether the experimental value is close enough to the model-calculated expected value. In normal situations the true σ-value might be between half and triple the calculated value (roughly the 97% limits with 10 data). However, formal Chi-Square or F-statistic testing is not required to assess uncertainty on the variability. I think an intuitive decision as to whether the model-calculated and experimental s-values are close enough to claim they are the same is often all that is warranted.

So: In uncertainty analysis, there is uncertainty in the outcome. Do not expect to predict exact values. One must be willing to judge whether a calculated value is close enough or not. A one-third to four times ratio is not unexpected.

Sources of Variation

We have many names for the uncertainty of numerical values. These include error, fuzz, noise, bias, fluctuation, and variance. Since the term "error" often connotes a mistake, and "bias" denotes a systematic deviation, the term uncertainty is a better representation. Sources of uncertainty are not necessarily human mistakes. They include the naturally occurring variability on measurements due to either systematic bias or random fluctuation. They also include process-to-model mismatch that results from idealizations, truncation, or the use of tabulated data. This monograph will use the term uncertainty; however, both the terms "error" and "propagation of error" remain as commonly used labels.

Following are a few sources of uncertainty encountered in engineering analysis.

Estimation: Often the basis of a calculation is an estimate. "Oh, I expect we'll be able to sell 25 metric tons per year." But, a plant that large may be a large capital risk for the company resources. Accordingly, the investment managers will want to know the likelihood of only being able to sell 15 or 20, or of the potential to sell 30 or 35 metric tons per year. Whether termed "givens" or "basis" such estimates of the situation are uncertain. Similarly, economic profitability indices such as Net Present Value or Long-Term Return on Investment, are based on estimates of future sales, tax rates, raw material costs, etc. Who can precisely predict the future? Again such basis for investment choices have uncertainty.

Discrimination: No measurement device is capable of infinitely small measurement intervals. For instance, a pressure gauge may be marked in increments of 10 psi, and one might be able to estimate a gage reading to the nearest 2 psi. Discrimination limits the reading to $\pm$ 1 psi. Similarly, if a 0 to 12,800 gpm flow rate reading is processed by an 8-bit computer, the discrimination ability is 2^{-8}, about 0.4% of full scale, about 50 gpm. Consequently, as the flow steadily increased from 900 to 1000 gpm, the computer would continue to report 900 gpm during the period in which the flow was changing from 900 to 910, to 920, etc. At 950 gpm it would report (jump-to and hold) 950 gpm until the flow reached 1000 gpm. Similarly, discrimination error in reading numerical data from charts and diagrams depends on the thickness of the pencil line or the scale of the axis. From tabulated data discrimination error is related to the number of digits displayed in the data. As a common example, my cell phone indicates time in hours and minutes. It displays 7:23 am until it becomes 7:24. Even though time progressively increases, the reading remains at 7:23 until it jumps to 7:24.

Calibration Drift: A metal ruler lengthens and contracts due to temperature change. A wooden ruler also changes length due to humidity. Temperature affects spring stiffness in pressure gauges or weigh scales. Temperature changes the resistance and capacitance in the electric circuitry of sensors and transmitters, and may be due to either ambient conditions or unit warm-up in uses. In general, calibration drift is affected by instrument age, temperature, humidity, pressure, oxidation, actinic degradation, fouling, and catastrophic events such as dropping the device.

Accuracy: Sensors and transmitters are not exactly linear. Calibrations are usually performed by adjusting a transmitter zero and span so that the instrument reports a value "close enough" to the "true" value of two standards. Then one assumes that the instrument response between the standards is linear even though the response is not exactly linear. (In fact, the local calibration standards are not perfect, either.) The *accuracy* of a device, its reading deviation from the true value, is often called *systematic error* or *bias*.

Technique: The measurement procedure may measure the wrong thing. For example, if knowledge of the steam pressure is required to size a reboiler, one must have the pressure downstream of the steam flow control valve and not the steam header pressure. As another example, a thermowell might be measuring the temperature in a hot spot in a furnace, but this would not represent the average temperature over all of the space which drives the heat transfer.

Constants and Data: Most calculations involve fundamental constants, but their values come from previous measurements that are not exactly known. Examples of constants include the gas law constant, the gravitational constant, the speed of light, and the molecular weight of sodium. Typically, these values are known to errors of only 0.001% or 0.01%. Examples of data include a tabulated viscosity or thermal conductivity. Typically these data are known only to 0.1% to 1% error. Often, data are obtained from correlations or graphs, such as pipe flow friction factors, thermodynamic properties of mixtures, and convective heat transfer coefficients. Such values may have a 10% to 30% uncertainty.

Noise: Process flow turbulence, changing electromagnetic fields around equipment, equipment mechanical vibrations, etc. may cause the measured value to randomly vibrate or fluctuate about its average value. Due to the fluctuation, any one instantaneous reading may not reflect the time-local average. One has to average many values to temper the uncertainty of random noise. Noise reduces your ability to obtain a precise measurement. Precision, repeatability, is related to such random influences.

Model and Equations: Nonidealities are usually not expressed in models of phenomena, and, accordingly, inaccuracy is reflected in the equations that are used to calculate the results. For instance, the ideal gas law ($PV = nRT$) is a model of particle dynamics that neglects the effects of intermolecular forces and molecular size. Although the law is often used for gases at ambient conditions, it can introduce a 5% error at those conditions and up to an 80% error at conditions near the critical point of real gases. The volume of a tank may be calculated by $V = \pi r^2 h$, an equation that is based on a right-circular cylinder model. Such a model neglects real surface irregularities such as the effects of dents or ribs, sensor intrusions into the side of the tank, and the curved bottom and the length of drain pipe above the bottom discharge valve. The square-root relation between differential pressure and flow rate, derived from the ideal Bernoulli Equation for inviscid streamline flow, is commonly accepted in orifice calibration equations, and may lead to a 5% error. The Beer-Lambert relation for the concentration dependent light absorption is commonly accepted in spectrophotometric devices. Models are often incorrect because they are incomplete; therefore, the calculated values (whether part of the measurement device or off-line) are also imperfect.

Humans: Humans are often part of the data acquisition and transmission process. We might transliterate, reversing the position of adjacent digits. We might decide when a process has achieved steady-state for sampling, but a desire to finish the project might override waiting long enough for the transient to settle. A noisy signal might be biased by using a convention that reports the mental average of the upper (or lower) extreme that was observed, or the most frequent value that was observed. A human might judge that a data point is faulty because it is inconsistent with expected trends in the data, and may discard that point, when in fact the data was good, and the understanding of the process was wrong.

Data and Process Models

Models could be classified as either data models or process models. The distinction between data model and process model is in its use. It is not fundamental to a model.

Data models are used in the experiments to convert sensor data into the composite measurement. For example, consider a bucket and stopwatch method to measure flow rate. Weigh the full and empty bucket and time interval to collect material. These are the measurements, then use the data model to calculate volumetric flow rate, $F = (W_{full} - W_{empty})/(t_{end} - t_{start})/\rho$. In this case the data model is the definition of flow rate. Here, the estimated uncertainty on the five data values (W, W, t, t, ρ) would be used to evaluate uncertainty on the calculated F.

We often think of the calculated flow rate as the measurement, because it is the property or response that we want to see. But, F is truly not the measurement, it is the calculated value from the measurements. If we are interested in how valve position affects flow rate, then we plot the data-model-calculated flow rate "measurement" w.r.t. the experimental control. But, the "measurement", was not measured. Weights and times were the measurement. The "measured response" was calculated by a data model.

Similarly, in getting a yield stress measurement from a tensile test, the actual measurements are breaking load and cross sectional dimensions, and stress is calculated from a definition, the data model. Again, similarly, when using conductivity as a sensor to infer salt composition, the measurement is conductivity, and the inverse of the calibration model is the data model that is used to report composition.

Contrasting the data model, the process model is the model that describes the response behavior of a process or product or procedure. For example, a process model might be that used to calibrate an orifice flow meter using the F results from above. For instance, the model might be $F = a(i - i_0)^b$, where i (transmitted mA signal) will be used to calculate flow rate for display. (Ideally, the Bernoulli relation would indicate that $b = 0.5$, and you may have expected to see a square root relation in the orifice equation. But, numbers in the $b = 0.46$ range often provide a better representation of the nonideal reality.) In calibration, the objective is to adjust the model coefficients (a, b) to make the calibration model best match the (F, i) experimental data. Here, uncertainty on the data-model calculated F, and sensor transmitted (i, i_0) values will affect certainty on the model coefficients (a, b). In this regression application, the response value is known, not calculated. It is the model coefficients that are the unknowns.

