

**APPLICATIONS OF ERROR PROPAGATION ANALYSIS TO
THE UNCERTAINTIES OF REGRESSION MODELS
IN EXPERIMENTAL THERMAL
AND FLUIDS ENGINEERING**

Sheldon M. Jeter

Georgia Institute of Technology

Introduction

Regression models are common in experimental thermal and fluids engineering. Typical applications are calibration of instruments, correlation of thermodynamic properties, and development of transport models. For the models to be used confidently and competently, students and practitioners must understand both the development of the models and the evaluation of the uncertainty of the models. The latter understanding is apparently not adequately developed in typical undergraduate statistics courses. Engineering students usually have adequate familiarity with the development of simple regression models. An example of such a model is a linear or a polynomial calibration curve. A specific example is a fitting the data for the thermoelectric potential of a thermocouple to a polynomial of temperature over the range of calibration. Students are somewhat familiar with the uncertainty of the data with respect to such a model, but even in such a simple case, students are usually unfamiliar with evaluating the uncertainty of model itself.

A further interesting complication is introduced when the model is developed not in terms of the variables directly measured but instead in terms of transformed variables calculated from the measured variables. An example of such a model is the widely applied Clausius-Clapeyron expression for the logarithm of the vapor pressure as a polynomial function of the inverse absolute temperature. In the literature of experimental uncertainty, such as the report by Taylor and Mohr³, the result of reading a graduated instrument is called a direct measurement; and the result of a calculation using direct measurements is called an indirect measurement. For consistency in the absence of obvious distinguishing terms in the literature, this paper will refer to models based on direct measurements as “direct models”, and models involving indirect measurements will be called “indirect models”.

The well understood theory and practice of developing simple direct models are easily expanded to include the development of indirect models so long as they are still linear in their parameters. In the case of even a relatively simple indirect model, however, most students are unable to evaluate even the uncertainty of the data and are certainly unable to evaluate the uncertainty of the model. The students are, of course, even less familiar with the development

and the evaluation of more complex indirect models that involve linearization. An example is the commonplace power law model for the Nusselt number in terms of the Reynolds number for forced convection. The shortcoming can be easily addressed. The basic principles and techniques of error propagation analysis (EPA) can be readily and concisely explained to engineering undergraduates, and these tools can be used to develop the desired uncertainty limits.

This paper reviews the principles of linear regression analysis and EPA and demonstrates applications to developing uncertainty limits for data and for models. The paper also includes laboratory examples of evaluating the uncertainties of direct and indirect linear and polynomial models.

Synopsis of Linear Regression

Linear regression fits a model that is linear in its parameters to the supporting data such that the variation of the data with respect to the model is minimized. The detailed statistical theory of linear regression is available in numerous elementary and intermediate textbooks. The excellent and comprehensive text by Draper and Smith¹ was used as a source in developing this paper. Since excellent references on this well developed topic are available, a lengthy discussion here is not warranted; however, for completeness some overview is needed. In addition, even though numerous references on regression are available, none is in itself completely satisfactory for the experimentalist or student of experimental engineering for at least three reasons. One, the readily available references do not consider the error of the model, or they limit detailed consideration only to linear models. Two, the references are only concerned with direct models and not with transformed or indirect models. Three, there are no practical examples including examples that consider all aspects of uncertainty. This section will review linear regression. The other topics are addressed later.

For an overview of regression, specifically consider a linear model

$$y_{\text{est}} = c + b x \quad (1)$$

The variation of the data with respect to the model is called the residual variation, RSS, which is

$$\text{RSS} = \sum_{i=1}^n (y_i - y_{\text{est}})^2 = \sum_{i=1}^n (y_i - c - b x_i)^2 \quad (2)$$

The two so-called normal equations that express the conditions for a minimum of the RSS are

$$\frac{\partial(\text{RSS})}{\partial c} = \sum_{i=1}^n 2(y_i - c - b x_i)(-1) = 0 \quad (3)$$

$$\frac{\partial(\text{RSS})}{\partial b} = \sum_{i=1}^n 2(y_i - c - b x_i)(-x_i) = 0 \quad (4)$$

or in expanded form

$$n c + \left(\sum_{i=1}^n x_i \right) b = \left(\sum_{i=1}^n y_i \right) \quad (5)$$

$$\left(\sum_{i=1}^n x_i \right) c + \left(\sum_{i=1}^n x_i^2 \right) b = \left(\sum_{i=1}^n x_i y_i \right) \quad (6)$$

The two equations can be written in matrix form as

$$\begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} \begin{bmatrix} c \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix} \quad (7)$$

The explicit solution for the coefficient b can be found by using Cramer's rule to be

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i} = \frac{\sum_{i=1}^n (x_i - x_{\text{ave}}) y_i}{\sum_{i=1}^n x_i^2 - n x_{\text{ave}}^2} \quad (8)$$

Where

$$x_{\text{ave}} = \frac{\sum_{i=1}^n x_i}{n} \quad (9)$$

With the coefficient b known, it is easy enough to solve Equation 5 for the constant c as

$$c = y_{\text{ave}} - b x_{\text{ave}} \quad (10)$$

Where, of course

$$y_{\text{ave}} = \frac{\sum_{i=1}^n y_i}{n} \quad (11)$$

Equations 8 and 10 give the parameters for a least RSS linear model. These regression principles can be readily extended to more complex models. Next the uncertainty of the data with respect to the model should be considered.

