Quantifying Model Error

Kevin McGrattan[1]
Fire Research Division

Blaza Toman
Statistical Engineering Division

National Institute of Standards and Technology
Gaithersburg, Maryland, 20899

ABSTRACT: A methodology is proposed for reporting the results of fire model validation studies that consist mainly of comparisons of model predictions and experimental measurements. The difference between the two is a combination of model error and experimental uncertainty. Assuming that one can quantify the experimental uncertainty, the model error can be estimated using a relatively simple form of Bayesian analysis.

KEY WORDS: fire modeling, validation.

1   Introduction

There are various definitions of model validation. For example, ASTM E 1355, "Standard Guide for Evaluating the Predictive Capability of Deterministic Fire Models," [1] defines it as "the process of determining the degree to which a calculation method is an accurate representation of the real world from the perspective of the intended uses of the calculation method." The accuracy of a model is a complicated function of its inputs and its equations. For the sake of clarity, we will use the word *uncertainty* in regard to the inputs, and the word *error* in regards to the equations. The reason for this is two-fold. For one, it helps to differentiate various terms that are used to quantify model accuracy; but, more importantly, it highlights the subtle distinction between the roles of the model and of the modeler. For example, if the modeler chooses to input 10 MW as the heat release rate of a hypothetical fire, we do not say it is wrong, but rather uncertain because there is no easy way to know for sure just how big the fire is going to be. The mathematical model, on the other hand, includes assumptions and approximations that make its predictions less accurate for the sake of practical implementation.

Our focus at present is not the input uncertainty, but rather the quantification of the model error. Our strategy is *not* to decompose the model into its many parts and systematically assess the error of each. Such an exercise would be extremely difficult, especially for a computational fluid dynamics (CFD) model, and the results would probably be subject to the same degree of error as the model itself. Instead, we take the traditional approach in which model predictions are compared with experimental measurements. However, unlike many model validation studies performed to date, we suggest a method by which the model *error* can be distinguished from the model input *uncertainty*. The difference between model prediction and experimental measurement

---

[1]Corresponding author, kevin.mcgrattan@nist.gov

is a combination of the two, and because the input uncertainty is easier to quantify than the model error, we can estimate the error by "subtracting off," in the broad sense, the input uncertainty from the reported difference between prediction and measurement. The quotation marks in the previous sentence indicate that model error and input uncertainty do not simply add together, and to decouple them requires some statistical analysis that will be described below.

2 Model Error

A deterministic fire model is based on fundamental conservation laws of mass, momentum and energy, applied either to entire compartments or smaller control volumes that make up the compartments. A CFD model may use millions of control volumes to compute the solution of the Navier-Stokes equations. However, it does not actually solve the Navier-Stokes equations, but rather an approximate form of these equations. The approximation involves simplifying physical assumptions, like the various techniques for treating subgrid-scale turbulence. The most critical approximation is the discretization of the governing equations. For example, the partial derivative of the density, ρ, with respect to the spatial coordinate, x, can be written in approximate form as:

$$\frac{\partial \rho}{\partial x} = \frac{\rho_{i+1} - \rho_{i-1}}{2\,\delta x} + O(\delta x^2) \tag{1}$$

where δx is the grid spacing chosen by the model user. The second term on the right represents all the terms of order δx^2 and higher in the Taylor series expansion and are known collectively as the *discretization error*. These extra terms are simply dropped from the equation set, the argument being that they become smaller and smaller with decreasing grid cell size, δx. The effect of these neglected terms is captured, to some extent, by the subgrid-scale turbulence model, but that is yet another approximation of the true physics. What effect do these approximations have on the predicted results? It is very difficult to determine based on an analysis of the discretized equations. At best, one can demonstrate convergence towards some analytical solution of the non-turbulent Navier-Stokes equations at a rate determined by the spatial and temporal differencing schemes, but this still does not answer a question like, "What is the error of the model prediction of the gas temperature at a particular location in the room at a particular point in time?"

To make matters worse, there are literally dozens of subroutines that make up a CFD fire model, from its radiation solver, solid phase heat transfer routines, pyrolysis model, empirical mass, momentum and energy transfer routines at the wall, and so on. A careful, systematic error analysis of the combination of all these routines is nearly impossible. Indeed the complexity of the analysis would approach that of the model itself. For this reason, we rely on comparisons of model predictions to as many experiments as possible as a means of quantifying model error. However, before we can definitively quantify the error, we first must quantify the uncertainty of the measurements against which the model predictions are to be compared. This is discussed below.

