

SAND99-1256  
Unlimited Release  
Printed May 1999

## **Statistical Validation of Engineering and Scientific Models: Background**

Richard G. Hills  
Department of Mechanical Engineering  
New Mexico State University  
Las Cruces, New Mexico 88003

Timothy G. Trucano  
Computational Physics Research and Development  
Sandia National Laboratories  
P. O. Box 5800  
Albuquerque, New Mexico 87185-0819

### **Abstract**

A tutorial is presented discussing the basic issues associated with propagation of uncertainty analysis and statistical validation of engineering and scientific models. The propagation of uncertainty tutorial illustrates the use of the sensitivity method and the Monte Carlo method to evaluate the uncertainty in predictions for linear and nonlinear models. Four example applications are presented; a linear model, a model for the behavior of a damped spring-mass system, a transient thermal conduction model, and a nonlinear transient convective-diffusive model based on Burger's equation. Correlated and uncorrelated model input parameters are considered. The model validation tutorial builds on the material presented in the propagation of uncertainty tutorial and uses the damp spring-mass system as the example application. The validation tutorial illustrates several concepts associated with the application of statistical inference to test model predictions against experimental observations. Several validation methods are presented including error band based, multivariate, sum of squares of residuals, and optimization methods. After completion of the tutorial, a survey of statistical model validation literature is presented and recommendations for future work are made.

## **Acknowledgements**

This report is an account of contract research (Doc. # AX-0620) performed on the part of the first author during the 1998 Fiscal Year. The second author served in the capacity as the technical project monitor. The authors thank Rob Easterling, Ben Blackwell, Ed Boucheron, and Dan Carroll for reviewing this manuscript.

# Table of Contents

Abstract .....	i
Acknowledgements .....	ii
Table of Contents .....	iii
List of Figures .....	v
List of Tables .....	vii
1.0 Introduction .....	1
1.1 Purpose and Organization of this Document .....	1
2.0 Propagation of Uncertainty .....	3
2.1 Introduction .....	3
2.2 Characterizing Uncertainty .....	3
2.3 Uncertainty in Two Model Parameters: Linear Model .....	9
2.4 Resulting Uncertainty in Model Prediction .....	13
2.4.1 Monte Carlo Method .....	13
2.4.2 Direct Evaluation for a Linear Model .....	16
2.4.3 Sensitivity Analysis .....	17
2.5 Simple Nonlinear Model .....	18
2.6 Diffusion Equation .....	24
2.7 Hyperbolic Equation .....	30
2.8 Summary .....	35
3.0 Model Validation .....	37
3.1 Two Approaches to Validation .....	39
3.2 Model Validation for the Damped Spring-Mass System .....	40
3.3 Simple Comparisons .....	41
3.4 Monte Carlo Model Testing using One Measurement .....	45
3.5 Model Testing for Multiple Measurements .....	46
3.6 Normally Distributed PDF .....	48
3.7 The Problem with the Sum of Squares of Residuals .....	52
3.8 An Alternative Approach .....	52
3.9 Engineering Validation and Approximate Models .....	56
3.10 What Does This All Mean? .....	59
4.0 Literature Review .....	61
4.1 Introduction .....	61
4.2 Measures .....	61
4.3 Residuals .....	62
4.4 Regression .....	64
4.5 Error Bounds .....	64
4.6 General Methods .....	65
4.7 Propagation of Uncertainty .....	65

4.8 Propagation of Uncertainty for Model Validation .....	66
4.9 Optimization Methods .....	67
5.0 Conclusions and Recommendations .....	70
6.0 References .....	74

## List of Figures

Figure 2.1. Histogram of Thermal Conductivity.....	6
Figure 2.2. Model for Histogram of Thermal Conductivity .....	8
Figure 2.4. Probability Density Functions of Parameters $a$ and $b$ . .....	10
Figure 2.5. Joint Probability Density Plots for the $a$ and $b$ : $a$ and $b$ are uncorrelated. ....	11
Figure 2.6. Joint Probability Density Plots for the $a$ and $b$ : $a$ and $b$ are correlated. ....	11
Figure 2.7. Histograms for Sample Counts of $p(T)$ .....	14
Figure 2.8. Probability Density Plots for Normal Distributions Defined in Table 2.4. ....	15
Figure 2.9. Predicted $X/F$ : Solid line based on means of model parameters; points are based on means of Monte Carlo predictions. ....	21
Figure 2.10. Predicted Standard Deviation for $X/F$ : Sensitivity analysis results are represented by the solid line; Monte Carlo results are represented by the points. ....	22
Figure 2.11. Predicted Populations for $X/F$ : $w=1.0$ ; Monte Carlo predictions are shown by the histogram; scaled Normal PDF based on sensitivity analysis statistics are shown by the solid curve. ....	22
Figure 2.12. Predicted Populations of $X/F$ : $w=1.5$ ; Monte Carlo predictions are shown by the histogram; scaled Normal PDF based on sensitivity analysis statistics are shown by the solid curve. ....	23
Figure 2.13. Predicted Maximum $X/F$ at 99% Confidence Level: Monte Carlo results represented by points; sensitivity analysis results represented by the solid curve. ....	24
Figure 2.14. Predicted Mean Centerline Temperature: Monte Carlo results represented by points; sensitivity analysis results represented by the solid curve.....	27
Figure 2.15. Predicted Standard Deviation for Centerline Temperature: Monte Carlo results represented by points; sensitivity analysis results represented by the solid curve. ....	27
Figure 2.16. Predicted Populations for Centerline Temperature: Monte Carlo results represented by histograms; scaled sensitivity analysis results represented by solid curve. ....	28
Figure 2.17. Predicted Mean $T = 0.5$ Isotherm Location: Monte Carlo results represented by points; sensitivity analysis results represented by the solid curve.....	28
Figure 2.18. Predicted Standard Deviation for $T = 0.5$ Isotherm Location: Monte Carlo results represented by points; scaled sensitivity analysis results represented by solid curve. ....	29
Figure 2.19. Predicted Populations for $T = 0.5$ Isotherm Location: Monte Carlo predictions -represented by histograms; scaled sensitivity analysis represented by solid curve. ....	29
Figure 2.20. Mean Velocity as a function of $x$ : $t = 0.0, 0.5, 1.0, 1.5, 2.0, 2.5$ .....	33
Figure 2.21. Velocity as a function of $x$ : One Monte Carlo sample, $t = 0.0, 0.5, 1.0, 1.5,$ $2.0, 2.5$ .....	34
Figure 2.22. Mean Front location as a Function of Time: Monte Carlo results represented by the points (8192 simulations); sensitivity analysis results represented by the solid curve. ....	34