Uncertainty of the Data with Respect to the Regression Model

Assuming that the model is a reasonable approximation of the systematic dependence of the dependent variable on the independent variable, then the random scatter of the data is the scatter of the data with respect to the model. This scatter is characterized by the Standard Error of Estimate (SEE) given by

$$SEE = \sqrt{\frac{\sum (y - y_{est})^2}{DF}} \quad (12)$$

Related experimental terminology will be reviewed later, but for now note that in experimental terms the SEE is the Standard Uncertainty A of the y data, $u_{A,y}$. Since the regression is based on a small sample, the statistics should be governed by the t-distribution. The t-distribution has an index of sample size called the statistical Degrees of Freedom (DF). When the DF is very large, the t-distribution is identical to the Normal distribution. For lower values of the DF , the t-distribution is slightly broader. In regression analysis the DF is the number of data less the number of parameters in the model. For a linear model the number of parameters, n_p , is 2, and for n data,

$$DF_{linear} = n - n_p = n - 2 \quad (13)$$

Experimentalists need hardly be concerned with the details of the formula for the SEE since statistics packages and spreadsheet programs calculate it as part of the regression analysis.

As noted, the SEE is in experimental terms the Standard Uncertainty A or $u_{A,y}$ for the data, meaning the uncertainty due to random variation that can be analyzed statistically. The Expanded Uncertainty of the Data or $U_{A,y}$ is calculated to be the 95 % confidence interval using the appropriate coverage factor, k_c , in the formula

$$U_{A,y} = k_c u_{A,y} = k_c SEE \quad (14)$$

This Expanded Uncertainty can be used to define the error limits of the data with respect to the model. Examples of such limits are the straight dotted lines labeled ELD (for Error Limit on Data) in Figure 1 below. Note that in standard regression analysis, the independent variables are assumed to be known exactly, and all of the variation is assigned to the dependent variable. The scatter in the dependent variable cannot be known exactly, but it is best approximated by the Standard Uncertainty in Equation 12 and the Expanded Uncertainty in Equation 14.

The uncertainty in the data with respect to the model is seen to be rather simple to evaluate. In contrast, evaluating the uncertainty of the model is relatively sophisticated. To explain how this uncertainty is developed, the principles of error propagation analysis will be reviewed in the next section. These principles will then be applied to show how the uncertainties in the parameters of a model and then the uncertainties in the model itself can be developed.

Recap of Error Propagation Analysis

The first conceptual step of error propagation analysis (EPA) is recognizing the idea of direct and indirect measurements. An indirect measurement is merely a calculation based on one or more direct measurements. Assume that m independent direct measurements, identified as a set of w_i s, contribute to an indirect measurement, z . The measurement formula is then merely the calculation formula,

$$z = z(w_1, w_2, \dots, w_m) \quad (15)$$

The operational concepts of EPA are essentially incorporated in two equations. The first of these two basic equations of EPA concerns how uncertainty in some dependent variable or indirect measurement z is caused by the uncertainty in some independent or directly measured variable, w . Call this uncertainty $u_{z,w}$. The formula relates the resulting uncertainty $u_{z,w}$ to the uncertainty in w as follows,

$$u_{z,w} = \frac{\partial z}{\partial w} u_w \quad (16)$$

The other basic operational formula shows how independent uncertainties are combined when several direct measurements contribute to an indirect measurement. Analysis shows that the squares of the contributing uncertainties sum to give the squared combined uncertainty, or

$$u_z^2 = u_{z,1}^2 + u_{z,2}^2 + \dots + u_{z,m}^2 = \left(\frac{\partial z}{\partial w_1} u_{w_1} \right)^2 + \left(\frac{\partial z}{\partial w_2} u_{w_2} \right)^2 + \dots + \left(\frac{\partial z}{\partial w_n} u_{w_m} \right)^2 \quad (17)$$

Each partial derivative is recognized to be the influence factor showing how each direct measurement, w_i , influences the indirect measurement, z . After reviewing the types of uncertainty, the two techniques presented above can be used to develop the formulas for the uncertainties of the parameters in a model and the uncertainty of the model itself as shown in the following sections.

Types of Uncertainty and Combining Uncertainties

Consensus standards representing the broad-based judgment of experienced professionals recognize two types of uncertainty. Uncertainty A is uncertainty that can be evaluated by

statistical analysis of the experimental data, such as the analysis of regression data in the preceding section. Uncertainty A is essentially the result of random variation. In contrast, Uncertainty B must be evaluated by analysis of the entire measurement system. Uncertainty B does not result in random variation but is rather the possible range of systematic error in the measurement system.

The Expanded Uncertainty is the half-width of the error band for the measurement. Typically this is the 95 % error band. The Expanded Uncertainty, U , is typically computed using the appropriate coverage factor, k_c , and the appropriate statistic called the Standard Uncertainty, u , as

$$U = k_c u \quad (18)$$

Typically the Standard Uncertainty is approximated using the appropriate standard error, and the coverage factor is computed using the t-distribution, which addresses small experimental samples.

To compute the overall Combined Uncertainty, first compute the Expanded Uncertainty A of the measurement by statistical analysis. Then compute Expanded Uncertainty B of the measurement by applying EPA to the overall measurement system. Since the random Uncertainty A and the Uncertainty B due to possible bias are obviously independent sources of error, compute the Combined Uncertainty of the model, as

$$U_C^2 = U_A^2 + U_B^2 \quad (19)$$

A spreadsheet block illustrated in Attachment A has been prepared to assist in calculating and plotting the uncertainties discussed in this paper. Careful inspection or use will reveal that the spreadsheet block includes a column for the Uncertainty B of the model. This uncertainty can be a constant; or, preferably, it can vary as a function of the independent variables. The spreadsheet block² is posted on line for the use of interested teachers and researchers.