3 The Validation Process

Because it is impractical to dissect a complicated fire model into its many components and assess the accuracy of each individually, a fire model validation study typically consists of comparing
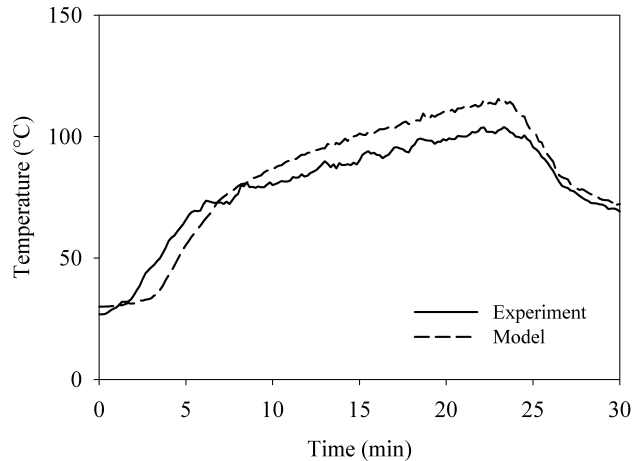
Figure 1: Example of a typical time history comparison of model prediction and experimental measurement.

point measurements from a wide variety of fire experiments with corresponding model predictions. Figure 1 is a typical result for a single point measurement, and given that usually dozens of such measurements are made during each experiment, and potentially dozens of experiments are conducted as part of a test series, we can expect hundreds of such plots for any given quantity of interest. In many cases, these plots in themselves provide sufficient information for a potential model user or authority having jurisdiction (AHJ) to decide if the model is sufficiently accurate for the given application. Because there is no consensus "acceptance criteria" within the fire protection community to indicate what constitutes sufficient accuracy, the judgment is left to end user or AHJ. It can be argued that this is as it should be because each situation is different. For example, some regulatory authorities prefer models that consistently over-predict the severity of the fire, in which case the model serves as a "screening tool." Design engineers, however, usually prefer that the model predict the outcome of the fire as accurately as possible, even though they may apply some form of safety factor to the results. In any case, there is certainly no reason why we cannot develop a process to quantify the accuracy of the model that goes beyond simply publishing hundreds of plots like the one shown in Fig. 1.

To this end, we first have to decide how to condense all the data represented by these plots into a more tractable form. Specifically, we have to decide on a metric with which to compare two curves like the ones shown in Fig. 1. Peacock *et al.* [2] discuss various possible metrics. A commonly used metric is simply the difference between the peak measured and predicted values. If the data is spiky, some form of time-averaging can be used. Regardless of the exact form of the metric, what results from this exercise is a pair of numbers for each plot, (E_i, M_i), that can be depicted graphically as shown in Fig. 2. This plot condenses hundreds of plots into just one, but we still have the problem of quantifying the degree of scatter in the results. At this stage in the analysis, the diagonal line on the plot shown in Fig. 2 only indicates where a prediction and measurement agree. But because the model has error, and the measurement has uncertainty, it cannot be said that the model is perfect if its predictions agree exactly with measurements. There needs to be a way of quantifying both the error and uncertainty before any conclusions can be drawn. Such an exercise
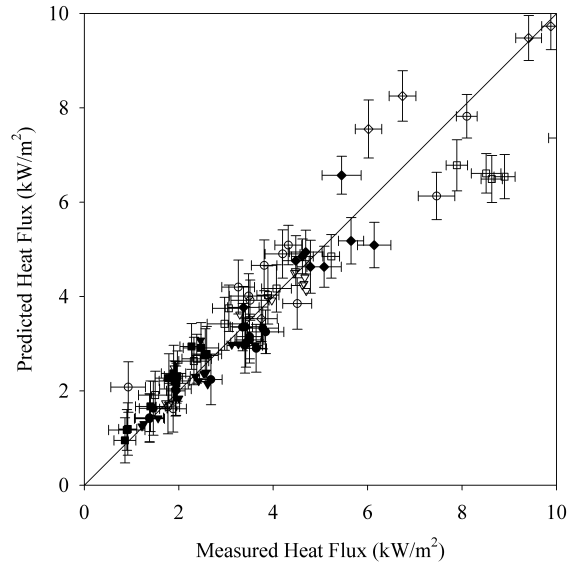
3

Figure 2: Example of a typical scatter plot of model predictions and experimental measurements.

would result in the error/uncertainty bars shown in the figure. At first glance, it would appear that the horizontal bar associated with each point represents the uncertainty in the measurement, and the vertical bar represents the error in the model. Unfortunately, it is not a simple task to quantify these values, nor can it be said that the vertical bar is the result solely of model error. As will be discussed below, any experiment involves measurements of elevated temperatures, heat flux, *etc.*, resulting from the fire, but also measurements of parameters, like the heat release rate and material properties, that are then fed into the model as input. Uncertainty in these input parameters combines with the error in the model to form the vertical uncertainty/error bar associated with each point in Fig. 2. Decoupling the error from the uncertainty is the subject of analysis below. Before presenting this analysis, however, we must first discuss in more detail the subject of experimental uncertainty.