Figure 2.23 Standard Deviation for Front location as a Function of Time: Monte Carlo results with 16384 simulations represented by long dashes, 8192 simulations represented by short dashes; sensitivity analysis results represented by the solid curve. ....	35
Figure 3.1. Measured and Predicted Spring/Mass/Damper Response: Predicted response represented by the solid line, measured response represented by the solid points. ....	42
Figure 3.2. Predicted Spring/Mass/Damper Response including Measurement Errors: Predicted response represented by solid line and the error bars, measured response represented by the solid points. ....	43
Figure 3.3. Measured and Predicted Spring/Mass/Damper Response including Parameter Uncertainty: Predicted response represented by solid line and error bars, measured response represented by the solid points. ....	44
Figure 3.4. Measured and Predicted Spring/Mass/Damper Response including Parameter Uncertainty and Measurement Error: Predicted response represented by the solid line and error bars, measured response represented by the solid points. ....	44
Figure 3.5. Histogram for Monte Carlo Predictions at $w=1.5$ : Predicted mean given by the thick line, experimental measurement given by the thin line, 95% confidence bounds given by the dashed lines.....	46
Figure 3.6. Contour Plot for Histogram Counts for Monte Carlo Predictions: Solid contour lines are for Counts = 125, 250, 500, 750. The dashed line represents the 95% confidence region. The heavy point represents the measurement pair ( $X/F(1.5)$ , $X/F(1.0)$ ). The Monte Carlo Predictions were counted into a 50 x 50 grid of bins.....	48
Figure 3.7. Equal Probability Contours for the Multivariate Normal Distribution: Solid contour lines are for probabilities = 25%, 50%, 75%. The dashed line represents the 95% confidence region. The light point represents the measurement pair ( $X/F(1.5)$ , $X/F(1.0)$ ). The heavy point represents the mean prediction. ....	50
Figure 3.8. Comparison of the 95% Confidence Regions: Monte Carlo results represented by the dashed curve. Normal approximation to the represented by the solid curve. The light point represents the measurement pair ( $X/F(1.5)$ , $X/F(1.0)$ ). The heavy point represents the mean prediction. ....	51
Figure 3.9. 95% Joint Confidence Region and Model Predictions: The point is at ( $c_{\text{mean}}$ , Measured $X/F(1.5)$ ). ....	54
Figure 3.10. Distance $r$ as a Function of $c$ : The critical value of $r$ is represented by the horizontal dashed line.....	55
Figure 3.11. Engineering Validation: The ellipse represent the 95% confidence region of the model prediction (including measurement error). The heavy point represents the measurement pair. The rectangle region represents the acceptable additional error ( $\pm 10\%$ of mean prediction). The light point represents the most probable model prediction lying in the rectangular region. ....	58

## List of Tables

Table 2.1.	Estimated $k_{mean}$ and $\mathbf{s}$ .....	7
Table 2.2.	Statistics for the Model Parameters .....	10
Table 2.3.	Random Numbers for $a$ and $b$ .....	13
Table 2.4.	Estimated $p_{mean}$ and $\mathbf{s}$ .....	15
Table 2.5.	Direct Evaluation of $p_{mean}$ and $\mathbf{s}$ .....	16
Table 2.6.	Statistics for the Oscillator Model Parameters: The parameters are normally distributed.....	19
Table 2.7.	Statistics for the Burger Model Parameters .....	31
Table 3.1.	Statistics for the Oscillator Parameters and Measurement Errors: All quantities are assumed to be normally distributed.....	41
Table 3.2.	Oscillator Parameters and Measurement Errors.....	57

(Page Left Blank)

## 1.0 Introduction

The use of numerical models for the simulation of physical systems has greatly affected our approach to engineering and science. These models are used to design commercial and military equipment, to design scientific experiments and analyze the results, and to perform what-if studies. The increased ease-of-use, the increased ability to model complex phenomena, and the lowering cost of modern computers have accelerated the use of numerical models. These models are reducing the design cycle time and cost, and are increasing the reliability of the resulting products.

The increased dependence on computer models leads to the natural question - how accurate is the model? Traditionally, modelers have tested their models against experimental data whenever possible. These comparisons often take the form of comparisons of predictions to measurements through simple x-y plots, scatter plots, or 2 dimensional contour plots. After plotting the results, we are faced with two questions: When is the agreement between experimental measurements and model predictions good enough? How does one measure this agreement?

The issue of model validation is very complex and there are probably as many opinions on model validation as there are workers in the field. In the present work, we will focus on one aspect of model validation - the actual process of comparing model predictions to experimental observations. As will be shown, this comparison is complicated by the presence of measurement error, by the inability to repeat the validation experiment many times to obtain significant statistical sampling, and by the effects of nonlinearity in the models. The focus on the comparison may seem, at first glance, to be restrictive. However, the rigorous quantification of this comparison process using scientific methodology allows us to make definitive statements about what we have really shown by a validation experiment. Perhaps more importantly, the development of these concepts and tools will allow us to better design future validation experiments.

### ***1.1 Purpose and Organization of this Document***

The primary purpose of this document has changed during its development. Initially, the purpose was to provide a literature survey of model validation methodology applicable to problems of interest to Sandia National Laboratories (SNL) and to define critical research needs and issues. Later, a brief tutorial discussing some of the basic concepts of propagation of uncertainty through models and statistical model validation for those scientists or engineers unfamiliar with probabilistic methods was to be added. However, as the tutorial examples were developed, we quickly realized that these examples could be structured to provide a more effective mechanism to communicate issues associated with uncertainty analysis and to introduce important statistical concepts associated with

model validation. As a result, the focus of this document has moved away from a traditional literature survey toward an example intensive tutorial.

The tutorial will focus on two topics: Propagation of uncertainty analysis and statistical model validation methodology. The chapter on propagation of uncertainty analysis has several objectives. First, we want to provide those unfamiliar with this field with an understanding of the basic issues and concepts associated with the propagation of uncertainty through a model. Second, we wish to demonstrate what such an analysis can provide. Finally, we provide a foundation for the following chapter on model validation methodology. Because Sandia National Laboratories has an ongoing program in uncertainty analysis, the discussion presented here will be at a level appropriate for those who are not active in the field.

The chapter on model validation builds on the previous chapter on uncertainty analysis. We present a tutorial using one of the simpler nonlinear models developed in the earlier chapter. While the methodology is demonstrated through simple examples, it is applicable to much more complex problems. Much of this material will be new to many readers and is based on ideas obtained from the hydrology, geophysics and the statistics literature. The methodology presented provides a more rigorous approach to model validation than that typically used in the past.

After the basic concepts have been developed in these two chapters, we will review the literature to establish the state-of-the-art relative to our interests. This review will help identify where we are and where we have to go to move the field of scientific model validation forward.

## **2.0 Propagation of Uncertainty**

### **2.1 Introduction**

Here we introduce some of the basic concepts related to the estimation of uncertainty in model predictions where statistical uncertainty in the model parameters occurs. Model parameters can be physical properties such as thermal conductivity or viscosity, the constants that appear in models for physical properties, or parameters that appear in boundary or initial conditions. Model parameters are generally estimated from experiment, or their values are controlled in some fashion (such as holding a boundary temperature at some desired parametric value). Those parameters representing properties, such as viscosity, can be dependent on the actual samples taken to measure these properties. Due to natural variation in the manufacturing process, uncertainty will exist as to the actual value of these parameters. This uncertainty is compounded by the uncertainty of experimental measurement. An important question is, what is the uncertainty in the model prediction given knowledge of the uncertainty in the model parameters?

In this chapter, we focus on two of the most common methods used to predict uncertainty of model predictions due to uncertainty in model parameters. These are the Monte Carlo method, which is based on simulating a population of possible samples; and the sensitivity method, which is based on moments of the sample population. We do this through the use of simple example models which possess the features of much more complex systems. These models range from simple linear algebraic models to models associated with nonlinear transport.

### **2.2 Characterizing Uncertainty**

A common feature of engineering models is the need to provide values for model parameters that define properties, or parameters which appear in boundary or initial conditions. For example, Fourier's law states that heat flux is proportional to the negative of the temperature gradient. The proportionality constant is called the thermal conductivity. Its value depends on the composition of the conducting material and can also depend on other quantities, such as temperature and pressure. The value for this model parameter is estimated using experimental techniques utilizing samples of the material. Unfortunately, the methods used to characterize thermal conductivity will always lead to errors in the measured conductivity due to experimental technique and equipment, due to uncertainty associated with uncontrolled environmental conditions, and due to uncertainty associated with variability from sample to sample because of the manufacturing process. A key question is how do we quantitatively characterize the sources and nature of this uncertainty.