Subsequently, in on-line process monitoring use, this calibration process model becomes the data model, converting sensor data (i) into "measured" flow rate. So the process model could be considered the data model in an alternate use.

The process model could also represent how the process responds. For instance, the question might be, "How does flow rate depend on valve position?" Perhaps the model could be an equal-% type, $F = ae^{o/R}$ where o is the controller output and (a, R) are model coefficients. Regression would be used to determine the (a, R) values. Once the model is obtained, the inverse of the process model could be used to determine the controller output needed to achieve a target flow rate: $o = Rln(\frac{F}{a})$.

Again, what is classified as the data model or process model depends on the application of the model, the same model could be viewed as representing either category, depending on the situation. The model could be used to calculate the response, or the inverse of the model used to calculate the influence to achieve a desired response, or the model could be used in regression to adjust coefficients to best fit data.

I'll use several common models as examples. The first is the calculation of volumetric flow rate with the bucket and stopwatch method, $F = (W_{full} - W_{empty})/[\rho(t_{full} - t_{start})]$. The second is calculation of sample breaking stress from a tensile test, $S = L/dw$. The third is moisture or volatile content as measured by drying in an oven, $\% = 100 * (W2 - W3)/(W2 - W1)$. Last is the volumetric flow rate calculated from the milliamp signal from an orifice flow meter, $F = a(i - i_0)^b$. The examples are all relatively simple, but reveal modest derivative complexity. This partly motivates the choice of numerical, not classic analytical approaches to uncertainty propagation.

Explicit and Implicit Models

Models could be classified as either explicit or implicit. An explicit model can be rearranged to isolate the response variable on one side of an equation. Then the functionality on the other side directs how to explicitly calculate the value.

However, some models contain functionality that prevents explicit solution of the relation. This might be an algebraic relation such as $\sqrt{x} + 0.1\ln(x) = 17$, which seems impossible to rearrange to isolate the variable x. For such, a numerical root finding procedure is an option. Additionally, implicit models might require solving an optimization, or determining the point along a path or time that meets a particular criterion. This could appear as $\underset{\{x\}}{min}\, J = f(x) = \frac{17}{x} + x^{0.1}$, or $17 = \int_{0.1}^{x} f(x)dx$, or any number of forms with path or transient model formats. But, whether implicit or explicit, uncertainty on the coefficients 0.1 and 17, create uncertainty on the x-value.

In addition, implicit models have a convergence criterion that determines when the solution is close enough to claim convergence. Implicit solutions do not give the exact solution, but iteratively approach it. Be sure that your threshold is chosen to have insignificant impact (two or three orders of magnitude less) on the answer, relative to other sources of uncertainty.

Significant Digits

Digits in a number that are significant are the digits that have values of which we are fairly confident. Whether explicitly stated or implied, you should reveal the uncertainty in reporting numerical values. The following is a reporting convention for integers and real numbers.

An integer has no decimal point. It is used to represent the number of whole events or whole things. A real number has a decimal point, and is used to represent the value of a continuum variable that can have fractional values. Some uncertainty is associated with both integers and real numbers. For instance, if you tried to count about 2,000 items, you might lose track, or you might eye-ball and remove groups of ten, and might have a count that is off by 10 or so. The number of items should then be reported as 2013 ± 10 to explicitly acknowledge the counting precision. Similarly, the speed of light in a vacuum is reported as 2.997925 x 10^8 m/s ± 0.000003 x 10^8 m/s to explicitly acknowledge uncertainty.

By custom, the uncertainty in numerical values is usually not explicitly reported, but is implied by the number of digits reported.

For continuum-valued numbers, the last (right-most) reported digit is the largest digit with uncertainty. When the precision is implied, we do not explicitly know whether the last digit is accurate to ±1 or ±2 or ± 3 or ±4. If the uncertainty were ±5 with a range of 10, the next digit to the left would be uncertain and would have been the last reported digit. If ±0.2 the last reported digit would be known with certainty, and the next digit to the right, the first fuzzy digit should be reported. Therefore, the uncertainty on the last digit reported could range from ±1 to ±4. Without specific guidance, we will assume a mid-value for the uncertainty of ±2.5. For continuously valued numbers, the last reported digit from the left, even if zero, is the fuzzy number. For example, a length may be reported as 100.0 in., which implies an error of about ± 0.25 in. Similarly, the atomic weight of Hydrogen = 1.00794 g/gmole (based on the ^{12}C isotope) implies that the value is known with an error of about ± 0.000025 g/gmole.

For integers, the last nonzero digit from the left represents the fuzzy number. For example, the 2010 census reported the population of the United States as 308,700,000. The last of the three nonzero digits, 7, is fuzzy and reflects a counting error of about 250,000 people. This method is the convention for reporting integers, but it has the unfortunate aspect that significant zeros are not identified. For instance,

if there are 2000 ± 10 items, one reader might interpret a reported 2000 as 2000 ± 250, and another as 2000 ± 2.5.

The +/- 2.5 is the average error if the last digit is the first (from the left) fuzzy (uncertain) digit. The maximum error would be twice that. Roughly, the sigma is the maximum uncertainty divided by 2.5. For example, if a table indicates a value of 0.624, and the "6" and "2" are certain, but the "4" is the first uncertain digit, then the nominal or average uncertainty on the data is expected to be +/- 0.0025, and the maximum uncertainty on the data value might be +/-0.005. But, the deviation is not always the maximum value. Sometimes the deviation might be very small. The standard deviation would be about 0.002.

By contrast, if the last reported digit was the last digit with a certain value, then the uncertainty would be ±0.25. Accordingly, both reporters and readers must use care in reporting and interpreting implied precision. Depending on the convention, the reader may make an order of magnitude error on interpreting an implied precision.

Estimating Uncertainty on Values

Before we can propagate uncertainty through calculations, or include uncertainty in the OF values, or report the uncertainty on calculated (or dependent) variables, we need to know the uncertainty on the numerical values that we use within the calculation. There are many legitimate ways to determine uncertainty on independent variables.

We often get data from tables (viscosity, thermal conductivity, density, etc.). If the table does not explicitly report an uncertainty, and you believe that the last reported digit is the largest uncertain digit, use ±2.5 on the right-most reported value as an average uncertainty. This means that the maximum uncertainty is ±5.0 on the right-most reported value is the maximum uncertainty, or roughly that $s = 2$ on the right-most digit.

If it is a measurement, replicate it enough times to be able to calculate the standard deviation. Usually, 10 replications will be ample to provide an adequately definitive estimate (the range on the true sigma could be within 0.6 to 2 times the 10-sample estimate), but 5 replications are often minimally adequate, and better than 3 which provides an uncertain estimate (the range on the true sigma could be within 0.4 to 14 times the true value). If the distribution is Gaussian (most measurements are close enough to this ideal distribution) then about 99% of the values, the range, fall within about ±2.5σ of the mean.

If the instrument has a calibration record or manufacturer specifications that indicate precision, use that for repeatability. (If it reports accuracy, use that for systematic error).

If the numerical value is calculated from a linear model with coefficients determined in a least-squares regression or correlation, use the "standard error of the estimate" as the standard deviation on the model coefficient value.

If regression of a nonlinear model to data, use bootstrapping to estimate uncertainty on the model coefficient and prediction values.

If an optimization procedure bounded the optimum, use convergence criteria to estimate the uncertainty in DV* as $\pm\frac{1}{2}\,\Delta x$ threshold.

Use your judgment and experience to estimate the possible error. For instance, in using a stopwatch to time collection of material in a bucket, you might estimate the start and stop time to each have a 0.5 second error. For another instance, in reading data from a curve, you can estimate the error that might happen due to pencil line thickness, your lines not being exactly parallel to the axis lines, discrimination ability on the two axes, or curvature of the graph due to photocopying.

To experimentally estimate uncertainty, only use replicate trials, independent trials that should provide the exact same result, because they nominally were run at exactly the same conditions. Do not use a σ that is calculated from all of the data.

Because uncertainty on input values to an equation are often estimates, one cannot expect the model-propagated value to be the true value. So, in uncertainty analysis, there is uncertainty in the model-calculated value as well as the experimental outcome. Again, one must be willing to judge whether a calculated value is close enough or not.

Perhaps, if the sigma experimental and model-calculated sigma are within a 3:1 or 0.5:1 ratio, you might be justified to claim that σ-modeled is equivalent to σ-experimental. But, a 10:1 ratio would be an indication that the analysis does not match the data. A 4:1 ratio would not lead to confidence in either assessment.