4   Experimental Uncertainty

For various reasons, documented fire experiments, especially full-scale experiments, contain fairly limited discussion of uncertainty. Some contain none at all. However, in order to assess the accuracy of the model whose predictions have been compared to these measurements, there must be some estimate of the combined effect of the uncertainty in the reported test parameters, like the heat release rate of the fire, and the reported measurements of the quantities of interest, like the hot gas layer (HGL) temperature, heat flux, and so forth. Otherwise, the difference between model and measurement would simply have to be attributed as model error and there is no point in continuing this exercise.

Rather than throwing out several decades worth of fire test data, there are ways that we can estimate the uncertainty of the measurements based on work that has been done to quantify the accuracy of such devices as heat flux gauges, oxygen-depletion calorimeters, thermocouples and

gas sensors. In a recent fire model validation study conducted by the U.S. Nuclear Regulatory Commission [3], Hamins estimated the combined uncertainty of the quantities of interest for the large scale fire experiments involved in the study. Referring to Fig. 2, there were two uncertainty estimates needed for each quantity. The first was an estimate, expressed in the form of a 95 % confidence interval, of the uncertainty in the measured quantity itself. For example, reported gas and surface temperatures were made with thermocouples of various designs (bare-bead, shielded, aspirated) with different size beads, metals, and so on. For each, one can estimate the uncertainty in the reported measurement. Next, the uncertainty of the measurements of the reported conditions was estimated, including the heat release rate, leakage area, ventilation rate, material properties, and so on. The effect of these uncertainties on the reported measured quantity had to be assessed as well. For example, most of the experiments used oxygen consumption calorimetry as the means of measuring the heat release rate. Estimates of the uncertainty in large scale calorimeters is on the order of 15 %. It has been shown [4] that the hot gas layer temperature rise due to a compartment fire is proportional to the HRR raised to the two-thirds power, thus a 15 % uncertainty in the HRR would lead to a $2/3 \times 15$ %=10 % uncertainty in the model prediction. This uncertainty now needs to be combined with the uncertainty in the thermocouple measurement itself. Let us say that this is also 10 %. Because the two forms of uncertainty are uncorrelated, they are combined by quadrature (summing of squares) to yield a combined uncertainty of 14 %. An other way to look at this is to recognize that the combined uncertainty is represented as the diagonal of the rectangle formed from the horizontal and vertical uncertainty/error bars in Fig. 2.

Hamins performed this exercise for ten quantities of interest in the U.S. NRC validation study. The results are summarized in Table 1, with each combined uncertainty reported in the form of a 95 % confidence interval (*i.e.* $2\widetilde{\sigma}_E$). The tilde above the σ denotes a *relative uncertainty*, which is a convenient way to report it because we are assuming that the uncertainty in the reported value is proportional to its magnitude. This assumption is made throughout the analysis for both measurement uncertainty and model error. The assumption is based on a qualitative assessment of dozens of scatter plots similar to that shown in Fig. 2 that show the scattered points to form an expanding "wedge" about the diagonal line, or some other off-diagonal line due to an assumed bias in the model predictions. This assessment is a critical component of the analysis described in the next section.

5   Calculating Model Error

We are now ready to describe a method for calculating the model error using the estimated experimental uncertainties listed in Table 1. We make two important assumptions:

1. All uncertainties and errors can be expressed in relative terms, and

2. All uncertainties and errors are assumed to be normally distributed.

Given the complexity of the models and the experiments, it would be difficult to justify any particular distribution. In fact, the very reason why we have developed this method of quantifying model error based on validation experiments is because the model algorithm itself is too complicated to work with directly as a means of estimating the error.

With these assumptions in mind, we proceed with the statistical analysis of the data. Assume that the set of model predictions and the corresponding set of experimental measurements are

Table 1: Summary of Hamins' uncertainty estimates [3].