Consider another example. We know that the chemical composition and the annealing process control the yield stress of steel. We would expect that there would be differences in the yield stress from manufacture to manufacture simply because of the different manufacturing processes used. We would also expect some differences in batch to batch from a single manufacture due to changes in the performance of equipment over time and small differences in chemical composition of the materials provided by the suppliers. We also expect to have variability in yield stress from within a batch. Since materials cool faster near their surfaces than in their interiors, the yield stress can vary significantly over the volume of a billet. These properties can change further due to the changes to grain structure that normally occurs when forming the steel billets into structural sections.

We see that the variability in our estimates for yield stress for our material depends on many factors. An additional factor is the spatial scale of the material that we are using in our system. This effect is especially important for heterogeneous material. For example, let us say we are interested in the thermal conductivity of a one-meter cube of earth. We expect that this cube, if naturally occurring, will possess significant variability in many of its properties across the cube. If we were to take many small samples from the cube, we would find that there is significant variability in the conductivity. If our analysis depends on only the mean behavior of conductivity on the scale of 1 meter, then we could perform an experiment designed to estimate the effective conductivity using the entire 1 meter cube of earth. Alternatively, we could attempt (not always an easy task) to develop a model that relates mean behavior at 1 meter to the conductivity of the many samples taken from the cube on a much smaller scale (say, 1 cm sections). This model could be as simple as the arithmetic mean, or it can be as complex as a weight average based on the dynamic behavior of the expected temperature gradient in the meter cube during the actual event being modeled.

Once we have identified which source of uncertainty is important for our application, we can use sample conductivities from the appropriate sources (i.e., many batches, one batch, or a sub-batch) to generate some model for the uncertainty. The most common and accepted method to accomplish this is to use methods from probability and statistics. To characterize this uncertainty, we randomly take representative test samples (of the appropriate scale) from the material or material batches of interest, perform experimental measurements, and estimate the thermal conductivity for each sample. We then plot a histogram of the number of samples that our procedure produced in different ranges of conductivity. Such histograms are shown in Figure 2.1 for the conductivities of 50, 500, and 5000 hypothetical samples. Note that there is variability in the conductivity with values ranging from 1.2 to 1.7 conductivity units. Also note that some values for thermal conductivity are more likely than others. For example, the histograms indicate that samples with a thermal conductivity near 1.5 are more likely than samples with a conductivity near 1.2. Finally, note that as the number of samples increases, the shape of the histogram tends to become smoother. As will be shown later, this histogram for

uncertainty can be used to estimate the uncertainty in a model's prediction due to the uncertainty in conductivity from sample to sample.

We can look at this histogram from two perspectives. First, we can consider the measured conductivities of the samples to give us the entire set of all possible thermal conductivities. Of course, we know that this is not true for a real quantity such as thermal conductivity. If we were to take a new sample, its measured conductivity would likely be slightly different from the conductivities of any of the existing samples. Because of this, we generally look at this finite set of samples from another perspective. We assume that the samples we gathered represent only a subset of all possible conductivities that could have been measured. If our sample set is a good representation of the entire set or population of conductivities, we should be able to use this set to characterize the entire population. This characterization could then be used to describe the uncertainty in conductivity of any sample taken from the entire population of samples for the material of interest.

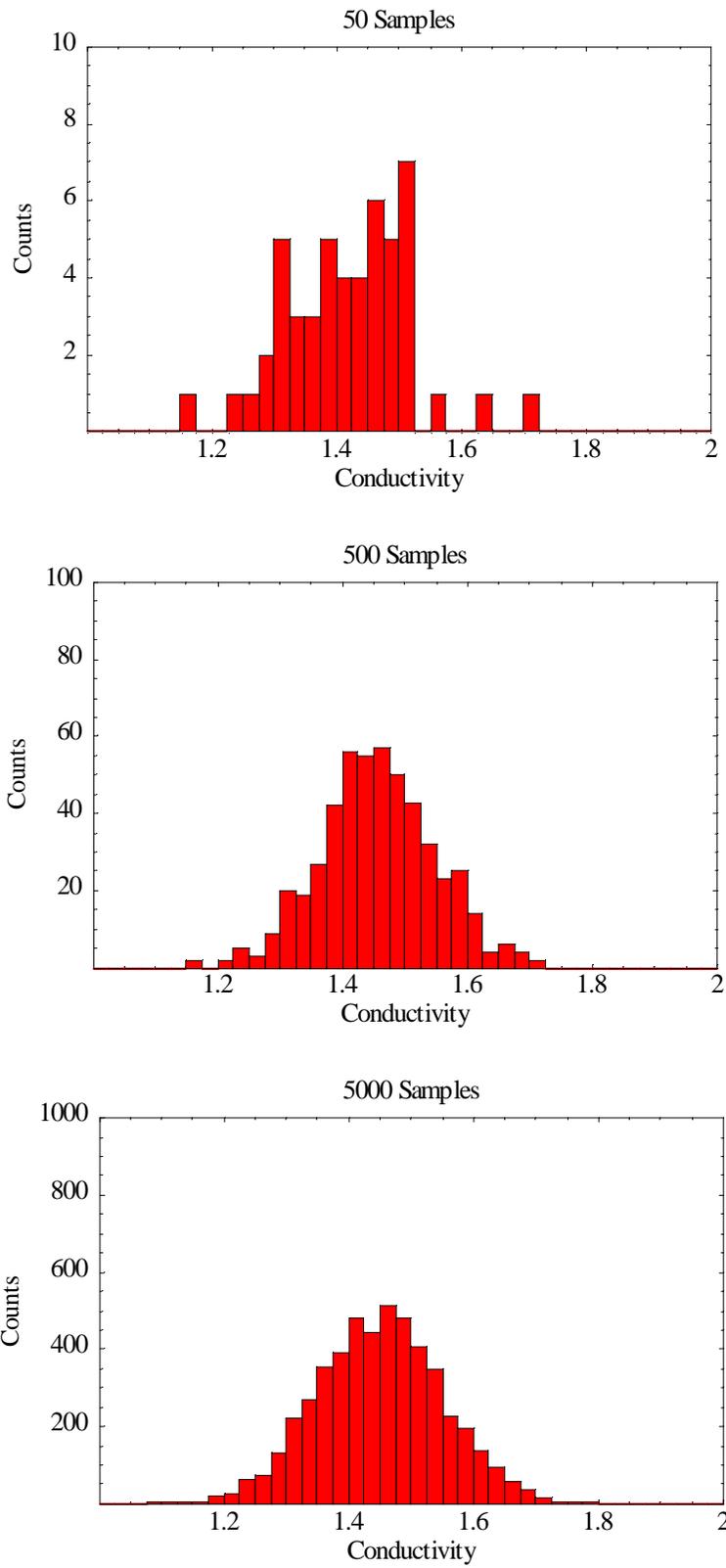
How do we characterize the distribution of conductivities for the entire population? One characteristic that we are almost always interested is some measure of the center of the distribution. For example, inspection of Figure 2.1 indicates that the center appears to be somewhere around 1.45. How do we mathematically define this center? The most common method is through the arithmetic mean. An estimate for this mean given  $n$  samples is

$$\langle k_{mean} \rangle = \frac{1}{n} \sum_{i=1}^n k_i \quad (2.1)$$

We use triangular brackets to indicate that this is an estimate for the mean of the entire population using a subset of samples from the population. There are many other mathematical measures for the center of the distribution. The median, for example, is that value for which there are as many samples with larger values as there are with lower values.