Uncertainty of the Coefficient

The coefficient cannot be known exactly because it is influenced by the uncertainty in the dependent variable. For the coefficient b , the information needed to evaluate the influence factor for one uncertain value of the dependent variable on the coefficient is contained in Equation 8, specifically

$$\frac{\partial b}{\partial y_i} = \frac{(x_i - x_{ave})}{\sum_{i=1}^n x_i^2 - n x_{ave}^2} \quad (20)$$

The data set contains n pairs of data, x_i and y_i . By convention every x_i is assumed to be known without uncertainty, but every y_i is assumed to have some uncertainty. Since the variation in

every y_i value effects b , the uncertainty of the coefficient is written as follows by substituting the influence coefficients from Equation 20 into Equation 17, which is the rule for combining uncertainties. The pertinent result is

$$u_b^2 = \sum_{i=1}^n \left(\frac{\partial b}{\partial y_i} u_y \right)^2 = \sum_{i=1}^n \left(\frac{(x_i - x_{ave})}{\sum_{i=1}^n x_i^2 - n x_{ave}^2} u_y \right)^2 = \frac{u_y^2 \sum_{i=1}^n (x_i - x_{ave})^2}{\left(\sum_{i=1}^n x_i^2 - n x_{ave}^2 \right)^2} \quad (21)$$

Then expanding the summed term in the numerator and regrouping gives

$$u_b^2 = u_y^2 \frac{\sum_{i=1}^n (x_i^2 - 2x_i x_{ave} + x_{ave}^2)}{\left(\sum_{i=1}^n x_i^2 - n x_{ave}^2 \right)^2} = u_y^2 \frac{\sum_{i=1}^n x_i^2 - 2x_{ave} \sum_{i=1}^n x_i + \sum_{i=1}^n x_{ave}^2}{\left(\sum_{i=1}^n x_i^2 - n x_{ave}^2 \right)^2} \quad (22)$$

Then simplifying the numerator gives

$$u_b^2 = u_y^2 \frac{\sum_{i=1}^n x_i^2 - 2x_{ave}(n x_{ave}) + n x_{ave}^2}{\left(\sum_{i=1}^n x_i^2 - n x_{ave}^2 \right)^2} = u_y^2 \frac{\sum_{i=1}^n x_i^2 - n x_{ave}^2}{\left(\sum_{i=1}^n x_i^2 - n x_{ave}^2 \right)^2} \quad (23)$$

Simplifying the fraction gives the essential result

$$u_b^2 = \frac{u_y^2}{\sum_{i=1}^n x_i^2 - n x_{ave}^2} \quad (24)$$

In practice the uncertainty in y is not and cannot known *a priori*, so it is estimated by the Standard Error of y Estimate. The resulting computational formula is

$$u_b^2 = \frac{SEE^2}{\sum_{i=1}^n x_i^2 - n x_{ave}^2} \quad (25)$$

The uncertainty of the coefficient is an important statistic used routinely in significance testing.

Uncertainty of the Constant

While the uncertainty of the coefficient is used routinely in significance tests, the uncertainty of the constant is not quite so important. It can be useful in some experimental work, however, so it is included here for completeness. For the constant c , the information defining the influence factor is contained in Equation 5, so

$$\frac{\partial c}{\partial y_i} = \frac{\sum_{i=1}^n x_i^2 - x_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - n^2 x_{\text{ave}}^2} \quad (26)$$

Since the uncertainty in every y_i contributes to the uncertainty in c , the total uncertainty in c is

$$u_c^2 = \sum_{i=1}^n \left(\frac{\partial c}{\partial y_i} u_y \right)^2 = \sum_{i=1}^n \left(\frac{\left(\sum_{i=1}^n x_i^2 - x_i \sum_{i=1}^n x_i \right)}{n \sum_{i=1}^n x_i^2 - n^2 x_{\text{ave}}^2} u_y \right)^2 \quad (27)$$

Which after expansion followed by simplification gives

$$u_c^2 = \frac{\left(n \sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n x_i^2 - n x_{\text{ave}}^2 \right)}{n^2 \left(\sum_{i=1}^n x_i^2 - n x_{\text{ave}}^2 \right)^2} u_y^2 \quad (28)$$

After a final simplification and introducing the SEE for the uncertainty in y , this formula for the uncertainty in c results,

$$u_c^2 = \frac{\sum_{i=1}^n x_i^2}{n \left(\sum_{i=1}^n x_i^2 - n x_{\text{ave}}^2 \right)} \text{SEE}^2 \quad (29)$$

When needed, this uncertainty will typically be readily available, as it is computed by all of the common regression packages.

Uncertainty of a Simple Linear Model

EPA is also readily applied to the model itself. A quick look at Equation 1 might lead one to think that the uncertainty in a simple one variable linear model can be determined by direct application of the combining rule since the uncertainties in the constant and the coefficient are known from the analyses in the sections above. If this conjecture were true, the squared uncertainty in the model would be sum of the squared uncertainty in the constant and the squared uncertainty in the model due to the uncertainty in the coefficient, or

$$u_{\text{model}}^2 = u_c^2 + x^2 u_b^2 \quad \text{erroneous!} \quad (30)$$

This result is intuitively unsatisfactory because it increases monotonically with x while one would expect the uncertainty to increase toward the end points of the range of x . The conjecture in Equation 30 is wrong because the parameters c and b are not independent. The proper approach is to first eliminate c using Equation 10. After eliminating c , the model equation is

$$y_{\text{est}} = c + b x = (y_{\text{ave}} - b x_{\text{ave}}) + b x \quad (31)$$

so

$$y_{\text{est}} = y_{\text{ave}} + b(x - x_{\text{ave}}) \quad (32)$$

Now, the uncertainty in the model is easy to represent by applying the combining rule to the preceding relationship as

$$u_{\text{model}}^2 = u_{y-\text{ave}}^2 + (x - x_{\text{ave}})^2 u_b^2 \quad (33)$$