Random and Systematic Error

A systematic error is a consistent bias, a deviation from true that persists for all measurements. A random error is an independently changing deviation for each measurement.

For instance, in using a deliver-to graduated cylinder that is marked in 1 ml increments to measure a volume for a recipe, the user might be able to fill the cylinder to +/- 0.2 ml. One liquid drop more or less might cause a noticeable deviation, and would be corrected. However, the fill point may randomly change by an unnoticeable half-a-drop with each use. However, the glass expands and contracts with temperature (room T, heat from the technicians, hand, sample T); and as a result, the marking of the volume might actually include +/- 0.1 ml more or less than the reading. Then, when the contents are poured out, the residual liquid in the cylinder may vary from 0.1 to 0.2 ml, due to user technique and temperature-induced liquid viscosity, for an additional +/- 0.1 uncertainty. Primitively, the random error on each use could be the sum of all sources, +/- 0.4 ml. Not all uses, however, would have that maximum error. In some uses, the three error sources could be counter to each other, and in some uses the individual errors would not be at the maximum values. If the errors could be plotted with use, ideally they would average zero, be symmetric, and have a Gaussian distribution with 95%, or so, of the effort values within the +/- 0.3 ml range. This would be random error.

However, the graduated cylinder is likely not a primary standard, and the marking of 132 ml, for instance, might actually deliver 0.8 ml less than expected. This error would be the same in each use. Now, if possible to assess, a plot of the error w.r.t. use would not average zero, but would average 0.8 ml with a

random fluctuation between 0.4 and 1.2 ml. This persistent offset of 0.8 ml would be the systematic error or bias.

The systematic bias may either have a known or unknown value. If, for instance, you know that your bathroom scales read 2-lbs too low, then add 2-lbs to the reading. This would be a systematic error of a known value. If you know the systematic error value, correct for it. However, you may be aware that no home scales are a perfect, primary standard device, know that there is a systematic bias, but not know its value. Without knowing the value, you cannot correct the reading. This uncertainty on the reading imparts uncertainty to the true value.

This concept idealization, the segregation of random and systematic error, is useful. But, if the graduated cylinder is progressively heating (or cooling) over sequential uses, then that influence would not be random w.r.t. use. It would appear as a systematic error that is progressively changing with each use, an autocorrelated bias.

The error then, might not be ideally a fixed bias with an independent sample-to-sample fluctuation. The error might be autocorrelated. Or it might change from day to day. Or if several devices are used for the measurement, the error might correlate to which device is used.

The device would be marked with the magnitude of the systematic error, for example +/-0.8 ml. But, you do not know whether the particular device is reading 0.8 ml high or low or some intermediate value. The value of the systematic error is unknown.

If the value of the systematic error were known, you would use that knowledge to correct the measurement. By contrast, if the systematic error value is unknown, then use uncertainty analysis to determine the impact of the bias on the measurement.

Both random error and unknown bias can be propagated using the same techniques. But, they should be independently analyzed and independently reported.

Coefficient Error Types

There are coefficients (numerical values) in equations. For some, the value is known with certainty. For some, there is a systematic (always the same) bias, and for others there is a random bias. Finally, for others there is random uncertainty. You need to be able to differentiate the types.

Coefficients with certain values would include the 100 when converting a fraction or proportion to a %, or the ½ that converts diameter to radius, or the 4 that multiplies the length of one side of a square to obtain the perimeter. There is no need to propagate uncertainty on such variables.

However, some apparently fundamental coefficients are truncated or rounded, and accordingly have uncertainty. If for instance you substitute 3.14 for the value of π then you are deciding to include a bias of 0.0015926535… Similarly if you decide to use 18 for the molecular weight of water, then you are including a bias of 0.01528 (based on ^{12}C isotope). Those errors are systematic bias errors. But, the value of the systematic error can be known. Since they are known, you could correct the error in the calculated values that they create. But, why use the wrong value, then post-correct the calculated value? The proper approach is to use the right value to begin with. However, many coefficients are irrational numbers and

it is not possible to use an infinite set. The rule of thumb is to use the number of places that makes the systematic bias negligible in the calculated value. Perhaps use enough digits so that the error on the model prediction is two or three orders of magnitude less than the uncertainty due to other contributors.

Similarly, do not truncate or round values of intermediate calculations to use the shortened version (systematic bias) in subsequent calculations. Keep two or three more digits than would be justified in reporting the value. Using a truncated value, would impose a bias to subsequent calculations, but the bias would be random, because it would change in subsequent applications due to the vagaries of the input data. However, you should report the shortened version to reveal your appreciation for significant digits.

Similarly, when numerical methods of root finding in implicit relations or optimization are used to determine a value, be sure that the convergence criterion is so small that the uncertainty it imparts to the solved for value is inconsequential to the result.

"Givens" have uncertainty. These are the basis, forecast loads, or assumptions in an application statement. Although, the givens may not seem uncertain when 'given' by authority. For example, the boss may say, "Twenty people are coming to the meeting. Be sure we have 20 seats." Or, the client might direct, "Design the process to produce 100 Kg of product per day, when the raw material is 23 wt-% useful." Don't be misled by authority or a problem statement to think that the givens have zero uncertainty.

Of course coefficients related to material properties (viscosity, molecular weight, thermal expansion, half-life, growth rate), fundamental constants (speed of light, gas law constant), unit conversions (kPa to psia, miles to km), material quantity (lbs per package, product purity), economics (tax rate, labor cost, utility price), probability (reliability, variance), and coefficients in equations ($Nu=0.023Re^{0.8}Pr^{0.4}$) are all uncertain. Don't accept them as absolute truth.

PROPAGATION OF UNCERTAINTY IN MODELS

There are two common metrics of the uncertainty: maximum error and probable error. For each, there are two methods to propagate uncertainty: analytical and numerical. And, for each there is systematic error and random error. ASTM recommends that systematic error and random error be propagated by the same methods, but independently, and separately reported. What follows is how to propagate uncertainty in general for either systematic or random error.

In general, consider this model:

$$w = f(x, y, z)$$

The input or independent variables are x, y, and z. However, they could be listed as x_1, x_2, and x_3. The independent variables, the inputs could also be considered the "givens", the basis for the calculation or the properties of materials. And there is no requirement that there are three inputs. There could be 1, or 2, or 10, or n, and represented as

$$w = f(x_1, x_2, x_3, x_4, \ldots) = f(\underline{x})$$

If you have more or fewer, adjust the formulas that follow. The response or dependent variable in this presentation is w. The influence variables are more than just experimental conditions, they include model coefficients and givens – any variable with uncertainty.

The objective is to develop a method to use the model and uncertainties on the input variables to predict the uncertainty on w, then to compare the model-predicted uncertainty to the actual experimentally obtained uncertainty.

Analytical Method for Maximum Uncertainty

A Taylor series approximation to the model about nominal (x, y, z) values is

$$w = f(x, y, z) = f(x_0, y_0, z_0) + \left.\frac{\partial f}{\partial x}\right|_0 (x - x_0) + \left.\frac{\partial f}{\partial y}\right|_0 (y - y_0) + \left.\frac{\partial f}{\partial z}\right|_0 (z - z_0) + \cdots$$

This is truncated to exclude quadratic and higher order terms, which means that (x, y, z) deviations from the base case values of (x_0, y_0, z_0) are small. Note that the derivatives are evaluated at the base case (x_0, y_0, z_0).

Representing deviations as $\varepsilon_x = (x - x_0)$, and removing the explicit notation on the derivatives

$$\varepsilon_w = f(x, y, z) - f(x_0, y_0, z_0) = \frac{\partial f}{\partial x}\varepsilon_x + \frac{\partial f}{\partial y}\varepsilon_y + \frac{\partial f}{\partial z}\varepsilon_z$$

Maximum uncertainty on the dependent variable would happen when each independent error is at its extreme value and each is "pushing" the independent variable in the same direction. So, use the absolute values and let ε_x represent the maximum expected deviation, not a particular realization.

$$\varepsilon_w = \left|\frac{\partial f}{\partial x}\varepsilon_x\right| + \left|\frac{\partial f}{\partial y}\varepsilon_y\right| + \left|\frac{\partial f}{\partial z}\varepsilon_z\right|$$

In general

$$\varepsilon_w = \sum_{i=1}^{n} \left|\frac{\partial f}{\partial x_i}\varepsilon_{x_i}\right|$$

This presumes

1. The impact of each independent variable has a linear impact on the dependent variable (which is usually acceptable, if the deviation is small relative to any curvature effects),
2. That the sensitivity (partial derivative) of one response is independent of the value of other inputs,
3. The (x, y, z) deviations are independent (uncorrelated),
4. The derivatives are all evaluated at the same base case $(x_{1_0}, x_{2_0}, x_{3_0}, \ldots)$,
5. That all influences are at their maximum value, and
6. That all influences are pushing the response in the same direction.