Another important property of a distribution is some measure of the dispersion or spread of the distribution about its center. The most common measure is the standard deviation,  $\mathbf{s}$ , which quantifies dispersion about the mean. The standard deviation of the entire population can be estimated from the samples as follows:

$$\langle \mathbf{s} \rangle = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (k_i - \langle k_{mean} \rangle)^2} \quad (2.2)$$



**Figure 2.1. Histogram of Thermal Conductivity**

The square of the standard deviation is called the variance.

$$\text{var}(k) = s^2 \quad (2.3)$$

Estimates for the population mean and standard deviation are given in Table 2.1 for the samples represented by the histograms of Figure 2.1. Note that as the number of samples change, the estimated values for the mean and standard deviation change. Statistical methods can be used to assess the quality of these estimates (see any standard statistics textbook). As the number of samples increase, we expect the accuracy of the estimates will also increase.

**Table 2.1. Estimated  $k_{mean}$  and  $s$**

# Samples	$k_{mean}$	$s$
50	1.419	0.1036
500	1.456	0.0926
5000	1.499	0.1007

While the mean and standard deviation are two of the most used quantities in statistics, they do not define the shape of the histogram. It is often useful to have a mathematical model for this shape. For example, the shape of the 5000 sample histogram in Figure 2.1 is well approximated by the equation

$$\text{Counts} = 495 e^{-49.3(k-1.499)^2} \quad (2.4)$$

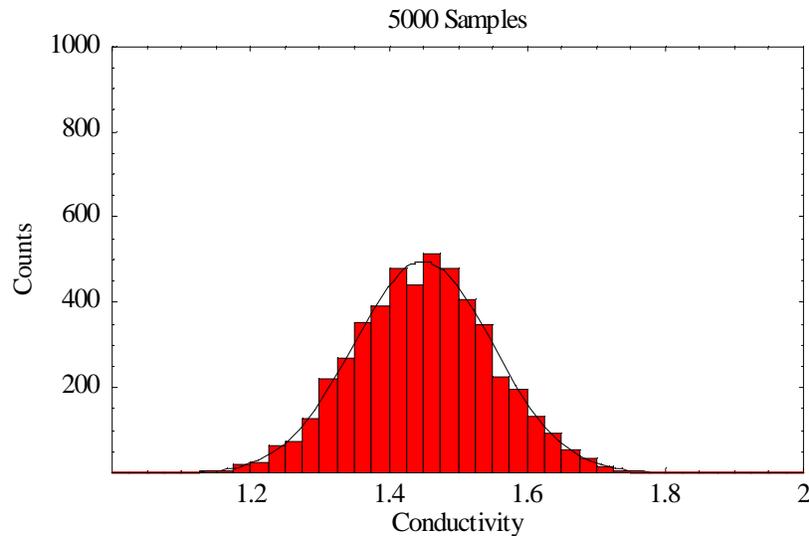
A comparison of this approximation and the 5000 sample histogram is shown in Figure 2.2. We can re-interpret the results in Figure 2.2 from a slightly different perspective. This histogram indicates that of the 5000 samples, approximately 400 were taken over the conductivity range from 1.5 to 1.525. We can thus say that the probability that a sample will result in a conductivity between 1.5 and 1.525 is approximately  $400/5000=0.08$ . This suggests that Eq. (2.4) can be used to develop a model for a probability distribution. Since the probability of obtaining samples over the entire range of conductivities must be unity, we can re-scale Eq. (2.4) such that the area under the curve is unity. Re-scaling Eq. (2.4) accordingly gives

$$\text{PDF}(k) = 3.962 e^{-49.3(k-1.499)^2} = \frac{1}{\sqrt{2ps}} e^{-\frac{1}{2}\left(\frac{k-k_{mean}}{s}\right)^2} \quad (2.5)$$

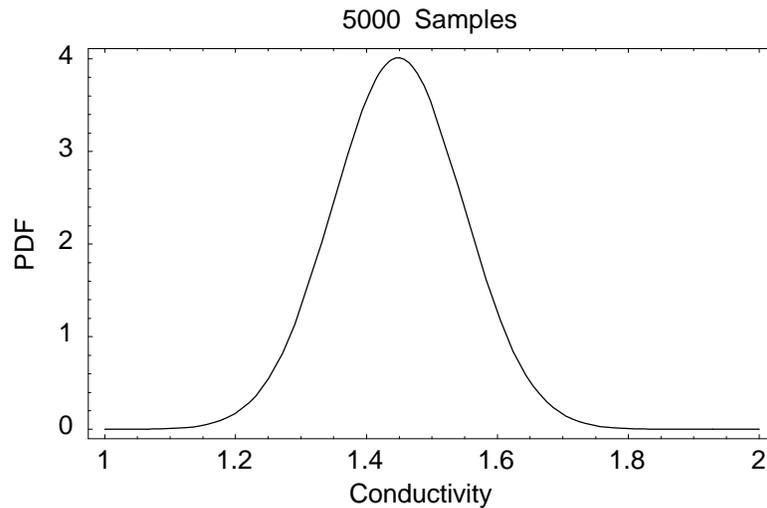
Note that Eq. (2.5) gives a value for each value of  $k$  rather than the probability of falling in a small, but finite range of  $k$ . To denote this difference, we call this function the probability density function (PDF). This particular probability density function (i.e., Eq.

(2.5)) is known as the Normal probability density function. It is the most common PDF and can be evaluated directly from the mean and standard deviation of the population as indicated in the last part of this equation. The resulting probability density function is plotted in Figure 2.3. Since Eq. (2.4) or (2.5) were evaluated using a subset of the entire population of conductivities, the resulting PDF represents an estimate for the true population probability density function. Note that Eq. (2.5) gives non-zero values for negative conductivities. Clearly this is not valid since conductivity cannot be negative. However, the probability density levels at negative conductivities are so small that this is not problematic for this case. For other cases where this can be a problem, the normal distribution is often truncated at non-negative bounds and renormalized so that the area under the truncated distribution is unity. Alternatively, another probability density function could be used which does not give non-zero probabilities at negative conductivities.

Here we chose the normal probability distribution as the example simply because this distribution is the most common one. Other distributions, such as the log normal distribution (see any text on statistics) may be more appropriate for some applications. The choice of which distribution to use is not always obvious and must be chosen with care.



**Figure 2.2. Model for Histogram of Thermal Conductivity**



**Figure 2.3. Estimated Probability Density Function for Conductivity**

### ***2.3 Uncertainty in Two Model Parameters: Linear Model***

In the previous example, we developed a model for the uncertainty for one model parameter. Most numerical models have many model parameters. If these parameters are independent, one can model the uncertainty in each parameter with its own probability distribution. If they are not independent, then we must account for the correlation between the parameters.

We start with a simple linear algebraic model to introduce the effect of two model parameters, the effect of correlation, and the propagation of uncertainty through a model. Let us assume that based on an experiment with a particular material, we have developed the following model for the value of some property  $p$  as a linear function of temperature  $T$ :

$$p(T) = a + bT \quad (2.6)$$

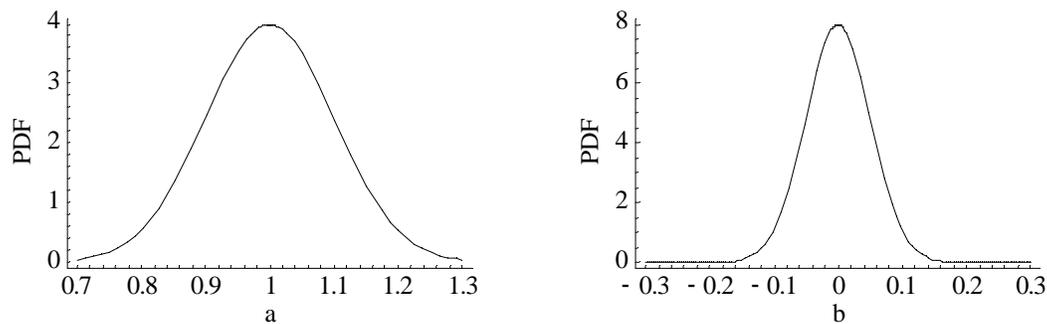
We consider  $p$  to be the model predictive variable,  $T$  to be an independent variable, and  $a$  and  $b$  to be the model parameters. Due to manufacturing tolerances, the property  $p$  of the material will vary from sample to sample. The experimental procedures used to estimate  $a$  and  $b$  will also have uncertainty associated with them. As a result, we cannot expect our model to provide the actual value for  $p$ . The most we can hope for is some prediction of central (i.e., mean, median, etc.) behavior of  $p$  as a function of  $T$ .