Measured Quantity	Combined Relative Uncertainty, $2\widetilde{\sigma}_E$
HGL Temperature	0.14
HGL Depth	0.13
Ceiling Jet Temperature	0.16
Plume Temperature	0.14
Gas Concentrations	0.09
Smoke Concentration	0.33
Pressure with Ventilation	0.80
Pressure without Ventilation	0.40
Heat Flux	0.20
Surface Temperature	0.14

denoted M_i and E_i, respectively, where i ranges from 1 to n and both M_i and E_i are positive numbers expressing the increase in the value of a quantity above its ambient. We assume that a given experimental measurement, E_i, is normally distributed[2] about the "true" value, θ_i, and there is no systematic bias:

$$E|\theta \sim N(\theta, \sigma_E^2) \tag{2}$$

The notation $E|\theta$ means that E is conditional on a particular value of θ. This is the usual way of defining a likelihood function. It is convenient to use the so-called delta method[3] to obtain the approximate distribution

$$\ln(E)|\theta \sim N\left(\ln(\theta), \widetilde{\sigma}_E^2\right) \tag{3}$$

The purpose of applying the natural log to the random variable is so that we can express its variance in terms of the relative uncertainty, $\widetilde{\sigma}_E = \sigma_E/\theta$. This is convenient because we have assumed that the relative uncertainty is constant for each quantity of interest. The quantities and uncertainty values are listed in Table 1. Using a Bayesian argument[4] with a non-informative prior on $\ln(\theta)$, we obtain the posterior distribution as

$$\ln(\theta)|E \sim N\left(\ln(E), \widetilde{\sigma}_E^2\right) \tag{4}$$

The assumption of a non-informative prior implies that we do not have sufficient information about the prior distribution (*i.e.* the true value) of θ to assume anything other than unity.

We cannot assume, as in the case of the experimental measurements, that the model predictions have no systematic bias. Instead, we assume that the model predictions are normally distributed

---

[2]$N(\mu, \sigma^2)$ denotes a normal (Gaussian) distribution with mean, μ, and standard deviation, σ.

[3]Given the random variable $X \sim N(\mu, \sigma^2)$, the delta method says that approximately $g(X) \sim N\left(g(\mu), (g'(\mu)\sigma)^2\right)$ [5].

[4]The form of Bayes theorem used here states that the posterior distribution is the product of the prior distribution and the likelihood function, normalized by their integral: $f(\theta|E) = p(\theta) f(E|\theta) / \int p(\theta) f(E|\theta) d\theta$. A constant prior is also known as a Jeffreys prior.

about the true values multiplied by a bias factor, δ:

$$M|\theta \sim N\left(\delta\theta, \sigma_M^2\right) \tag{5}$$

The standard deviation, σ_M, is the model-intrinsic uncertainty, *i.e.* model error. This and the bias factor, δ, are the parameters that we seek to quantify. Again, using the log transformation we obtain

$$\ln(M)|\theta \sim N\left(\ln(\delta) + \ln(\theta) - \frac{1}{2}\frac{\widetilde{\sigma}_M^2}{\delta^2}, \frac{\widetilde{\sigma}_M^2}{\delta^2}\right) \quad ; \quad \widetilde{\sigma}_M = \frac{\sigma_M}{\theta} \tag{6}$$

Note that the estimate of the mean includes an extra term in the Taylor series expansion to obtain a more precise estimate of δ. Because $\ln(\theta)$ is random, and we have its posterior distribution from Eq. (4), we can obtain the marginal distribution[5]

$$\ln(M) \sim N\left(\ln(\delta) + \ln(E) - \frac{1}{2}\frac{\widetilde{\sigma}_M^2}{\delta^2}, \frac{\widetilde{\sigma}_M^2}{\delta^2} + \widetilde{\sigma}_E^2\right) \tag{7}$$

Notice that we have expressed the *total relative uncertainty* of the model predictions as a combination of model error and experimental uncertainty. What we need now is a way to estimate this error and the bias factor, δ. First, define:

$$\overline{M} = \frac{1}{n}\sum_{i=1}^{n}\ln(M_i) \quad ; \quad \overline{E} = \frac{1}{n}\sum_{i=1}^{n}\ln(E_i) \tag{8}$$

The least squares estimate of the standard deviation of $\ln(M)$ is

$$u(M) = \sqrt{\frac{\widetilde{\sigma}_M^2}{\delta^2} + \widetilde{\sigma}_E^2} \approx \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left[(\ln(M_i) - \overline{M}) - (\ln(E_i) - \overline{E})\right]^2} \tag{9}$$

Note that $u(M)$ is the *total relative uncertainty* of M. Next, compute the estimate for the model error:

$$\frac{\widetilde{\sigma}_M}{\delta} \approx \sqrt{u(M)^2 - \widetilde{\sigma}_E^2} \tag{10}$$

An estimate of $\ln(\delta)$ is $\overline{M} - \overline{E} + \widetilde{\sigma}_M^2/2\delta^2$, and, thus, the estimate of δ is

$$\hat{\delta} \approx \exp\left(\overline{M} - \overline{E} + \frac{\widetilde{\sigma}_M^2}{2\delta^2}\right) \tag{11}$$

In essence, this procedure gives us a way to estimate the "true" value, θ, using only a predicted value, M, and the bias factor and relative error for that particular quantity that have been obtained using past validation experiments. The "truth" is a random variable with the following distribution:

$$\theta \sim N\left(\frac{M}{\hat{\delta}}, \left(\frac{M\widetilde{\sigma}_M}{\hat{\delta}}\right)^2\right) \tag{12}$$

This is essentially a recipe for estimating the error associated with a model prediction, M. Below, we describe how one might make use of this formula in practice. First, however, we need to verify the accuracy of the procedure just described.