A few simple examples reveal some caution issues for those a bit rusty in Calculus I: For the yield stress model, $S = L/dw$,

$$\varepsilon_S = \frac{1}{dw}\varepsilon_L + \frac{L}{d^2w}\varepsilon_d + \frac{L}{dw^2}\varepsilon_w$$

Note that the negative signs due to the derivatives w.r.t. d and w became positive after the absolute value, and that each term in the equation has the same units of stress.

For the bucket and stopwatch method, $F = (W_{full} - W_{empty})/\left[\rho(t_{full} - t_{start})\right]$,

$$\varepsilon_F = \frac{1}{\rho(t_{full} - t_{start})}2\varepsilon_W + \frac{(W_{full} - W_{empty})}{\rho^2(t_{full} - t_{start})}\varepsilon_\rho + \frac{(W_{full} - W_{empty})}{\rho(t_{full} - t_{start})^2}2\varepsilon_t$$

Note that the factor of 2, is the result of the two weight and two time measurements having independent variation with identical uncertainty and each having the same absolute value of the derivatives.

For the orifice flow meter, $F = a(i - i_0)^b$, in use there would be noise (uncertainty) on the transmitted mA, but also an uncertainty due to the development of the model. One could consider that the three model coefficients (a, b, i_0) are independently uncertain, but they would be correlated, so I'll lump the model uncertainty into one term such as would be quantified as the 95% limit in a bootstrapping analysis. Here, $F = a(i - i_0)^b + \varepsilon_{F-model}$. Then,

$$\varepsilon_F = ab(i - i_0)^{b-1}\varepsilon_i + \varepsilon_{F-model}$$

Note the exponent decrement $b - 1$ in the relation. Also note, if one were to consider uncertainty on the coefficient b, then there would be a term $aln(i - i_0)(i - i_0)^b\varepsilon_b$ in the expansion for ε_F.

Again, don't be misled by the nomenclature to limit x to experimental conditions. In the above, x represents any variable with uncertainty (givens, coefficient values, basis, experimental measurements, etc.)

This also presumes that you can take the analytical derivative. Many common models, like the simple ones above, stretch the calculus ability that you were supposed to have learned in the Calculus 1 class, many years ago. Further, many models are so complex that even with skill, there is a good chance of making an error in many lines of the calculus and algebra associated with getting the analytical equation of a derivative. If so, the analytical derivatives can be replaced by a finite difference estimate. If using the central difference formula:

$$\frac{\partial f}{\partial x} \cong \frac{f(x + \Delta x, y, z) - f(x - \Delta x, y, z)}{2\Delta x}$$

You choose the Δx value. It should be small relative to the curvature in the function. I often use $\Delta x = 0.1\varepsilon_x$. But, it needs to be not too small, so that numerical truncation error does not compromise the derivative estimate. Accordingly, I like to use double precision variables in the calculations.

Except in trivial cases, my default is to use the numerical approximation for the derivative.

Analytical Method for Probable Uncertainty

It is unlikely that each variable will be at its extreme and pushing in the same direction. More likely, the perturbation values are normally (Gaussian, bell-shaped) distributed, and independent of each other. In this case, propagation of variance will provide an estimate of the probable error.

I've seen about 4 ways to derive this. Here is my favorite. Start with the definition of variance

$$s_w{}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(w_i - \bar{w})^2$$

If the average is the true value at the base case, $\bar{w} = f(x_0, y_0, z_0)$, and if the linear Taylor series approximation is valid,

$$s_w{}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{\partial f}{\partial x}\varepsilon_{x_i} + \frac{\partial f}{\partial y}\varepsilon_{y_i} + \frac{\partial f}{\partial z}\varepsilon_{z_i}\right)^2$$

Recall that derivatives are evaluated at the base case, not at some i-th value.

Expanding the squared argument of the sum and explicitly showing some of the terms

$$s_w{}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\left(\frac{\partial f}{\partial x}\varepsilon_{x_i}\right)^2 + \left(\frac{\partial f}{\partial y}\varepsilon_{y_i}\right)^2 + \cdots + \frac{\partial f}{\partial x}\varepsilon_{x_i}\frac{\partial f}{\partial z}\varepsilon_{z_i} + \cdots\right]$$

And regrouping the sum

$$s_w{}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{\partial f}{\partial x}\varepsilon_{x_i}\right)^2 + \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{\partial f}{\partial y}\varepsilon_{y_i}\right)^2 + \cdots + \frac{1}{N-1}\sum_{i=1}^{N}\frac{\partial f}{\partial x}\varepsilon_{x_i}\frac{\partial f}{\partial z}\varepsilon_{z_i} + \cdots$$

The squared terms should be recognized as the variance estimate on each input. The cross product terms are called co-variance, and measure the correlation between independent variable values. If the variation in each input variable is independent then the co-variance terms will have as many + values as – values and the sum of them will tend to remain about a value of zero. The squared terms, however, will always be positive and the sum will be relatively large. If there is no correlation, then the co-variance terms can be ignored, and conventionally representing sample standard deviation with the population value, propagation of variance is:

$$\sigma_w = \sqrt{\left(\frac{\partial f}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial f}{\partial y}\sigma_y\right)^2 + \left(\frac{\partial f}{\partial z}\sigma_z\right)^2}$$

In general

$$\sigma_w = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\sigma_{x_i}\right)^2}$$

If you do not know the values of σ_i you could estimate them from $\sigma_i \cong \varepsilon_i/2.5$. If you cannot analytically take the derivatives, then use the numerical approximation.

This also presumes that

1. The impact of each independent variable has a linear impact on the dependent variable (which is usually acceptable, if the deviation is small relative to any curvature effects),
2. That the sensitivity (partial derivative) of one response is independent of the value of other inputs,
3. The (x, y, z) deviations are independent (uncorrelated),
4. N is large enough to make the covariance sums become negligible, and
5. The derivatives are all evaluated at the same base case $(x_{1_0}, x_{2_0}, x_{3_0}, \dots)$.

Contrasting propagation of maximum error, this does not presume that

1. That all influences are at their maximum value, and
2. That all influences are pushing the response in the same direction.

Using probable error as the 95% interval

$$\varepsilon_{w,0.95} = 1.96\sigma_w$$

However, the estimates on the σ_i values are probably +/- 50% of the true value, which does not justify three digit precision on the 1.96, justifying this relation

$$\varepsilon_{w,0.95} \cong 2\sigma_w$$

If wanting the 99% probable error use

$$\varepsilon_{w,0.99} = 2.5\sigma_w$$

Since the 2.5 coefficient is the same in the estimating relation between error and sigma, this reduces to a propagation of errors (as opposed to variance).

$$\varepsilon_{w,0.99} \cong \sqrt{\left(\frac{\partial f}{\partial x}\varepsilon_x\right)^2 + \left(\frac{\partial f}{\partial y}\varepsilon_y\right)^2 + \left(\frac{\partial f}{\partial z}\varepsilon_z\right)^2}$$

Or generically:

$$\varepsilon_{w,0.99} \cong \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\varepsilon_{x_i}\right)^2}$$

Again, considering the uncertainty on the σ_i or ε_i values, and variability in measured σ_w; if a model-predicted $\varepsilon_{w,0.99}$ is within a 4:1 or 0.4:1 ratio of the measured $\varepsilon_{w,0.99}$ then the analysis cannot be rejected. However, if outside of the 4:1 or 0.4:1 ratio, question the model or the sources or estimates of variation.

The propagation of variance will provide a more realistic estimate of probable error than the propagation of maximum error. However, often when safety or risk is involved, people want to know what might be the worst case scenario, a possible worst outcome.

Numerical Method for Maximum Uncertainty

The analytical method presumes that sensitivities of one variable are independent of the value of other variables, that influences are linear, and that you can obtain values for the sensitivities (derivatives). It also assumes that the individual uncertainties are normally distributed and independent of each other. The numerical procedure does not impose such conditions.

Here is pseudo-code for an approach to numerically propagate maximum uncertainty. It is a simple exhaustive search.