First we assume that, based on experience with estimating the parameters  $a$  and  $b$  from multiple samples of the material, the uncertainty in these parameters can be modeled as normally distributed random variables. Also assume that we have estimated the following means and variances for these distributions as shown in Table 2.2. Covariance will be introduced below. The corresponding PDFs for each parameter are shown in Figure 2.4. Note that different scales are used for each plot. Since these are PDFs, the area under each curve must be unity.

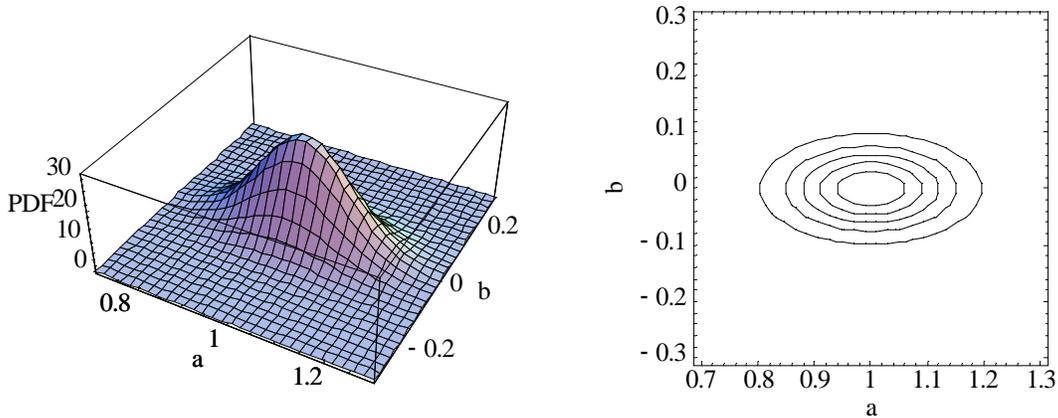
The joint probability is another important concept that is needed for uncertainty analysis. For example, we may be interested in the value of the probability density for particular values of  $a$  and of  $b$  occurring simultaneously. A typical joint probability distribution is shown in Figure 2.5. The volume under the joint PDF must also be unity. Note that the contour plot for this particular joint probability density function results in ellipses with a horizontal major axis.

**Table 2.2. Statistics for the Model Parameters**

Statistic	$a$	$b$
Mean	1	0
Variance	.01	.0025
Uncorrelated Covariance( $a,b$ )	0	
Correlated Covariance( $a,b$ )		.004

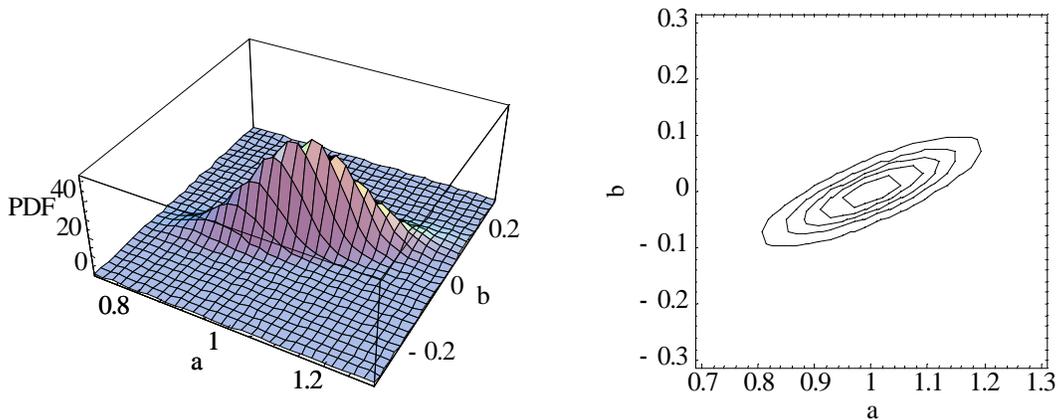


**Figure 2.4. Probability Density Functions of Parameters  $a$  and  $b$ .**



**Figure 2.5. Joint Probability Density Plots for the  $a$  and  $b$ :  $a$  and  $b$  are uncorrelated.**

If the uncertainty in  $a$  and  $b$  are positively correlated (i.e., samples with large values of  $a$  also tend to have large values of  $b$ ), then the probability density plots would take the form similar to that shown in Figure 2.6.



**Figure 2.6. Joint Probability Density Plots for the  $a$  and  $b$ :  $a$  and  $b$  are correlated.**

The ellipses shown in the contour plot of Figure 2.6 possess major and minor axis that are not parallel to the  $a$  and  $b$  axis. This is characteristic of correlated variables. If the correlation between  $a$  and  $b$  is negative (i.e., large values for  $a$  tend to occur with small values of  $b$ ), then the major axis would have a negative slope. In contrast, uncorrelated parameters will have a PDF which possess a major axis that is either horizontal or vertical as shown in Figure 2.5.

It is not unusual for two parameters to be correlated if they represent two parameters in a property model. For example, consider the case for which  $a$  represents a reference conductivity and  $b$  represents the slope of the conductivity with temperature. If the material is such that samples with high reference conductivities also tend to have high

sensitivities with temperature (i.e., large magnitude of the slope),  $a$  and  $b$  will be correlated.

The process by which  $a$  and  $b$  are estimated itself can introduce correlation between the uncertainty in  $a$  and  $b$ . If an experiment is executed multiple times to generate many measurements of  $p$  as a function of  $T$ , and if the  $a$  and  $b$  in Eq. (2.6) are chosen to give the best fit of this equation to the measurements in a least squares sense, the estimates for  $a$  and  $b$  will possess uncertainty which is correlated.

We have introduced the concepts of mean and standard deviation as measures of uncertainty for a single random variable or parameter. These also apply to more than one variable. However, we also need a measure of correlation between variables. One of the more common measures is the covariance which is closely related to the variance. Covariance of  $a$  and  $b$  can be estimated from multiple pairs of  $a_i$  and  $b_i$  sampled from a population as follows:

$$\langle \text{cov}(a, b) \rangle = \frac{1}{n-1} \sum_{i=1}^n (a_i - a_{\text{mean}})(b_i - b_{\text{mean}}) \quad (2.7)$$

where  $n$  is the number of samples. Note that the covariance of a random variable with itself is the variance or the square of the standard deviation of that variable (compare Eq. (2.7) to Eqs. (2.2) and (2.3)). The covariance of two uncorrelated random variables is zero.