The SEE is used as the estimate of the uncertainty in the y data. Then the uncertainty in the average of y , which is the average of n individual y data, is

$$u_{y-\text{ave}} = \frac{\text{SEE}}{\sqrt{n}} \quad (34)$$

Since the uncertainty in the coefficient has been determined in Equation 25, then the uncertainty in the model is

$$u_{\text{model}}^2 = \left(\frac{\text{SEE}}{\sqrt{n}} \right)^2 + (x - x_{\text{ave}})^2 \frac{\text{SEE}^2}{\left(\sum_{i=1}^n x_i^2 - n x_{\text{ave}}^2 \right)^2} \quad (35)$$

or simplifying slightly

$$u_{\text{model}}^2 = \text{SEE}^2 \left(\frac{1}{n} + \frac{(x - x_{\text{ave}})^2}{\left(\sum_{i=1}^n x_i^2 - n x_{\text{ave}}^2 \right)^2} \right) \quad (36)$$

This result is intuitively entirely satisfactory for at least two reasons. First, because it is a minimum at the average value of the x data where the information about the actual trend in y is should be the best. Second, because it increases monotonically and approximately quadratically toward the either end of the range in x . These features are the expected and observed behavior in uncertainty of a linear model. As is usual, the uncertainty in the preceding equation is taken to be the Standard Uncertainty of the model. Note that this is the uncertainty due to random error or Uncertainty A. To plot the 95 % error band, the appropriate coverage factor, k_c , should be used in computing the Expanded Uncertainty A, or

$$U_{\text{model}} = k_c u_{\text{model}} \quad (37)$$

For a large number of data, the coverage factor will be 2, and many investigators use 2 even for small samples. Rigorously, the coverage factor should be computed from the t-distribution using as the Degrees of Freedom (DF) the number of data less the number of parameters in the model. The limits defined by this expanded uncertainty are a pair of curved somewhat hourglass shaped broken lines. A typical example is shown in Figure 1.

The graph shown in Figure 1 is the result of a practical application of linear regression including the regression model along with the corresponding error limits on the data and the, possibly more interesting, error limits on the model. The figure displays the calibration relationship between the temperature indicated by a digital thermometer and the calibration temperature measured by a field standard RTD. The standard RTD is taken to have an independently determined Uncertainty B due to possible bias and residual random variation of 0.01 Celsius degree. This Uncertainty B can be combined with the Uncertainty A for the data or the model by the usual rule for combining uncertainties from independent sources of error,

$$U_{C,\text{model}}^2 = U_{A,\text{model}}^2 + U_B^2 \quad (38)$$

Error limits based on this combined uncertainty, U_C , are the hour glass shaped pair of curved lines drawn on the figure.

The regression parameters and the SEE for the data in Figure 1 were computed by the Excel regression package. The complete spreadsheet⁴, including actual data, for this calibration exercise is posted on the author's web page for use by interested students or instructors. The curves showing the error bounds were computed in a compact block of the Excel workbook illustrated below as Attachment 1. This block is available in a more convenient format as a separate spreadsheet² for calculating the error limits of a model (ELM) on the author's web page. This example has 12 data points resulting in a DF of 10 for a linear model that has 2 parameters.

For the resulting 10 degrees of freedom, the t-distribution rigorously requires a coverage factor of 2.23 as is calculated by the spreadsheet block. If this value were used in drawing the graph, the error band would be invisibly small; therefore an exaggerated coverage factor of 40 was used for legibility. Note that the error band for the data is significantly wider than the error bound on the model, reflecting the averaging effect of the regression modeling.