1. Assign base case values to each x_i variable, and ε_i, the maximum expected error, to each input variable.
2. Perturb the base case with each possible combination of $+\varepsilon_i$, 0, and $-\varepsilon_i$. With 3 variables and three possible perturbed values for each, there are $3^3 = 27$ cases. With n variables and three possible perturbations for each, there are 3^n number of combinations. Calculate the w-value for each combination.
3. Search through the 3^n w-values to find the maximum and minimum.

If the function is nonlinear and the ε_i values are large relative to curvature, it is possible that the extreme w-value could be at some intermediate input value, not one of the extremes, or zero. You could use a 9-point discretization of the input error: Zero, $\pm 0.25\varepsilon_i$, $\pm 0.50\varepsilon_i$, $\pm 0.75\varepsilon_i$, and $\pm 1.00\varepsilon_i$. Or, any such exploration that seems appropriate. The 9-point analysis would require 9^n cases.

Numerical Method for Probable Uncertainty

Again, the analytical method presumes that sensitivities of one variable are independent of the value of other variables, that influences are linear, and that you can obtain values for the sensitivities (derivatives). It also assumes that the individual uncertainties are normally distributed and independent of each other. The numerical procedure does not impose such conditions.

Again, propagation of maximum error provides the worst possible outcomes, but it is unlikely that each of the many input errors are at their extreme values and simultaneously pushing the response in the same direction. Intermediate confluences of perturbations are much more likely. When there are several (or more) independent influences, the probable error will be more representative of the outcome uncertainty.

Here is pseudo-code for a Monte Carlo approach to numerically propagate probable uncertainty.

1. Assign base case values to each x_i variable.
2. Perturb each base case x_i to a possible value within the probable range, a realization. Use whatever distribution is appropriate. If uncertain, use the normal distribution.
3. Calculate w from the perturbed x_i values.
4. Record the possible w-value from that realization.
5. Repeat Steps 2, 3, & 4 a large number of times. 100 is a good starting number, for the number of realizations, but you might find that the σ_w value is noticeably changing with sequential realizations. You might need 100,000 realizations to have the σ_w value settle to a steady value (unchanging with subsequent realizations). Observe how σ_w changes with additional realizations, and stop when its value seems to have settled.
6. Calculate the desired statistic from the list of many w-values. This could be the maximum and minimum values, the 99% limits, the 95% limits, quartiles, the standard deviation, etc.

In Step 2, if you want to create a uniformly distributed perturbation over the possible range then use

$$x_{i,j} = x_{i\,base} + 2(r_j - 0.5)\varepsilon_{x_i}$$

Here, r is a uniformly distributed random number in the 0-1 interval, and independent for each realization. The subscript j indicates realization number. Do the same for each influence.

However, influences are more likely to be Gaussian distributed. In this case I think the Box-Muller formula is a good approach.

$$x_{i,j} = x_{i\,base} + \sigma_{x_i}\sqrt{-2\ln\,(r_{1,j})}\sin\,(2\pi r_{2,j})$$

Here, r_1 and r_2 are each independent uniform random numbers, and independent for each variable and each realization. Here σ_x could be estimated as $\varepsilon_x/2.5$. If your random number generator provides r-values in the range $0 \le r < 1$, then it is possible to have the computer attempt to calculate $\ln\,(0)$. So, use $1 - r$ in the calculation rather than r.

If you feel that the distribution is Poisson, or log normal, or other, it is not too difficult to make perturbations from those. Use the random number generator to generate a random number, assign that as the CDF value of the distribution, then use the inverse of the distribution to get the random variable value. See a reference, such as Law and Kelton, Simulation Modeling and Analysis, 2nd Edition, McGraw Hill, 1991.

Also in Step 2, you could add a cross correlation to the independent variables. For instance, if you believe that what makes y high (perhaps humidity or operator technique) also makes z high, then you could calculate the perturbation on z from that on y

$$z = z_{base} + \frac{\sigma_z}{\sigma_y}(y - y_{base})$$

There are hundreds of options to model forms of correlation, such as including independent variation with the correlated perturbation, autocorrelation in time-series of values, and adjusting sigma values with variable value or other variables.

I like this numerical and propagation of variance approach best, and use of the Box-Jenkins formula. However, it does require some computer programming, but that is not very complex. Although it does not seem as mathematically sophisticated as the calculus-derived formulas of the analytical approaches, it has fewer assumptions about linear and independent relations, and it is much easier to change the underlying distribution of variation and to add correlated variation.

Key Sources of Uncertainty

One of the uses of uncertainty analysis is to determine what variables to work on to reduce variability of the dependent variable, to improve uniformity, to improve quality. Variance propagated is

$$\sigma_w{}^2 = \left(\frac{\partial f}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial f}{\partial y}\sigma_y\right)^2 + \left(\frac{\partial f}{\partial z}\sigma_z\right)^2$$

And the fractional impact of x variation on the total is

$$\frac{\sigma_{x\ on\ w}{}^2}{\sigma_{total\ w}{}^2} = \frac{\left(\frac{\partial f}{\partial x}\sigma_x\right)^2}{\left(\frac{\partial f}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial f}{\partial y}\sigma_y\right)^2 + \left(\frac{\partial f}{\partial z}\sigma_z\right)^2}$$

Do this for the y and z elements, also. The values will sum to unity. If something has less than 0.1 relative value, then there is no justification to seek ways to reduce its impact on the dependent variable uncertainty. Focus on reducing the variation of the input variable that has the largest relative impact.

If you want to assess the individual impact in a numerical approach, then make one of the ε_{input} values equal to zero, and repeat the realizations and analysis with realizations of the others. The reduction in ε_w, or σ_w would indicate the impact if the ith source of variation could be entirely eliminated. You could halve the ε_{input} value (or make another choice based on what you think might be a possible reduction in an influence), and look at the reduction in ε_w, or σ_w. Do this for each of the input variables to see where it is most to invest effort to seek improved quality.

A Special Case of Composite Uncertainty

Suppose there are several factors that contribute to uncertainty of one variable. For example, variation in flow rate may affect a heat transfer model. But variation in flow rate may be related to bucket and stopwatch calibration error. Or, width may be a factor in a stress model, but the measurement of width may be subject to perturbations in micrometer reading, micrometer tightness, off-normal angle of the micrometer, or sample location of the measurement. In these cases the composite uncertainty is the sum of each component uncertainty. Use either propagation of error or variance on the sum.

A Special Case of Uncontrolled Environmental Uncertainty

Alternately, for example, if a sample is affected by humidity (%RH), then one needs a model of how %RH affects the measurement, then can use that model to account for uncertainty due to the environmental elements. For instance, in calculating tensile breaking stress, the equation is S=L/A, but this does not include %RH. So, do experiments with "identical" material, and determine how L is affected by %RH. Perhaps generate an empirical equation, $L = L_0 + a(\%RH) + b(\%RH)^2$. Now the a and b factors can be used to propagate uncertainty. $\sigma_L = \sqrt{{\sigma_{L0}}^2 + [(a + 2b(\%RH0)\sigma_{\%RH}]^2}$.

PROPAGATING UNCERTAINTY OF DATA ONTO REGRESSION MODELS

Bootstrapping

In regression, model coefficients are selected to minimize the deviations between data and model. It could be a purely empirical model with many coefficients, or one that was developed from a phenomenological (mechanistic, either first-principles or rigorous) approach with only a few adjustable coefficients. Here, one set of experiments provides the data to best fit the model to data. However, a replicate set of experiments will provide similar but different data, due to the experimental vagaries, providing different model coefficient values. The question is, "How does irreproducibility in the data affect the uncertainty on the model?" Bootstrapping is a technique to estimate the uncertainty in model predictions due to uncertainty in the experimental data.