---

[5] $f(M) = \int f(M|\theta)f(\theta)\,d\theta$

6 Verifying the Procedure

The statistical analysis described in the previous section is difficult to understand without a fairly good background in Bayesian analysis. However, the calculation itself is no more difficult than determining means and standard deviations of a few columns of numbers and it can be easily done with a simple spreadsheet program.

To better illustrate the process, and also to verify this procedure, suppose we generate 1000 uniformly distributed random numbers, θ_i, between 0 and 1000. We say that these numbers represent a particular quantity of interest, like a gas temperature at a particular point and at a particular time, and we also say that these values have no uncertainty – they are the "truth." Of course, we can never know the true values, but for this hypothetical exercise assume we can. Next, for each θ_i, we randomly choose a value that is to represent a hypothetical measurement, E_i, from a normal distribution whose mean is θ_i and whose relative uncertainty, $\widetilde{\sigma}_E$, is assumed known. In the same way, we randomly choose a value that is to represent a hypothetical model prediction, M_i, from a normal distribution whose mean is $\delta\theta_i$ and whose relative error, $\widetilde{\sigma}_M$, is specified for this hypothetical exercise. What we have now created are 1000 pairs of (E_i, M_i) with which we can test the procedure outlined in the previous section. In other words, we ought to be able to take our 1000 pairs of values, and the experimental uncertainty, $\widetilde{\sigma}_E$, and estimate the model bias, $\hat{\delta}$, and relative error, $\widetilde{\sigma}_M$.

We consider two examples. For the first example, assume that the model has no bias $(\delta = 1)$ and that the relative uncertainty of the measurements and the relative error of the model are both 0.1, or 10 %. The scatter plot on the left side of Fig. 3 displays the model predicted values compared to the true values. The dashed lines indicate the 95 % confidence interval; that is, it is expected that 95 % of the points should fall between these lines, whose slopes are plus and minus 20 % $(2\widetilde{\sigma}_M)$ of the diagonal line. Of course, this plot cannot exist in a real situation, because the true values are never known. Instead, the only way to present the data is via the scatter plot on the right side of Fig. 3. Here, the measured values, E_i, are compared with the predicted values, M_i. The same dashed lines are carried over from the plot on the left. Because the predicted values are being compared with measurements that have uncertainty, it appears that the model error is greater than it actually is. In this hypothetical example, the model is assumed to be as accurate as the measurements, yet the comparison makes it seem as if the model is less accurate than the experiments. The procedure outlined above, which makes use of only the measured and predicted values, is able to extract from the data the fact that the hypothetical model has a relative error of 10 %, not the roughly 15 % that one would infer from the plot if the experimental uncertainty were not taken into account.

For the second example, we make our hypothetical model predictions and experimental measurements more realistic. Assume now that the measured value is the HGL temperature, and that the relative uncertainty of the measured values, $\widetilde{\sigma}_E = 0.07$, is obtained from Table 1. We assume the model has a bias factor of 1.03 and its relative error is 0.10 in order to generate a set of hypothetical model predictions. A graphical representation of the data is shown in Fig. 4. Note that the experimental uncertainty bounds are represented by thick, dashed lines, and the model bias and relative error is shown by the lighter, dashed lines. Each are expressed as 95 % confidence intervals $(2\widetilde{\sigma})$.