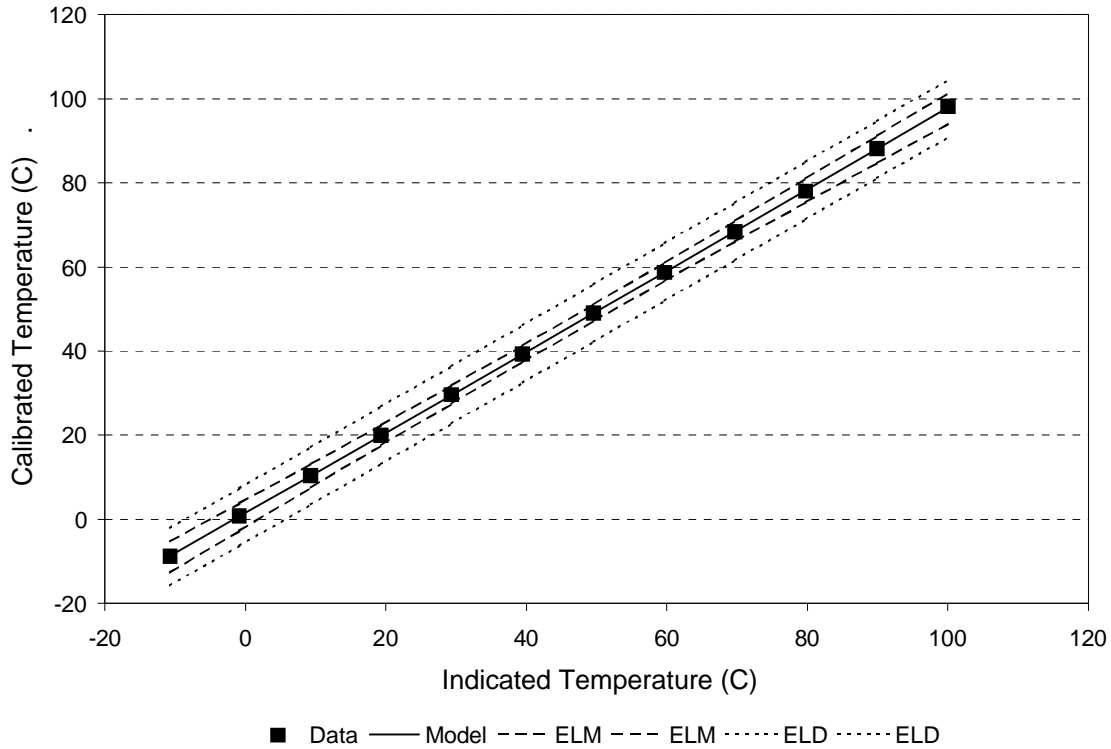


Figure 1. Temperature Calibration Data with Exaggerated Error Bands
 ELM is the Error Limit on the Model calculated using Equation 38.
 ELD is Error Limit on the Data calculated using Equation 14.

The practical benefit of knowing the uncertainty of the model is particularly important in a calibration application such as this example. The uncertainty of the calibration is fixed once the calibration is completed, and any part of the uncertainty caused by mere random variation during the calibration is from there on included in the now fixed combined uncertainty of the calibration. The combined uncertainty of the calibration must be considered as Uncertainty B when the calibrated instrument is used in an experiment since this uncertainty cannot be further evaluated by repeated experimental measurements. To minimize the uncertainty in the experimental application, the uncertainty of the calibration should be minimized. Since the calibration is intended to represent the systematic trend between the indicated temperature and the calibration temperature, the uncertainty of the model not the uncertainty of the data should be used. Increasing the number of well distributed calibration points will significantly reduce the uncertainty of the model since the averaging effect tends to balance out the fluctuations caused by mere random variation. In contrast, the uncertainty of the data is hardly changed by repeated

measurements since it represents the inherent random variation in the data. In this case the Expanded Uncertainty of the data is .38 Celsius degree, while the average Expanded Uncertainty of the model is a considerably smaller value of .15 Celsius degree.

Uncertainty of a More Complex Model

EPA is also readily applied to the uncertainty of a more complex model with multiple independent variables so long as it is linear in its parameters. It can be shown that centering the data by subtracting their averages from the dependent variable and the independent variables always eliminates the constant, so the model can always be written as

$$y_{\text{est}} = y_{\text{ave}} + b_1(x_1 - x_{1,\text{ave}}) + b_2(x_2 - x_{2,\text{ave}}) + \cdots + b_m(x_m - x_{m,\text{ave}}) \quad (39)$$

As before, the uncertainty in this centered model is easy to formulate by applying the combining rule to the preceding relationship, so

$$u_{\text{model}}^2 = u_{y-\text{ave}}^2 + (x_1 - x_{1,\text{ave}})^2 u_{b1}^2 + \cdots + (x_m - x_{m,\text{ave}})^2 u_{bm}^2 \quad (40)$$

The uncertainty in the average of y , is again computed using the SEE as the standard deviation in the formula for an average of n data, so

$$u_{y-\text{ave}} = \frac{\text{SEE}}{\sqrt{n}} \quad (41)$$

The uncertainty in this more complicated model, such as a polynomial model, is then

$$u_{\text{model}}^2 = \left(\frac{\text{SEE}}{\sqrt{n}} \right)^2 + (x_1 - x_{1,\text{ave}})^2 u_{b1}^2 + \cdots + (x_m - x_{m,\text{ave}})^2 u_{bm}^2 \quad (42)$$

An example of application of the previous equation is a polynomial regression model such as typically used for a thermal anemometer calibration. The general form of the uncertainty for a polynomial model is

$$u_{\text{model}}^2 = \left(\frac{\text{SEE}}{\sqrt{n}} \right)^2 + (x - x_{\text{ave}})^2 u_{b1}^2 + \cdots + (x^m - x_{\text{ave}}^m)^2 u_{bm}^2 \quad (43)$$

Since the uncertainties of the coefficients are always computed by standard commercial regression packages, it is straightforward to calculate and plot the Expanded Uncertainty A according to the usual formula,

$$U_{A,\text{model}} = k_c u_{\text{model}} \quad (44)$$

The spreadsheet block called ELM illustrated in Attachment 1 is designed for applications with as many as four independent variables. The results for an example quadratic calibration model for a thermal anemometer are shown in the Figure 2 below. The complete spreadsheet³, including actual data, for this thermal anemometer calibration is also posted on the author's web page for review or use by interested students or instructors. This complete spreadsheet includes the block called ELM, which made it easy to compute and plot the Uncertainty A for this and indeed any regression model.