One assumption in bootstrapping is that the experimental data that you have represents the entire population of all data realizations, including all nuances in relative proportion. It is not the entire possible population of infinite experimental runs, but it is a surrogate of the population. A sampling of that experimental data then, represents what might be found in an experiment. Another assumption is that the model cannot be rejected by the data, that the model expresses the underlying phenomena. In bootstrapping:

1. Sample the experimental data with replacement (retaining all data in the draw-from original set) to create a new set of data. The new set should have the same number of items in the original, but some items in the new set will likely be duplicates, and some of the original data will be missing. This represents an experimental realization from the surrogate population.
2. Determine the model coefficient values that best fit the data set realization from Step 1. This represents the model that could have been realized. It does not matter whether the objective function is vertical or total least squares, SSD or rms, maximum likelihood, etc. Use whichever is appropriate.
3. Record the model coefficient values.
4. For independent variable values of interest, determine the modeled response. You might determine the y-value for each experimental input x-set. This would be simple. Alternately, if the model is needed for a range of independent variable values, you might choose key x-values within the range and calculate the model y for each x-set. This is likely more useful.
5. Record the modeled y-values for each of the desired x-values.
6. Repeat Steps 1-5 many times (perhaps over 1,000, perhaps 100,000).
7. For each desired x-set, create the CDF of the model predictions. This will reflect the distribution of model prediction values due to the vagaries in the data sample realizations. The variability of the prediction will indicate the model uncertainty due to the vagaries within the data.

8. Choose a desired confidence interval value. The 95% range is commonly used.
9. Use the cumulative distribution of model predictions to estimate the confidence interval on the model prediction. If the 95% interval is desired, then the confidence interval will include 95% of the models; or, 5% of the modeled y-values will be outside of the confidence interval. As with common practice, split the too-high and too-low region of values into equal probabilities of 2.5% each, and use the 0.025 and 0.975 CDF values to determine the y-values for the 95% confidence interval. With 1,000 trials, the 25^{th} and 975^{th} values represent the 95% interval.

Visit http://www.r3eda.com/bootstrapping-uncertainty-assessment/ to obtain an Excel VBA file to implement the technique.

This bootstrapping approach presumes that the original data has enough samples covering all situations so that it represents all features of the entire possible population of data. Then the new sets (sampled with replacement) represent legitimate realizations of sample populations. Accordingly, the distribution of model prediction values from each re-sampled set represents the distribution that would arise if the true population were independently sampled.

If you are seeking to model transient data, you might have set up one situation then taken 100 measurements over time. But, this is one realization, not 100. All ensuing 100 samples are related to the same initialization. If the set-up would have mixing, composition, temperature, etc. variation from one set-up to another; then the 100 sequential samples do not represent that true variation. They are simply one realization. You should run many replicate trials of the 100 responses to include the set-up to set-up variability.

Similarly, if you create a large sample, or set-up a situation, then perform multiple tests on the same material (or set-up) all of the tests reflect a single set-up. These replicate part of the source of variability, but not all of it. The data excludes the variation due to set-up, and is in effect a single realization. For example, in rheology we make a mixture, place a sample of it in a device, and measure shear stress for a variety of shear rates (perhaps 20 different shear rates). Then we plot stress w.r.t. shear rate and use regression to match a rheology model to the 20 sets of data. But, there is error in the sample preparation, and the 20 sets of data represent one sample preparation realization, not the population. To make the data reflect the population, perform the tests on replicate set-ups (material preparations).

Bootstrapping assumes: the limited data represents the entire population of possible data, that the experimental errors are naturally distributed (there are no outliers or mistakes, not necessarily Gaussian distributed, but the distribution represents random natural influences), and that the functional form of the model matches the process mechanism. Then a random sample from your data would represent a sampling from the population; and for each realization, the model would be right.

If there are N number of original data, then sample N times with replacement. Since the Central Limit Theorem indicates that variability reduces with the square root of N, using the same number keeps the variability between the bootstrapping samples consistent with the original data. In Step 1, the assumption is that the sample still represents a possible realization of a legitimate experimental test of the same N. If you use a lower number of data in the sample, M, for instance, then you increase the variability on the model coefficient values. You could accept the central limit theorem and rescale the resulting variability by square root of M/N. But, the practice is to use the same sample size as the “population”, to reflect the population uncertainty on the model.

In Step 6, if only a few re-samplings, then there are too few results to be able to claim the variability range with certainty. As the number of Step 6 re-samplings increases the Step 9 results will asymptotically approach the representative 95% values. But, the exact value after infinite re-samplings is not the truth, because it simply reflects the features captured in the surrogate population of the original N data, which is not actually the entire population. So, balance effort with precision. Perhaps 100 re-samplings will provide consistency in the results. On the other hand, it is not unusual to have to run 100,000 trials to have Monte Carlo results converge. After 100 or so re-samplings observe how the CDF evolves, and keep adding bootstrapping trials until the CDF reasonably settles.

Bootstrapping assumes that the data have no systematic bias. If there is a systematic bias, then all of the data would be shifted up or down. Bootstrapping is an analysis of random fluctuation. ASTM advises to report the two independently. However, if you want to determine the joint impact of random and systematic errors on the model prediction then use the square root of the sum of variances.

One can estimate the number of re-samplings, n, needed in Step 6 for the results in Step 9 to converge from the statistics of proportions. From a binomial distribution the standard deviation on the proportion, p, is based on the proportion value and the number of data:

$$\sigma_p = \sqrt{p(1-p)/n}$$

Desirably, the uncertainty on the proportion will be a fraction of the proportion:

$$\sigma_p = fp$$

Where the desired value of f might be 0.1.

Solving for the number of data required

$$n = \left(\frac{1}{p} - 1\right)/f^2$$

If p=0.025 and f=0.1, then n≈4,000.

Although n=10,000 trials is not uncommon, and n=4,000 was just determined, I think for most engineering applications 100 to 1,000 re-samplings will provide an appropriate balance between computational time and precision. Alternately, you might calculate the 95% confidence limits on the y-values after each re-sampling, and stop computing new realizations when there is no meaningful progression in its value, when the confidence limits seem to be approaching a noisy steady state value.

In Step 9, if you assume that the distribution of the $\tilde{y}$-predictions are normally distributed, then you could calculate the standard deviation of the $\tilde{y}$-values and use 1.96 times the standard deviation on each model prediction to estimate the 95% probable error on the model at that point due to errors in the data. Here, the term error does not mean mistake, it means random experimental normal fluctuation. The upper and lower 95% limits for the model would be the model value plus/minus the probable error. This is a parametric approach.

By contrast, searching through the n=1,000, 4,000, or 100,000 results to determine the upper and lower 97.5% and 2.5% values is a non-parametric approach. The parametric approach has the advantage that it uses values of all results to compute the standard deviation of the $\tilde{y}$-prediction realizations, and can get relatively accurate numbers with much fewer number of samples. Perhaps, n=20. However, the parametric approach presumes that the variability in $\tilde{y}$-predictions is Gaussian (bell-shaped and symmetric). It might not be. The nonparametric approach does not make assumptions about the underlying distribution, but only uses 2 samples to interpolate each of the $\tilde{y}_{0.025}$ and $\tilde{y}_{0.975}$ values. So, it requires many trials to generate truly representative confidence interval values.

Unfortunately, the model coefficient values are likely to be correlated. This means, if one value needs to be higher to best fit a data sample, then the other will have to compensate (perhaps be lower). If you plot one coefficient w.r.t. another for the 100 re-samplings and see a trend, then they are correlated. When the variability on input data values are correlated, the classical methods for propagation of uncertainty due to coefficient uncertainty are not valid. They assume no correlation (no co-variance) in the independent variables in the propagation of uncertainty.

Also, Step 9 has the implicit assumption that the model matches the data, that the model cannot be rejected by the data, that the model expresses the underlying phenomena. If the model does not match the data, then bootstrapping still will provide a 95% confidence interval on the model; but you cannot expect that interval to include the 95% of the data. As a caution: If the model does not match the data (if the data rejects the model) then bootstrapping does not indicate the range about your bad model that encompasses the data, the uncertainty of your model predicting the true values.

The figure reveals the results of a Bootstrapping analysis on a model. The circles represent data, the inside thin line is the modeled value from the entirety of original $N = 15$ data, and the darker lines indicated the 95% limits of the model based on 100 realizations of the data set.

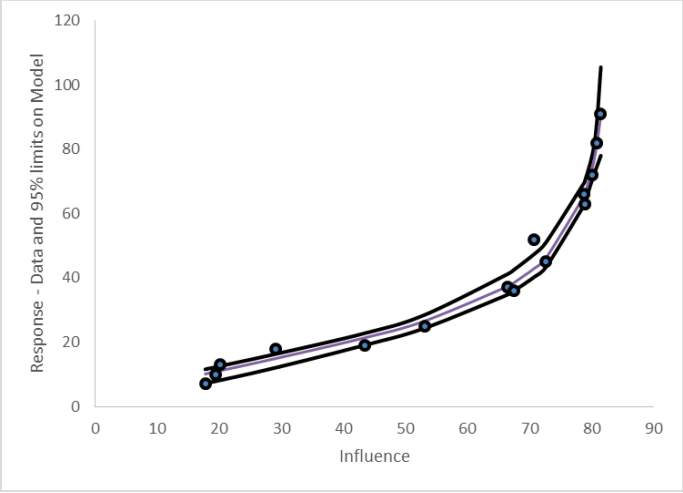

Figure: Bootstrapping Estimate of Model Uncertainty Due to Data

If the experimental procedure is valid, the deviations between replicate sets would be small and the model-to-model prediction variability would be small, but it would exist.