The statistical procedure outlined above ought to provide estimates of the values of the bias factor, δ, and the model error, $\widetilde{\sigma}_M$, whose values we used in creating the hypothetical model predictions. The results of seven random trials are shown in Table 2. It is not expected that the estimates
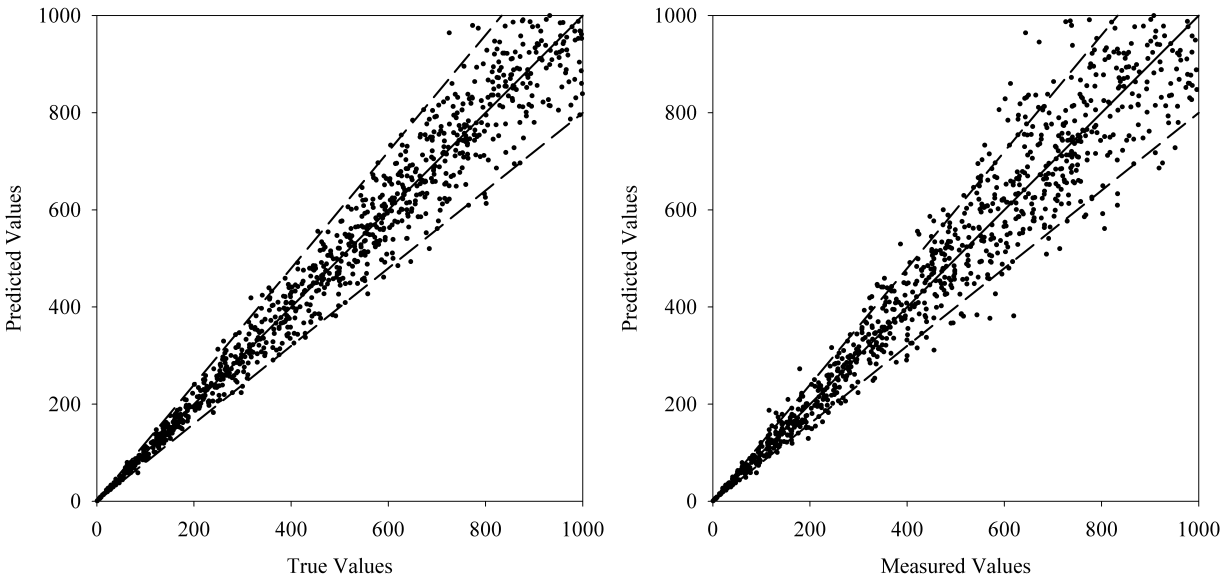
8

Figure 3: (Left) A comparison of predicted and "true" values for 1000 hypothetical experiments in which the model predictions and experimental measurements have the same accuracy. (Right) The same data, except now the predicted values are compared to measured values. On both plots, the uncertainty bounds apply to both the predicted and measured values.

will be exact because they are based on truncated Taylor series approximations. However, given the fact that the experimental uncertainty estimate, $\widetilde{\sigma}_E$, is often only a gross approximation in its own right, the accuracy of the procedure is more than adequate.

7   Making Use of the Model Error

The previous sections describes a method of quantifying the model error by comparing its predictions with experimental measurements. But how does one make use of the computed model bias and relative error? This is best answered with an example. Suppose the model is being used to estimate the likelihood that electrical control cables could be damaged due to a fire in a particular compartment. Damage is assumed to occur when the surface temperature of any cable reaches 400 °C. It is also assumed that the fire is ignited within an electrical cabinet and the heat release rate of the fire is a specified function of time, and that all other input parameters for the model are known and provided. Finally, it is assumed, for the time being, that there is no uncertainty associated with any of these assumptions. What is the likelihood that the cables would be damaged if that fire were to occur? The calculation is performed, and the model predicts that the maximum surface temperature of the cables is 350 °C. Does this mean that there is no chance of damage, assuming that the input parameters and assumptions are not in question at the moment? The answer is no, because we know that the model itself is subject to error. So what is the chance that the cables could actually reach temperatures greater than 400 °C?

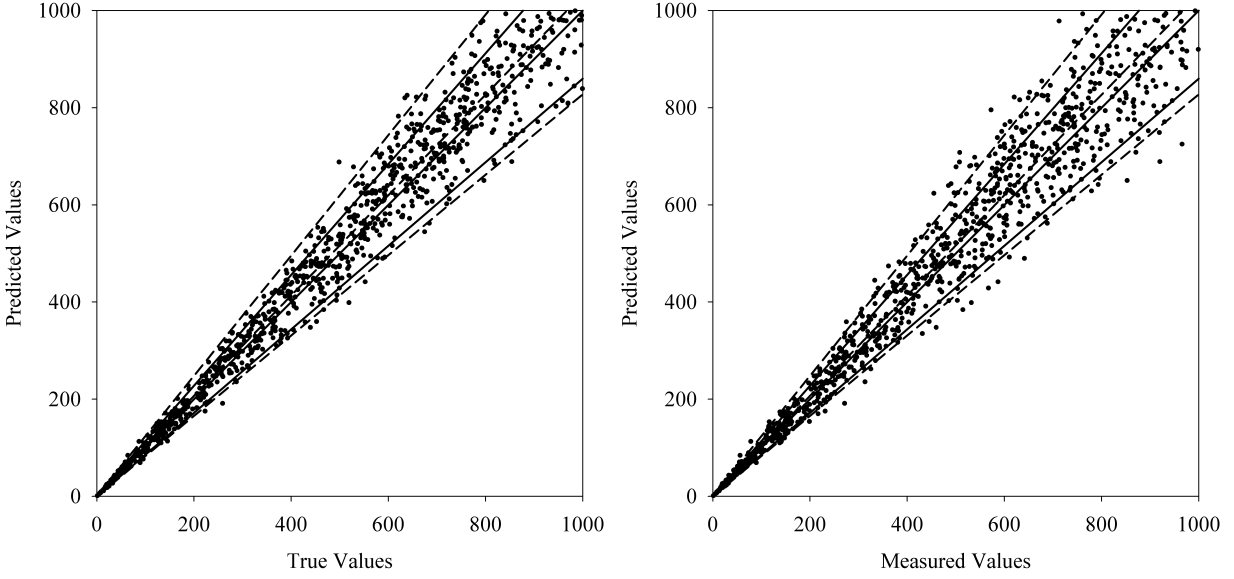Before we can answer this question, we need to consider past experiments for which model