If our model requires more than two model parameters, a convenient method to write the variances and covariances is in terms of the covariance matrix. For the parameters  $a$ ,  $b$ , and  $c$ , the covariance matrix takes the form:

$$\mathbf{V} = \begin{bmatrix} \text{cov}(a, a) & \text{cov}(a, b) & \text{cov}(a, c) \\ \text{cov}(b, a) & \text{cov}(b, b) & \text{cov}(b, c) \\ \text{cov}(c, a) & \text{cov}(c, b) & \text{cov}(c, c) \end{bmatrix} \quad (2.8)$$

For the case of Figure 2.6, the covariance matrix is (see Table 2.2)

$$\mathbf{V} = \begin{bmatrix} 0.01 & 0.004 \\ 0.004 & 0.0025 \end{bmatrix} \quad (2.9)$$

Given a covariance matrix, we can show the functional form of the Multi-normal PDF used to generate Figures 2.5 and 2.6. The general form for this two-dimensional PDF is

$$\text{PDF}(a, b) = \frac{1}{2\pi \sqrt{|\mathbf{V}|}} \exp \left( - [a - a_{\text{mean}} \quad b - b_{\text{mean}}] \mathbf{V}^{-1} \begin{bmatrix} a - a_{\text{mean}} \\ b - b_{\text{mean}} \end{bmatrix} \right)$$

where  $\mathbf{V}^{-1}$  denotes the inverse of  $\mathbf{V}$  and  $|\mathbf{V}|$  indicates the determinant of  $\mathbf{V}$ .

## 2.4 Resulting Uncertainty in Model Prediction

Now that we have some idea as to how to characterize uncertainty in model parameters, we are ready to see what effect this uncertainty has on model predictions.

**Question** - Given the model of Eq. (2.6) and the uncertainty in  $a$  and  $b$ , what is the uncertainty in the predicted  $p(T)$ ?

There are several methods by which the uncertainty in  $p(T)$  can be evaluated. We focus on the two most common methods; the Monte Carlo method and the sensitivity analysis method.

### 2.4.1 Monte Carlo Method

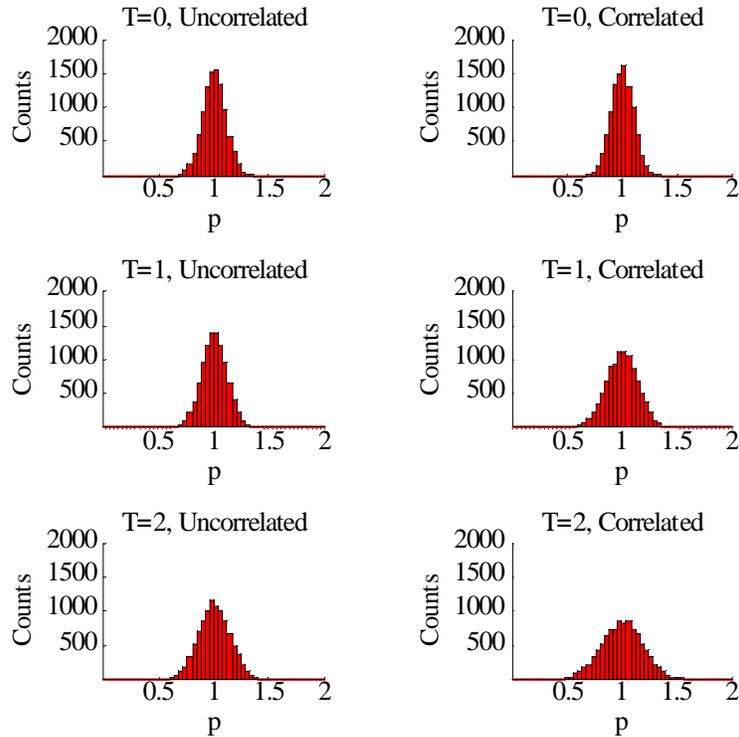
Conceptually, the easiest method to evaluate prediction uncertainty is to simply generate random numbers for  $a$  and  $b$  from the probability density function illustrated in Figures 2.5 and 2.6. Computer algorithms for the generation of random number for various PDFs are commonly available. For example, 10 random numbers for  $a$  and  $b$  generated using the statistics given in Table 2.2 for normal PDFs are shown in Table 2.3.

**Table 2.3. Random Numbers for  $a$  and  $b$**

Not Correlated		Correlated	
a	b	a	b
1.0180	-0.07091	1.0813	0.04698
1.0054	-0.12924	0.8889	-0.02649
1.1906	0.07094	0.9366	0.01859
1.1215	-0.04411	0.9347	-0.06073
0.9559	-0.03382	1.0399	0.06674
0.9633	0.01498	1.0247	-0.05535
1.0268	-0.05891	0.8890	-0.05640
1.1866	0.11282	0.8711	-0.01384
1.1441	0.06003	1.0348	-0.07123
0.8399	-0.02804	1.0151	0.06108

By repeated sampling of  $(a, b)$  pairs, and using these in Eq. (2.6), a corresponding population of predicted  $p$ 's can be obtained for different values of  $T$ . Figure 2.7 shows the histograms for  $p$  obtained for  $T=0, 1, 2$  for both the uncorrelated and correlated model parameters. The present analysis used 10,000 sample pairs for each plot. Note that as the temperature  $T$  increases, the product  $T$  times  $b$  in Eq. (2.6) becomes larger. This increases the sensitivity of the predicted  $p$  to  $b$ , which is reflected by broader distributions in Figure

2.7 at larger  $T$ . Also note that the effect of correlation between  $a$  and  $b$  is to also increase the uncertainty in  $p(T)$ . This is due to the fact that large  $a$ 's tend to go with large  $b$ 's, and small  $a$ 's tend to go with small  $b$ 's. The net result is that the variability in  $p(T)$  is increased due to this correlation.

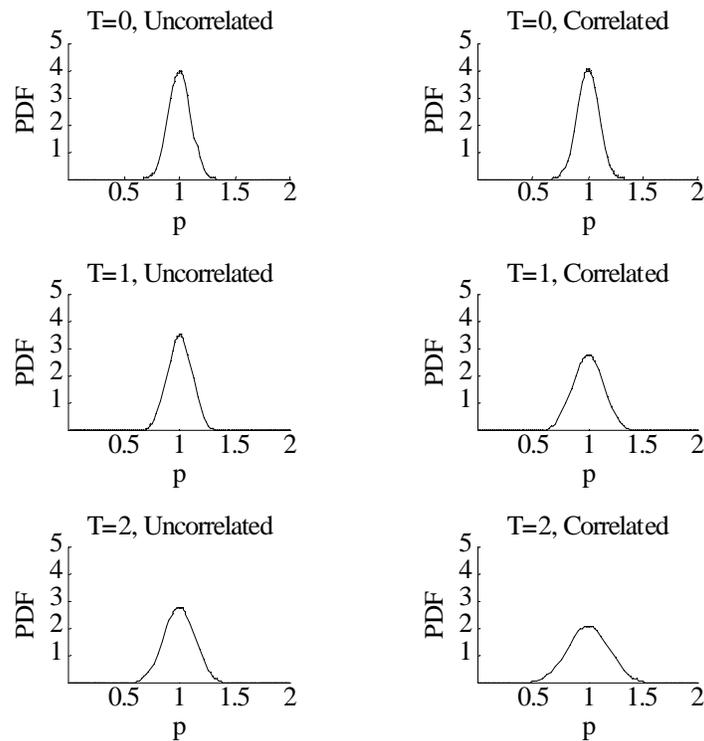


**Figure 2.7. Histograms for Sample Counts of  $p(T)$**

We can use the results of Figure 2.7 to estimate the equivalent probability distributions for  $p(T)$ . The appearance of the histograms suggests that normal probability distributions may be good approximations for each value of  $T$ . As discussed in a previous section, we can use estimates of the mean and the standard deviation of the predicted population for  $p(T)$  to obtain the corresponding estimated PDF for each value of  $T$ . The resulting estimates for mean and standard deviation are shown in Table 2.4 and the resulting PDFs are shown in Figure 2.8 for several values of  $T$ . Equations (2.1) and (2.2) were used to estimate the means and standard deviations.