If the model matches the underlying phenomena, then ideal natural experimental vagaries (the confluence of many, small, independent, equivalent sources of variation) should result in residuals that have a normal (Gaussian, bell-shaped) distribution. If this is the situation 1) model matches phenomena, 2) normally distributed residuals, and 3) model coefficients are linearly expressed in the model, and 4) experimental variance is uniform over all of the range (homoscedastic), then analytical statistical techniques have been developed to propagate experimental uncertainty to provide estimates of uncertainty on model coefficient values, and on the model. It gives the 95% probable range, or so, for the model. However, if the variation is not normally distributed, if the model is nonlinear in coefficients, if variance is not homoscedastic, or the model does not exactly match the underlying phenomena, then the analytical techniques are not applicable. In this case numerical techniques are needed to estimate model uncertainty. Bootstrapping is the one I prefer. It seems to be understandable, legitimate, simple, and is widely accepted

The advantage of bootstrapping over conventional propagation of uncertainty is that you do not have to estimate the uncertainty (error) or make assumptions about error distributions on individual elements in the model. Bootstrapping uses the uncertainty in the data, as Nature decided to present it; and provides model-prediction uncertainty corresponding to the data uncertainty.

However, bootstrapping does not account for the component of uncertainty that would be contributed by your estimates of coefficient values (such as the gas law constant, the speed of light, the value of Pi, a tabulated viscosity, etc.), or givens (heat exchanger duty, production rate, etc.). Use bootstrapping to determine the impact of experimental uncertainty on the model prediction, ε_y. Then use propagation of uncertainty to combine that with a propagation of uncertainty from model parameters and givens to generate an estimate of total model error.

Bootstrapping generates a set of model coefficient values, one for each data sampling realization. The variability or range in individual coefficient values can be an indication of the sensitivity of the coefficient to the data. A model coefficient that has a large range, perhaps relative to its base case value, could indicate any of several features: 1) The model parameter has little impact on the model; so, the specific phenomena that it represents should be reconsidered, and either modeled differently or the inconsequential phenomenal removed. 2) The model parameter is sensitive to the data variability, and experimental design should be reconsidered to generate data with sufficient precision.

In bootstrapping, the model coefficient values will be correlated. In a simple case consider a linear (y, x) model, $y = a + bx$. If a best model for a sampling has a high intercept, it will have a low slope to compensate and keep the model in the proximity of the other data. Since the parameter values are correlated, one cannot use the range (or alternate measures of variability) of the parameter values from bootstrapping to individually estimate the uncertainty on the model due to the parameter value. Estimate model uncertainty from the ensemble – each of the N model predictions from each of the N sets of coefficient values.

TEST FOR VARIANCE EXPECTATIONS

If the model functionality matches the process, and the experiment to generate data is properly performed and understood, then the residuals (model to data deviations) should have a variance that matches the propagation of variance on the experimental data model.

The variance in the residuals should match that expected from propagation of uncertainty in the data. One could compare the range of residuals to either replicate trials or the propagation of probable error on the data model (the method used to calculate a data value from measurements). If the residual range is much larger than expected, this could imply that the process model is wrong, or that the experiment was not controlled or understood as expected. If the residual range is much smaller than expected, this could mean that the experimental uncertainty was much smaller than expected. In any case, something was not properly represented.

A statistically proper way to compare variances is with an F-statistic. However, for the comparison of residuals to those expected from a propagation of variance, Item 1 above, the propagation of variance is just an estimate based on linearization, independence, and human estimates of several of the component uncertainties. It is a reasonable estimate not the truth. And, my experience has usually been that the number of data in question is too low (too few residuals) to lead to a definitive test, and the human choice as to how to partition the data could lead to challenges.

Although use of an F-statistic to test variances might be the "should", either test would be infected with human choices. Accordingly, I think that a human judgment as to whether the residual variance matches the expectation and whether the data is homoscedastic best balances the perfection/sufficiency values.

Probable error from propagation of uncertainty on data model should be equivalent to the standard deviation (or effectively the rms) of the residuals. The F-test indicates that a 4:1 ratio might not be unexpected for experiments of $N = 15$ or so data. So, this test will only be revealing if the ratio is larger than about 5:1. If the expected probable error is much smaller than the rms then you may be underestimating your experimental uncertainty, or the model might not be matching the data leading to large deviations. If the expected probable error is much larger, then you may be overestimating the experimental uncertainty, your model may have too many adjustable coefficients and be fitting the noise, or you might not have reached steady-state when you took the data. This is a check on your understanding of the data generating process.

PROPAGATION OF UNCERTAINTY ON OPTIMUM DV VALUES

The question is, "How does uncertainty on model coefficients affect the value of x^* and consequently OF^*?" Several approaches to propagating uncertainty, are either analytical or numerical. If the models are analytically differentiable then it can be done analytically. More often a numerical approach is needed.

Contrasting the purpose to propagate uncertainty to the DV* values, most propagation of uncertainty relates to a model, $y = f(\underline{x}, \underline{c})$, and addresses the issue of uncertainty on y, due to uncertainty on the c coefficients or the x influences. Here, however, the interest is in the uncertainty on x^*, not on $y = f(\underline{x}, \underline{c})$.

Analytical Method

Consider the optimization statement

$$\min_{\{x\}} J = f(x, c)$$

In which x is the DV and c is a model coefficient that has an uncertain value. At the optimum

$$\left.\frac{\partial f}{\partial x}\right|_{x^*} = g(x^*) = 0$$

This figure reveals the derivative of the OF as the solid line, and the dashed line represents the shift in the derivative due to a possible alternate coefficient value.

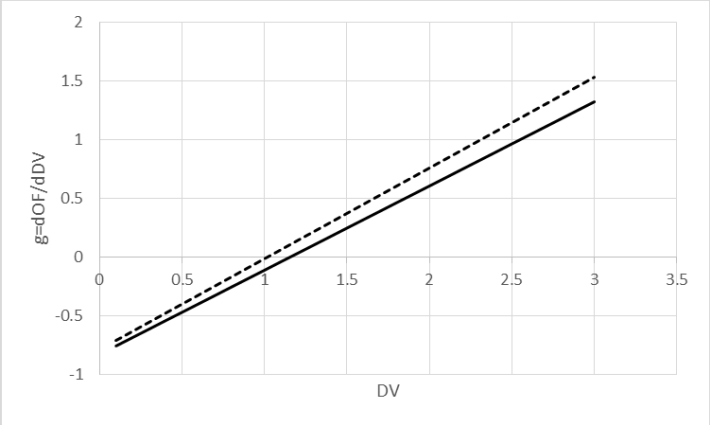

Figure. Shift in the derivative due to uncertainty in model coefficient values

Effectively the slope of the $g(x, c)$ curve is unchanged with coefficient value, but at the original x^* value, x^*_0, for which $g(x^*_0, c) = 0$ the value of the new $g(x^*_0, c + \Delta c)$ is not zero. Assuming the slope of the $g(x, c)$ curve is unchanged with coefficient value, the shift in x^* value due to the shift in c value can be estimated as

$$\Delta x^* \cong \frac{g(x^*_0, c + \Delta c)}{\frac{\partial g(x^*_0, c)}{\partial x}} \cong \left.\frac{\frac{\partial g}{\partial c}\Delta c}{\frac{\partial^2 f}{\partial x^2}}\right|_{x^*_0, c} = \left.\frac{\frac{\partial}{\partial c}\frac{\partial f}{\partial x}}{\frac{\partial^2 f}{\partial x^2}}\right|_{x^*_0, c} \Delta c$$

Representing the finite differentials with an uncertainty value of half the uncertainty range, $\varepsilon = \pm(R/2)$

$$\varepsilon_{x^*} \cong \left.\frac{\frac{\partial}{\partial c}\frac{\partial f}{\partial x}}{\frac{\partial^2 f}{\partial x^2}}\right|_{x^*_0, c} \varepsilon_c$$

If there are multiple coefficients with uncertain value, and they have independent effects, and local linearization of each is valid, then using propagation of maximum uncertainty:

$$\varepsilon_{x^*} \cong \sum \left\{ \left| \left. \frac{\frac{\partial}{\partial c_i}\frac{\partial f}{\partial x}}{\frac{\partial^2 f}{\partial x^2}} \right|_{x^*_0,c} \varepsilon_{c_i} \right| \right\}$$

Using propagation of the 95% probable uncertainty

$$\varepsilon_{x^*,0.95} \cong \sqrt{\sum \left\{ \left(\left. \frac{\frac{\partial}{\partial c_i}\frac{\partial f}{\partial x}}{\frac{\partial^2 f}{\partial x^2}} \right|_{x^*_0,c} \varepsilon_{c_i} \right)^2 \right\}}$$

Note: This could be a trivially simple analysis, as it is for a power series model, but generally, determining the derivative values is more complicated. The derivative values can be estimated by numerical, finite difference, methods. But, even if using numerical estimates for the derivatives, this is still the analytical method, with the local linear and independent assumptions.