9

Figure 4: (Left) A comparison of predicted and "true" values for 1000 hypothetical experiments. The short dashed lines indicate the bias, $\delta = 1.03$, and the relative error, $2\widetilde{\sigma}_M = 0.20$, of the model predictions. The solid, off-diagonal lines indicate the relative uncertainty, $2\widetilde{\sigma}_E = 0.14$ of the experimental measurements. (Right) The same data, except now the predicted values are compared to measured values. On both plots, the uncertainty bounds are the same.

predictions have been compared to measured surface temperatures of objects with similar thermal characteristics as the cables in question. How "similar" the experiment is to the hypothetical scenario under study can be quantified by way of various parameters, like the thermal inertia of the object, the size of the fire, the size of the compartment, and so on. Next, the results of the validation study can be analyzed following the procedure spelled out above, which provides us with an estimate of the bias factor, δ, and relative error, $\widetilde{\sigma}_M$, for the model predictions of this particular quantity. Let us say, for the sake of argument, that the bias factor is 1.03; that is, on average, the model has been shown to over-predict surface temperatures by 3 %. Let us also say that the relative error has been calculated and it is 0.10 . These are the same values that were assumed in the second example of the previous section. Now, consider the graph shown in Fig. 5. The vertical lines indicate the "threshold" temperature at which damage is assumed to occur (400 °C), and the temperature predicted by the model (350 °C). Given an ambient temperature of 20 °C, the predicted temperature rise, M, is 330 °C. The bell curve is a normal distribution obtained from Eq. (12) above. Its mean and standard deviation are

$$20 + \frac{M}{\delta} = 20 + \frac{330}{1.03} \quad ; \quad \widetilde{\sigma}_M \frac{M}{\delta} = 0.10 \frac{330}{1.03} \tag{13}$$

respectively. The shaded area beneath the bell curve is the probability (0.03 in this case) that the "true" temperature can exceed 400 °C. This means that there is an 3 % chance that the cables could become damaged *based solely on the fact that the model is not a perfect representation of reality.*

The obvious question to ask at this point is what do we do if we cannot assume that the speci-

10

Table 2: Estimated bias and relative error from six trials used to verify the analysis.

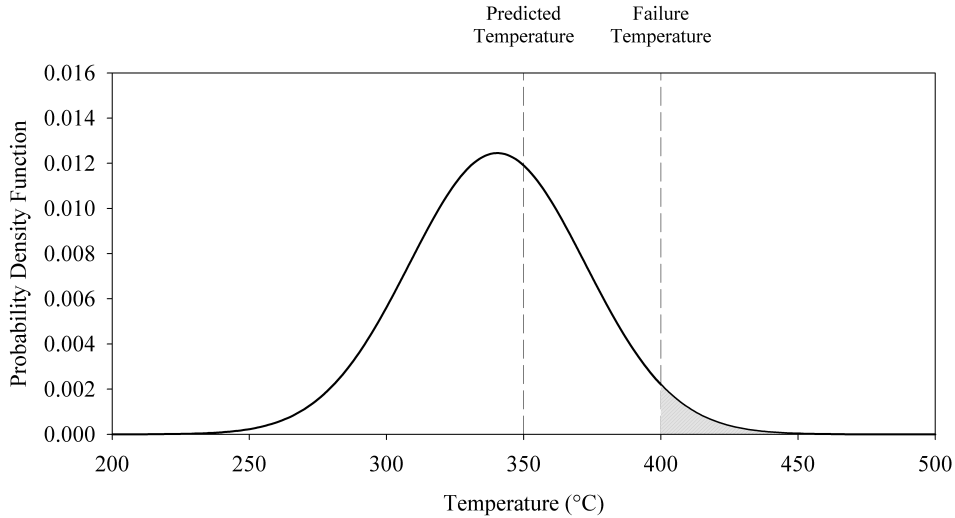| Trial	Bias Factor	Relative Error
Exact	1.030	0.100
1	1.030	0.103
2	1.035	0.102
3	1.033	0.105
4	1.037	0.104
5	1.041	0.105
6	1.034	0.101
7	1.034	0.103