**Table 2.4. Estimated  $p_{mean}$  and  $s$**

$T$	Not Correlated		Correlated	
	$p_{mean}$	$s$	$p_{mean}$	$s$
0	0.9994	0.1002	1.0001	0.0988
1	0.9987	0.1124	0.9994	0.1416
2	0.9979	0.1423	0.9988	0.1876



**Figure 2.8. Probability Density Plots for Normal Distributions Defined in Table 2.4.**

Note that the agreement between the shapes of the histogram plots of Figure 2.7 and the probability plots of Figure 2.8 is quite good. These plots show how the probability density function of the predicted quantity varies as a function of  $T$  for uncorrelated and correlated model parameters  $a$  and  $b$ . The results indicate that while the mean value of  $p$  does not vary much from 1.0, the width of the PDF function, and hence the uncertainty, increases with  $T$  and is larger for the correlated model parameters.

## 2.4.2 Direct Evaluation for a Linear Model

The mean and standard deviation of the population of possible predictions can be evaluated directly from the means and covariance matrix of the model parameters if the model is linear in the parameters. For example, consider

$$z = c_1x + c_2y \quad (2.10)$$

where  $x$  and  $y$  are random variables, which cause the model prediction  $z$  to be a random variable. The expected value and variance can be evaluated directly, and exactly, as follows (Beck and Arnold, 1977):

$$z_{mean} = c_1x_{mean} + c_2y_{mean} \quad (2.11)$$

$$\text{var}(z) = c_1^2 \text{var}(x) + c_2^2 \text{var}(y) + 2c_1c_2 \text{cov}(x, y) \quad (2.12)$$

For our case, we have

$$p_{mean}(T) = a_{mean} + b_{mean}T \quad (2.13)$$

$$\text{var}(p) = \text{var}(a) + T^2 \text{var}(b) + 2T \text{cov}(a, b) \quad (2.14)$$

Using the known values for mean, standard deviation, and covariance for our random variables  $a$  and  $b$  (see Table 2.2), we find

$$p_{mean}(T) = 1 \quad (2.15)$$

$$\mathbf{s}^2 = \text{var}(p) = 0.01 + 0.0025T^2, \quad a \text{ and } b \text{ not correlated} \quad (2.16)$$

$$\mathbf{s}^2 = \text{var}(p) = 0.01 + 0.008T + 0.0025T^2, \quad a \text{ and } b \text{ correlated} \quad (2.17)$$

The results for three temperatures are shown in Table 2.5.

**Table 2.5. Direct Evaluation of  $p_{mean}$  and  $\mathbf{s}$**

$T$	Not Correlated		Correlated	
	$p_{mean}$	$\mathbf{s}$	$p_{mean}$	$\mathbf{s}$
0	1	0.1	1	0.1
1	1	0.1118	1	0.1432
2	1	0.1414	1	0.1897

Comparison of the results obtained by direct calculation in Table 2.5 to those obtained using the Monte Carlo method in Table 2.4 shows that the differences are small. These differences are due to the fact that we are estimating the population statistics from a finite number of samples when using the Monte Carlo method. Additional Monte Carlo simulations are likely to provide improved estimates of these quantities.

### 2.4.3 Sensitivity Analysis

Another technique to estimate prediction uncertainty is to evaluate the sensitivities of the model predictions to the model parameters through a truncated Taylor's series, and use these sensitivities to estimate the corresponding prediction variance. Consider the general linear or nonlinear model of the form

$$z = f(a, b, x) \quad (2.18)$$

where  $a$  and  $b$  are model parameters and  $x$  is an independent variable (or vector of independent variables). We can approximate the change in  $z$  from its mean value due to changes in the model parameters using a truncated Taylor's series expansion:

$$z_i - z_{mean} = \Delta z_i \cong \frac{\partial f}{\partial a} \Delta a_i + \frac{\partial f}{\partial b} \Delta b_i \quad (2.19)$$

or

$$\Delta z_i = \mathbf{c}^T \Delta \mathbf{d}_i \quad (2.20)$$

where

$$\mathbf{c}^T = \left[ \frac{\partial f}{\partial a} \quad \frac{\partial f}{\partial b} \right] \quad (2.21)$$

$$\Delta \mathbf{d}_i^T = [a_i - a_{mean} \quad b_i - b_{mean}] = [\Delta a_i \quad \Delta b_i]^T \quad (2.22)$$

The subscript  $i$  indicates that the  $z$  is for the  $i^{\text{th}}$  sample of  $a$  and  $b$ . The derivatives that appear in Eq. (2.21) are known as sensitivity coefficients. They are used to relate small changes in  $f$  to small changes in the parameters  $a$  and  $b$ . If we have  $n$  samples, we can create the new matrix system of responses as follows:

$$\Delta \mathbf{z}^T = \mathbf{c}^T \Delta \mathbf{d} \quad (2.23)$$

where

$$\Delta \mathbf{z}^T = [\Delta z_1, \Delta z_2, \dots, \Delta z_n] \quad (2.24)$$

$$\Delta \mathbf{d} = \begin{bmatrix} \Delta a_1 & \Delta a_2 & \dots & \Delta a_n \\ \Delta b_1 & \Delta b_2 & \dots & \Delta b_n \end{bmatrix} \quad (2.25)$$

Because  $a$  and  $b$  may have different units, scaled sensitivity coefficients are often used. These are given by

$$\mathbf{c}_s^T = \left[ a_r \frac{\partial f}{\partial a} \quad b_r \frac{\partial f}{\partial b} \right]$$

where the  $r$  subscripts denote reference values for the parameters  $a$  and  $b$ . We can now estimate the variance of  $z$  using Eq. (2.23) as follows:

$$\begin{aligned}
\text{var}(z) &= \frac{1}{n-1} \sum_{i=1}^n \Delta z_i^2 = \frac{1}{n-1} \Delta \mathbf{z} \Delta \mathbf{z}^T \\
&= \frac{1}{n-1} \mathbf{c}^T \Delta \mathbf{d} \Delta \mathbf{d}^T \mathbf{c} \\
&= \mathbf{c}^T \begin{bmatrix} \frac{1}{n-1} \sum_{i=1}^n \Delta a_i^2 & \frac{1}{n-1} \sum_{i=1}^n \Delta a_i \Delta b_i \\ \frac{1}{n-1} \sum_{i=1}^n \Delta a_i \Delta b_i & \frac{1}{n-1} \sum_{i=1}^n \Delta b_i^2 \end{bmatrix} \mathbf{c}
\end{aligned}$$

The term in brackets is simply our estimate of the covariance matrix of the model parameters (see Eq. (2.7) and (2.8)). Finally, we can write

$$\langle \text{var}(z) \rangle = \mathbf{c}^T \langle \mathbf{V} \rangle \mathbf{c} \quad (2.26)$$

where we denoted  $\text{var}(z)$  as an estimate since  $\mathbf{V}$  is an estimate. For our present model

$$p = [1 \quad T] \begin{bmatrix} a \\ b \end{bmatrix} \quad (2.27)$$

So

$$\begin{aligned}
\text{var}(p) &= [1 \quad T] V(a, b) \begin{bmatrix} 1 \\ T \end{bmatrix} \\
&= \text{var}(a) + 2T \text{cov}(a, b) + T^2 \text{var}(b)
\end{aligned} \quad (2.28)$$

This is exactly the result obtained in Eq. (2.14) through direct calculation. Since this model is linear, the truncated Taylor's series expansion of Eq. (2.19) is exact, leading to exact results for the variance of  $p$ . The advantage of using Eqs. (2.21) and (2.26) is this analysis allows the variance of predicted variables to be estimated for nonlinear models. However, this method is good only as long as the truncated Taylor series expansion, Eq. (2.19), provides a good approximation. The extension of this method to highly nonlinear models should be done with care.

The application of the sensitivity method provides estimates for the variance (or standard deviation) of the prediction population due to uncertainty in the model parameters. This method does not provide the shape of the resulting prediction PDF. We show the effects of this in the next section.