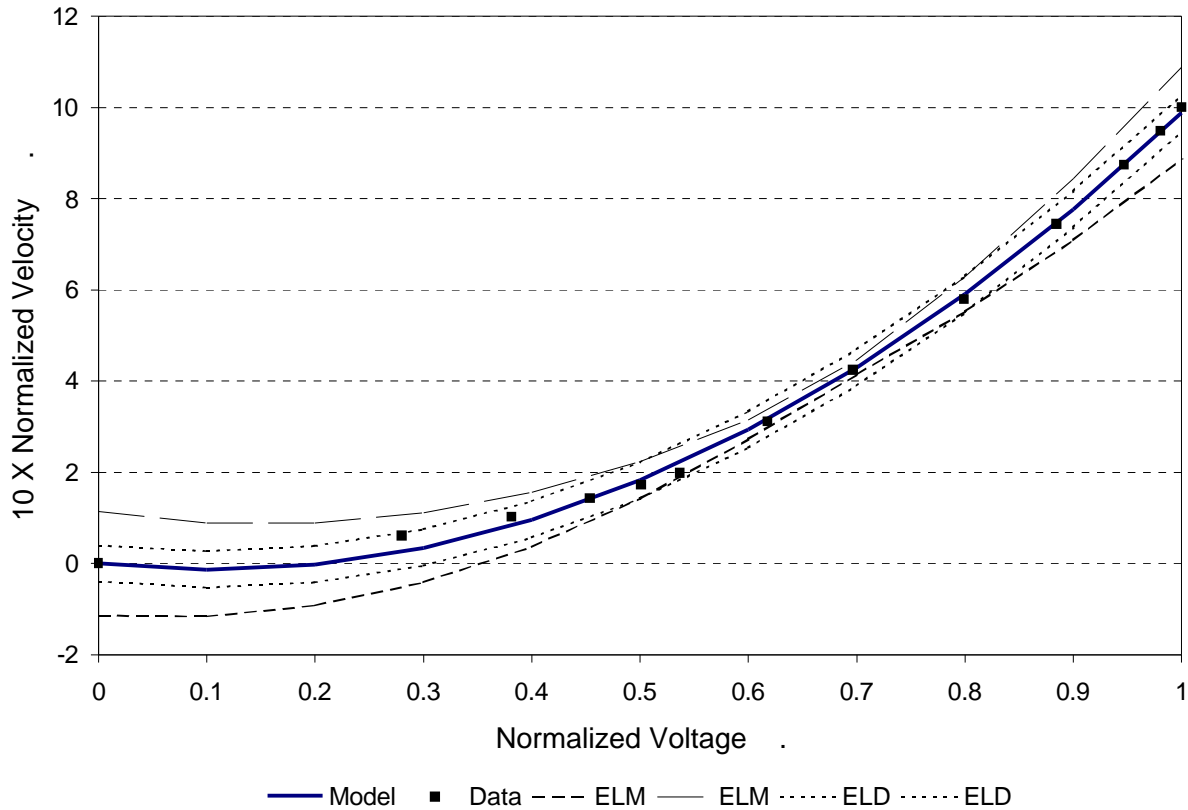


Figure 2 Example Plot of a Quadratic Model with the Uncertainty Limits for the Data (ELD) and the Uncertainty Limits for the Model (ELM) Included.

As noted above, to compute the Combined Uncertainty first compute the Expanded Uncertainty A of the model as detailed above. Then compute Expanded Uncertainty B of the indirect measurement by applying EPA to the overall measurement system. Then compute the Combined Uncertainty of the model, as

$$U_{C,model}^2 = U_{A,model}^2 + U_B^2 \quad (45)$$

For this calculation the spreadsheet block called ELM, illustrated in Attachment A, has a column for the Uncertainty B of the model. This uncertainty can be a constant; or, preferably, it can vary as a function of the independent variables.

In Figure 2, the uncertainty band for the data is delimited by dotted lines plotted the Expanded Uncertainty A of the data above and below the model. For the data the Uncertainty A is a constant. Indeed, the Expanded Uncertainty A of the data is the constant computed by the usual formula

$$U_{A,data} = k_c \text{ SEE} \quad (46)$$

The uncertainty band for the model is delimited by broken lines plotted the Expanded Uncertainty of the Model above and below the regression model. The Uncertainty A of the model varies with x in a roughly quadratic fashion. The Uncertainty B of the model could be estimated by error propagation analysis and included, but in this case it has been ignored for simplicity.

Uncertainties of an Indirect Model

The model in Figure 2 is actually a simplistic case of an indirect model since the regression variables are normalized. A more interesting and realistic example of an indirect model is a model with transformed variables such as the familiar Clausius-Clapeyron model for the vapor pressure. The simplest case is a model linear in the inverse temperature,

$$\ln\left(\frac{P_v}{P_0}\right) = c + b \frac{1}{T} \quad (47)$$

Here P_0 is the unit pressure, which was taken to be 1 kPa in this particular example. Normalizing the vapor pressure by dividing by unit pressure ensures that the argument of the logarithm is dimensionless. Note that the transformed dependent variable is the logarithm of the normalized transformed model, and note that the transformed independent variable is the inverse absolute temperature.

The transformed logarithmic data and logarithmic model can be plotted with error limits for the data and model on a log-linear plot. The linear scale would be used for the inverse temperature. The result is hardly different in concept from Figure 1, so it is not shown.

Developing the uncertainties for the dimensional data and the dimensional model, *i. e.* the uncertainties in terms of the vapor pressure itself, is more interesting. The transformation of the logarithmic results to dimensional results is straightforward. First recall Equation 16, which is the relevant basic relationship from EPA. This equation is used to estimate the uncertainty in a calculated variable, otherwise called an indirect measurement z , when the uncertainty in the direct measurement, w , is known. The partial derivative in this equation is calculated for the function relationship between the two variables, otherwise called the measurement formula. In this case the relationship is

$$P_v = P_0 \exp(\ln(P_v/P_0)) \quad (48)$$

Here the indirect measurement, z , is

$$z = P_v \quad (49)$$

Note carefully in this case that the direct measurement is the normalized pressure, or

$$w = \ln(P_v/P_0) \quad (50)$$

So the partial derivative needed in the basic EPA relationship, Equation 16, is

$$\frac{\partial z}{\partial w} = P_0 \exp(\ln(P_v/P_0)) = P_v \quad (51)$$

Therefore an uncertainty for the vapor pressure can be inferred from an uncertainty for the logarithm of the vapor pressure as follows.

$$U_{PV} = P_v U_{LPV} \quad (52)$$

Where the symbol identified as U_{LPV} can stand for either the Uncertainty A for the logarithmic data, U_{ALD} , or the different uncertainty for the logarithmic model, U_{ALM} , as appropriate. Obviously the uncertainty determined for a logarithmic representation is a fractional uncertainty for the underlying data.