Numerical Method

Numerical methods to propagate uncertainty are not dependent on the condition assumptions of the analytical method – independent (uncorrelated), symmetric (distribution of error), and small (locally linear effects). They reveal the distribution of outcomes, not just a maximum or probable error. This makes them more appropriate tools. However, they require computer programming methods, which requires additional skill.

The procedure would be:

1. Define the distribution (range and functionality) of the givens (model and coefficients).
2. Sample a realization of coefficient values to create the possibility of an OF model. A realization is a possible coefficient set. You might sample from a uniform distribution, normal distribution, or other. You might have coefficients independently perturbed, or correlated in some manner. What you choose depends on how you understand the application.
3. For that realization, determine the DV* and OF* values. Use any appropriate optimization method, with convergence criterion that contributes orders of magnitude smaller error than the impact of the uncertainty in the givens.
4. Repeat Steps 2 and 3 for many realizations, until the distribution of DV* and OF* values seems stationary. This may be 100 realizations (and corresponding optimizations) for a model that just has uncertainty in one or two givens; but more frequently, it will be on the order of 1,000 to 100,000 realizations.
5. From the list of DV* and OF* values, select appropriate statistics such as the 95% extremes, upper and lower quartiles, or such for reporting the results.
6. Include the range, uncertainty, variability, when reporting the results. Don't imply that one DV* is the right value. Reveal the impact of the uncertainty.

In Step 2, If the sampling of a coefficient value is uniform over the range:

$$c_i = c_{nominal} + (u_i - .5)c_{range}$$

If the sampling is Gaussian (normal), I recommend the Box-Muller method:

$$c_i = c_{nominal} + \sigma_c\sqrt{-2\ln(u_{1,i})}\sin(2\pi u_{2,i})$$

Where c_i is the i[th] realization for the coefficient value, and u_i is a uniformly distributed, independent, random number $0 < u_i \leq 1$, typically $u_i = 1 - r_i$, where r_i is the standard pseudo-random number.

Although this has a Monte Carlo approach, each realization of model coefficients provides one deterministic optimization exercise, not a stochastic OF.

Note: Contrasting the analytical approach, the numerical (Monte Carlo) method reveals when the uncertainty range is not symmetric about the nominal value, when there is interaction between the coefficients, and is valid even if the uncertainty range is large enough to violate the locally linear assumption in the analytical method.

Note: The numerical approach can include the impact of uncertainty in model coefficients in the uncertainty of x* and their joint impact on uncertainty in OF*, as well as the impact of convergence criterion.

Note: The numerical approach does not require the user to be able to take the derivatives and second derivatives of the function.

Accordingly, I believe that the numerical method is better. It is more general and gives more representative results. However, it does require computer programming.

DESIGN OF EXPERIMENTS

How much data?

The question is, "How many experiments to run?" If a data point has too much variability, then run several replicates and use the average. For example, in the bucket and stopwatch method to measure flow rate, when the flow rate is small, it may take a long time to fill the bucket, and as a result, the uncertainty on the start and stop time may be insignificant. One sample may have a small enough uncertainty to be fully adequate. However, at high flow rates, with short collection time to completely fill the bucket, the same uncertainty on the start and stop time may have a large impact on the calculated flow rate uncertainty. The central limit theorem can be derived by propagating variance on the average, and indicates that $\sigma_{average} = \sigma_{individual}/\sqrt{N}$. So, by averaging N replicates, the uncertainty on the average of N is reduced by the square root of N. By either propagating uncertainty to estimate $\sigma_{individual}$, or actually calculating $\sigma_{individual}$ from replicates, then specifying a desired $\sigma_{average}$, one can calculate the required number of replicates by $N = \left(\sigma_{individual}/\sigma_{average}\right)^2$.

Bootstrapping can provide the uncertainty on a model due to the variation in the data (if the model adequately fits the data functionality). If the model uncertainty is undesirably large, then adding

experimental data sets will decrease the model uncertainty. One can appropriate the central limit theorem to describe either the rms of the residuals, or the model uncertainty. If model uncertainty is defined by the 95% range, resulting from n experimental sets, then the number that should be run to reduce the uncertainty to a desired range is $= n(R_{.95\ from\ n}/R_{.95\ desired})^2$. Rather than replicating points, I prefer to add data at in-between locations to more completely cover the range, to eliminate data-sparse regions, to increase the likelihood that the experiments reveal all process attributes, and to provide a basis for model validation.

Determining Steady-State or Equilibrium

We often run experiments to equilibrium or steady state. Equilibrium is when a process is not being acted upon by external influence and settles to a steady value, the thermodynamic equilibrium. Similar, but in contrast, a steady state does not change in time, but it may not be in equilibrium. For example, the contents of a continuously fed reactor may be at steady state, may not be changing composition, but may not be at equilibrium, because new material is continuously being input. At steady state the value is unchanging in time. But, instead of measuring exactly the same value in a time series, measurement vagaries will add random perturbations to the data.

Often a procedure is to take samples periodically, and when three in a row are within expected uncertainty, then claim at SS. This is common in volatiles/moisture content testing, or determining when a process has come to equilibrium.

Use propagation of uncertainty in data models to determine the measurement range that would permit a sequence to be claimed at steady state, when data-to-data variation simply reflects normal variability, not real change in the variable value in time.

TROUBLE SHOOTING AND ANALYSIS OF OUTCOMES

1. Model-predicted variation is ≅ experimental (within 4:1 or 1:3 ratio).
   1.1. Probably confirms the analysis, and affirms understanding of the process.
   1.2. Can use the model to see key contributors and estimate the improvement if effort is invested to reduce the source of variation.
   1.3. Can use the model to see if alternate processing conditions would reduce variation (shorter samples for tensile testing (weakest link in a chain analysis), longer cure times to ensure approach to equilibrium or completeness, etc.
   1.4. The variation on an influence, a model input, an independent variable, may be due to several factors; if important, use propagation of uncertainty to see what aspect is most important. If no model is possible, run tests to seek to eliminate variation. Perhaps use a single technician to prep samples from common chemical batches on a single oven, etc. This may eliminate variability. If the resulting sigma is improved, then seek the course of variation and seek to eliminate that source.
2. Model-predicted variation is >> experimental (greater than 4 times experimental)
   2.1. Estimate of some source of variability is too large
   2.2. Experimental is standard error, $s/\sqrt{N}$, not standard deviation, s.
   2.3. Less than 10 samples in experimental s calculation.

2.4. Variation in influences may be correlated (if a tech rounds up then factors in the numerator and denominator will tend to cancel)

3. Model-predicted variation is << experimental (less than experimental/3)
   3.1. Model does not include an influence
      3.1.1. Such as T or %RH on Load – generate a model, perhaps by regression, then include the new variables in the analysis.
      3.1.2. Variation in experimental is due to multiple machines, multiple dual operators, etc. Isolate data from separate devices or procedures.
      3.1.3. Variation in experimental is due to sample preparation (composition, materials, technique). Run an experiment to have one person make up samples from one batch of chemicals, cooked in one oven, etc. Design a procedure to eliminate what might be sources of variation. See if this reduces variation.
      3.1.4. Variation is due to sloppy technique (sample handling, placement in machine, etc.).
   3.2. Estimate of some source of variability is too small.
   3.3. Experimental data is not limited to replicates, but combines disparate data and calculates sigma from things that should not be identical.
   3.4. Less than 10 samples in experimental σ calculation.
   3.5. Variation in influences may be correlated (if a tech rounds up then all dimensions will be high and area will be doubly high).