Figure 5: Plot showing how one might make use of the model error.

fied fire, material properties, and other input parameters are known exactly? How does that affect our estimate of the likelihood of cable damage? The procedure above has only provided a way of expressing the model error. What about the so-called "user effects"? What if the specified fire in the electrical cabinet is actually chosen from a distribution of heat release rates? What if the material properties of the cables are not known exactly? Assuming one could quantify the uncertainty of all of the input parameters, and assuming that the model error and the input uncertainty are uncorrelated, it is possible to combine the two following the procedure that was described above that is used to determine the combined experimental uncertainty. Of course, the uncertainty associated with the input parameters would need to be quantified and propagated through the model. The end result would be a widening of the distribution shown in Fig. 5 and an increase in the likelihood of cable damage, assuming that the parameters used in the "base case" were all taken as the mean values of their respective distributions. In fact, depending on the scenario, the uncertainty associated with the input parameters far outweighs the model error.

8 Limitations

The above verification exercises are valuable in assuring that the procedure works as designed, but it also points out a few issues that need to be addressed. First, any statistical procedure is based on the law of averages, or, in other words, more data is better than less. It is usually not possible to conduct a large number of fire experiments to assess the accuracy of the model in predicting each quantity of interest. Sometimes we only have a few data points with which to estimate the model error. Worse yet, we may not have the necessary information about the experimental procedure to estimate the uncertainty of the reported measurements. In such cases, it may be better to simply present the comparison of model prediction and experimental measurement as a series of plots like Fig. 1, or in the form of a scatter plot (Fig. 2) without any uncertainty or error bars. The value of the entire validation process as outlined above is that it is possible at any step along the way to simply stop and accept the raw output of the study as the basis for making an assessment of the model. The uncertainty analysis and error quantification is valid only to the extent that sufficient experimental data with quantified uncertainty estimates is available. It does more harm than good to attempt to quantify the model error with insufficient means to do so.

Another concern with the above procedure is that an over-estimate of the experimental uncertainty will naturally result in an under-estimate of the model error. Keep in mind that the model can never be declared more accurate than the experimental measurements against which it is compared. This rule is demonstrated mathematically by Eq. (10) where it is observed that an over-estimate of the experimental uncertainty, $\tilde{\sigma}_E$, can result in the square root of a negative number, an imaginary number. In the case that the computed model error is less than the estimated experimental uncertainty, then the latter must be re-evaluated, or the number of data points needs to be seriously questioned.

9 Conclusion

We have outlined a procedure to estimate model error by way of comparisons of model predictions with experimental measurements whose uncertainty has been quantified. For clarity of presentation, issues associated with the selection of experiments, metrics of comparison, and presentation of results, have not been discussed in detail because these decisions are application-specific and best left to the end user or AHJ. The overall procedure is not a dramatic departure from current methods of model validation, but it does present a very tractable method of distinguishing the model *error* from the *uncertainty* of the model inputs and the measurements against which the model is compared. Too often these various forms of uncertainty are lumped together, in which case there is no way to know what part of the model, if any, needs improvement. For this reason, this methodology is of tremendous value to model developers because, in a sense, it strips away what are often referred to as "user effects" to reveal inadequacies of the model itself. The goal of the model developers is to reduce the model error as much as possible, even though uncertainties in the input parameters may never be eliminated completely.

References

[1] American Society for Testing and Materials, West Conshohocken, Pennsylvania. *ASTM E 1355-04, Standard Guide for Evaluating the Predictive Capabilities of Deterministic Fire Models*, 2004.

[2] R.D. Peacock, P.A. Reneke, W.D. Davis, and W.W. Jones. Quantifying fire model evaluation using functional analysis. *Fire Safety Journal*, 33:167–184, 1999.

[3] K. Hill, J. Dreisbach, F. Joglar, B. Najafi, K. McGrattan, R. Peacock, and A. Hamins. Verification and Validation of Selected Fire Models for Nuclear Power Plant Applications. NUREG 1824, United States Nuclear Regulatory Commission, Washington, DC, 2007.

[4] W.D. Walton and P.H. Thomas. *SFPE Handbook of Fire Protection Engineering*, chapter Estimating Temperatures in Compartment Fires. National Fire Protection Association, Quincy, Massachusetts, 4th edition, 2008.

[5] G. Oehlert. A Note on the Delta Method. *The American Statistician*, 46(1):27–29, 1992.