To illustrate the uncertainty in the vapor pressure data and model, consider Figure 3. This figure shows the dimensional vapor pressure data and the vapor pressure regression model in the form of P_v versus T . Error limits for the data and the model are also included. The error limits for the data are computed from

$$U_{PVD} = P_v U_{ALD} \quad (53)$$

Here the uncertainty of the logarithmic data is computed with the SEE of the logarithmic model. Note that even though the uncertainty of the logarithmic data is constant the uncertainty of the vapor pressure data varies with the vapor pressure. The error limits for the vapor pressure model are computed from

$$U_{PVM} = P_v U_{ALM} \quad (54)$$

The results are shown in Figure 3 below. Note how all error limits are now curved since the uncertainties increase with the vapor pressure. The complete spreadsheet⁵, including actual data, for this vapor pressure correlation is also posted on the author's web page for review or use by interested students or instructors.

Conclusion

This paper has reviewed the principles of linear regression analysis and error propagation analysis as related to the uncertainty analysis of regression models. Various applications of

interest in experimental thermal and fluids engineering have been demonstrated. Methods for evaluating uncertainty limits for data and for models are presented. The paper also includes laboratory examples for evaluating the uncertainties of direct and indirect linear and polynomial models. It is seen that the uncertainty limits for data and models can be easily generated with data that are readily available from standard regression packages.

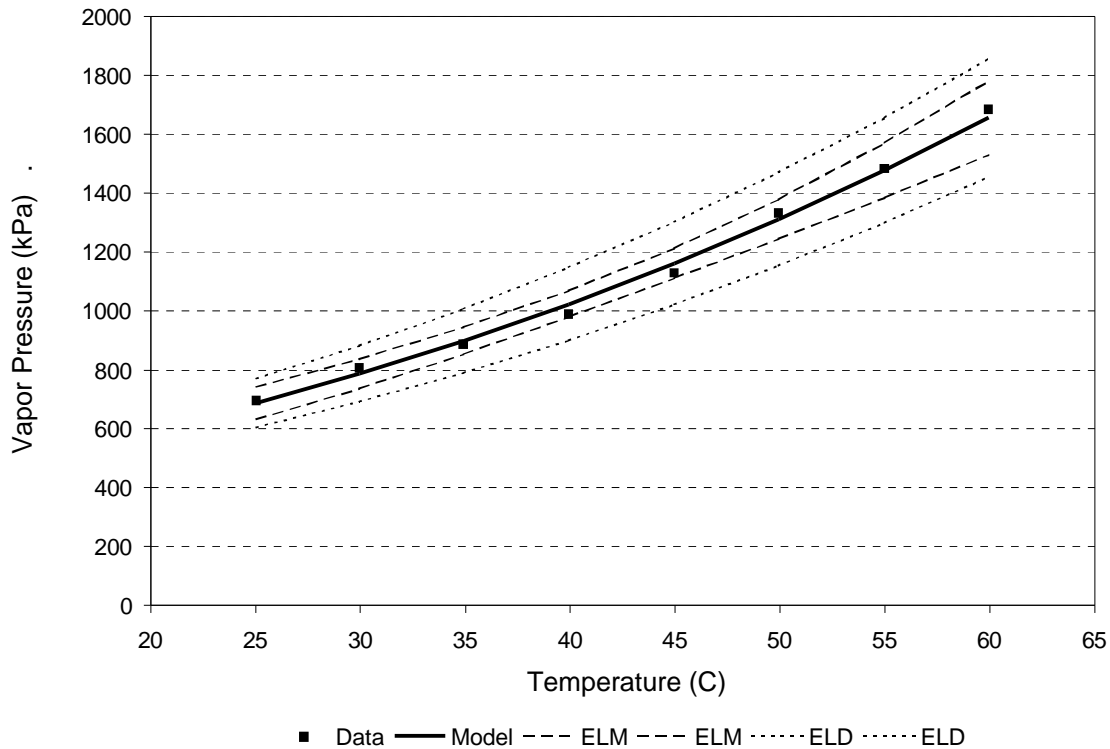


Figure 3. Plots of Vapor Pressure Data, Model, and Uncertainty Limits. The model and the uncertainty limits were inferred from a logarithmic model.

References

1. Draper, N. R. and H. Smith, *Applied Regression Analysis*, third edition, John Wiley and Sons, New York, NY.
2. Jeter, S. M., "Spreadsheet ELM for Computing the Error Limits of a Model", ME 4053 Engineering Systems Laboratory, the George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, 4 January 2002, available on line at me.gatech.edu/~sjeter.
3. Jeter, S. M., "Spreadsheet TACAL_2002: New Calibration Spreadsheet for IFA Thermal Anemometer", ME 4053 Engineering Systems Laboratory, the George W. Woodruff School of

Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, 14 January 2002, available on line at <me.gatech.edu/~sjeter>.

4. Jeter, S. M., “Spreadsheet TCCAL_2002: Calibration Spreadsheet for Thermocouples”, Thermal Hydraulics Laboratory, the George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, 14 January 2002, available on line at <me.gatech.edu/~sjeter>.

5. Jeter, S. M., “Spreadsheet VPRESS-Sp-2002_Example: Vapor Pressure of HFC-134a”, ME 4053 Engineering Systems Laboratory, the George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, 12 February 2002, available on line at <me.gatech.edu/~sjeter>.

6. Taylor, B. N. and P. J. Mohr, 1999, “The NIST Reference on Constants, Units, and Uncertainty”, NIST Physics Laboratory, NIST, Gaithersbery, MD, 23 July 1999, available online at <<http://physics.nist.gov/cuu/Uncertainty/index.html>>.

Biography

SHELDON M. JETER is Associate Professor of Mechanical Engineering at the George W. Woodruff School of Mechanical Engineering at Georgia Tech. He has degrees from Clemson University, the University of Florida, and Georgia Tech. He has been on the academic faculty at Georgia Tech since 1979. His research interests are thermodynamics His research interests are thermodynamics, heat and mass transfer, and energy systems.

Appendix: Spreadsheet Block ELM for Computing Uncertainty of a Model

Sheet to Compute and Plot the Uncertainty of a Model, SMJ Nov 2001, updated 11 Jan 2002

User must insert and/or update data coded with yellow. User should format these and other cells for neatness or legibility as

- Instructions:**
- (1) Copy this sheet to your experimental *.xls workbook.
 - (2) Insert the experimetal data into block 2.
 - (3) Insert the regression results into block 2.
 - (4) Select the desired coverage factor.
 - (5) Update the cell ranges to compute the averages in Block 3.
 - (6) Identify by cell formula the maximum and minimum X1 values in the cells on the next row.
 - (7) Insert data or formulas for other XN data in Block 5.
 - (8) Insert optional data for U_b in block 5.
 - (9) Plot experimental data points with green block in Block 4.
 - (10) Plot model and limits with green block in Block 5.

1. Summary Data:

The averaged U_a of model = 0.152 The averaged U_c of model = 0.152
 The constant U_a of the data = 0.380

2. Block of Data from Regression

User must insert the following data from the regression block.

Constant = 1.43743
 Coefficient B1 = 0.96190 0.00142 = Std error of B1
 Coefficient B2 = 0.00000 0.00000 = Std error of B2
 Coefficient B3 = 0.00000 0.00000 = Std error of B3
 Coefficient B4 = 0.00000 0.00000 = Std error of B4
 Std Error of y Est = 0.17056
 n, number of data = 12
 p, number of parameters = 2
 coverage factor, kc, by t-dist = 2.23

3. Block of Calculations

User must specify the desired coverage factor to be used. Rigorous value is in cell D16.

coverage factor, kc, used = 2.23 <----User must select.

User must update the cell ranges in the following four formulas to calculate the correct averages.

Average value of X1 = 44.54383
 Average value of X2 = 3194.737
 Average value of X3 = 0.00000
 Average value of X4 = 0.00000

4. Block of Experimental Data and Results, plot the data with the block in green

User must insert the complete set of y and x data from the experimental data set into following block. Add rows as required.

User may insert column of data for Expanded Uncertainty B, uncertainty due to possible bias, if desired.

Experimental Data, insert at least one zero in every otherwise unused XN column

Y data	X1 data	X2 data	X3 data	X4 data	U_b of model	U_a of model	U_c of model	X1 data	Y data	regress model
	X1	X2	X3	X4						
-8.737	-10.698	114.45	0.000	0.000	0.010	0.206	0.206	-10.698	-8.737	-8.853
0.781	-0.880	0.77			0.010	0.180	0.181	-0.880	0.781	0.591
10.382	9.302	86.53			0.010	0.156	0.156	9.302	10.382	10.385
19.988	19.309	372.84			0.010	0.136	0.136	19.309	19.988	20.011
29.583	29.308	858.96			0.010	0.120	0.120	29.308	29.583	29.629
39.266	39.434	1555.04			0.010	0.111	0.111	39.434	39.266	39.369
48.948	49.531	2453.32			0.010	0.111	0.111	49.531	48.948	49.081
58.663	59.688	3562.66			0.010	0.120	0.120	59.688	58.663	58.852
68.340	69.723	4861.30			0.010	0.135	0.136	69.723	68.340	68.504
78.093	79.810	6369.64			0.010	0.156	0.157	79.810	78.093	78.207
88.056	89.923	8086.15			0.010	0.180	0.181	89.923	88.056	87.935
98.048	100.076	10015.21			0.010	0.207	0.207	100.076	98.048	97.701

5. Block Uniformly Spaced Data for Plotting wrt X1, plot model and limits with the block in green

User must insert cell references to identify the maximum and minimum X1 values in the cells on the next row.

max X1 = 100.076 min X1 = -10.698 1.11E+01 = computed delta X1

Spreadsheet will compute uniformly spaced X1 values. User must code columns for corresponding values of X2, X3, and X4.

User may insert column of data for Expanded Uncertainty B, uncertainty due to possible bias, if desired.

X1	X2	X3	X4	U_b of model	U_a of model	U_c of model	X1 data	regress model
-1.1E+01	114.45			0.010	0.206	0.206	-1.1E+01	-8.853
3.79E-01	0.14			0.010	0.177	0.178	3.8E-01	1.802
1.15E+01	131.26			0.010	0.151	0.152	1.1E+01	12.458
2.25E+01	507.79			0.010	0.130	0.130	2.3E+01	23.113
3.36E+01	1129.74			0.010	0.115	0.115	3.4E+01	33.769
4.47E+01	1997.11			0.010	0.110	0.110	4.5E+01	44.424
5.58E+01	3109.89			0.010	0.115	0.116	5.6E+01	55.079
6.68E+01	4468.09			0.010	0.130	0.131	6.7E+01	65.735
7.79E+01	6071.71			0.010	0.152	0.152	7.8E+01	76.390