## 2.5 Simple Nonlinear Model

The differential equation for a simple damped spring mass system under the influence of a harmonic forcing function is give by

$$m \frac{d^2x}{dt^2} + c \frac{dx}{dt} + kx = F \sin(\omega t) \quad (2.29)$$

where  $x$  is displacement,  $t$  is time,  $F$  and  $\omega$  are the magnitude and circular frequency of the forcing function,  $m$  is mass,  $c$  is the damping coefficient, and  $k$  is the spring constant. The amplitude of the resulting oscillation for large time is given by

$$X = \frac{F}{\sqrt{(k - m\omega^2)^2 + (c\omega)^2}} \quad (2.30)$$

In this example, we assume that  $\omega$  is a known variable, and that the model parameters  $m$ ,  $k$ , and  $c$  contain uncertainty. We also assume that  $m$ ,  $k$ , and  $c$  have been measured through independent experiments so that their uncertainty is uncorrelated.

Note that even through Eq. (2.30) results from a model, Eq. (2.29), which is linear in the independent variable  $t$ , the amplitude of the oscillation is a nonlinear function of the model parameters  $m$ ,  $k$ , and  $c$ . This non-linearity in the parameters is a common occurrence when dealing with differential equations. Assume that the uncertainty in  $m$ ,  $k$ , and  $c$  are normally distributed with the statistics given in Table 2.6.

**Table 2.6. Statistics for the Oscillator Model Parameters: The parameters are normally distributed.**

Parameter	Mean	Standard Deviation	Standard Deviation / Mean
m	1.00	0.001	0.001
k	2.00	0.05	0.025
c	0.25	0.05	0.200

Note that the relative uncertainty (last column in Table 2.6) in the mass is very small whereas the relative uncertainty in the damping coefficient is large. We can easily measure mass with high precision. Of the three model parameters shown in Table 2.6, we assumed that the damping coefficient is the most difficult to measure with accuracy.

Since the model for amplitude is not linear in the parameters, the direct calculation method used for the linear model does not apply. Here we apply the sensitivity method and the Monte Carlo method to the prediction of the ratio of amplitude to the magnitude of the forcing function.

$$\frac{X}{F} = f(m, k, c, \omega) = \frac{1}{\sqrt{(k - m\omega^2)^2 + (c\omega)^2}} \quad (2.31)$$

We start with the sensitivity analysis. Note that

$$\Delta(X / F) \cong \frac{\mathcal{J}f}{\mathcal{J}m} \Delta m + \frac{\mathcal{J}f}{\mathcal{J}k} \Delta k + \frac{\mathcal{J}f}{\mathcal{J}c} \Delta c = \mathbf{c}^T \mathbf{d} \quad (2.32)$$

where

$$\mathbf{c}^T = \begin{bmatrix} \mathcal{J}f \\ \mathcal{J}m \\ \mathcal{J}k \\ \mathcal{J}c \end{bmatrix} \quad (2.33)$$

$$\mathbf{d} = \begin{bmatrix} m - m_{mean} \\ k - k_{mean} \\ c - c_{mean} \end{bmatrix} = \begin{bmatrix} \Delta m \\ \Delta k \\ \Delta c \end{bmatrix} \quad (2.34)$$

For our case, we have

$$\frac{\mathcal{J}f}{\mathcal{J}m} = \frac{(k - m\omega^2)\omega^2}{\sqrt[3]{(k - m\omega^2)^2 + (c\omega)^2}} \quad (2.35)$$

$$\frac{\mathcal{J}f}{\mathcal{J}k} = \frac{-(k - m\omega^2)}{\sqrt[3]{(k - m\omega^2)^2 + (c\omega)^2}} \quad (2.36)$$

$$\frac{\mathcal{J}f}{\mathcal{J}c} = \frac{-c\omega^2}{\sqrt[3]{(k - m\omega^2)^2 + (c\omega)^2}} \quad (2.37)$$

From Eq. (2.26) and Table 2.6, we can write

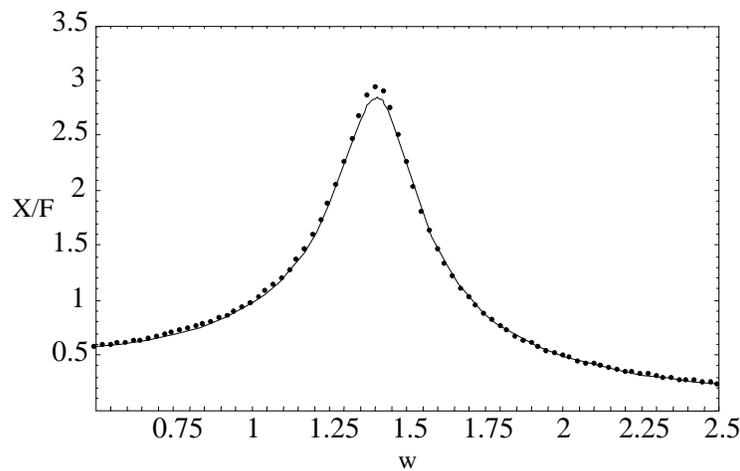
$$\text{var}(X / F) = \mathbf{c}^T \begin{bmatrix} 0.001^2 & 0 & 0 \\ 0 & 0.05^2 & 0 \\ 0 & 0 & 0.05^2 \end{bmatrix} \mathbf{c} \quad (2.38)$$

The off-diagonal terms are zero since the parameters are not correlated. Note that the terms in the vector  $\mathbf{c}$  are nonlinear functions of the frequency  $\omega$ .

The Monte Carlo analysis is performed simply by generating normally distributed, uncorrelated, random values for  $m$ ,  $k$ , and  $c$ , for the statistics given in Table 2.6; using these values in Eq. (2.31); and utilizing the results for statistical analysis. Equation (2.31) was evaluated as a function of  $\omega$  for each set of the random numbers for  $m$ ,  $k$ , and  $c$ . In this example, we used 10000 random sets (i.e. triples) of the model parameters. While we can obtain good results using less than 10000 samples, the CPU requirements to generate 10000 samples for this model were minimal. The solid curve of Figure 2.9 shows the predicted ratio  $X/F$  (see Eq. 2.31) as a function of frequency  $\omega$  using the mean values for the input parameters. The points represent the means of the predicted  $X/F$  from the Monte

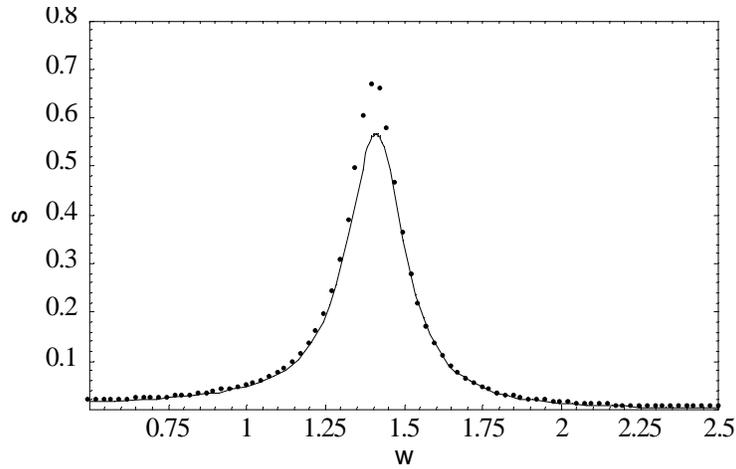
Carlo simulations. Note that there are some discrepancies between the two but these discrepancies are small.

Figure 2.10 illustrates the estimated standard deviations (square root of the variances) for the uncertainty in the prediction of  $X/F$  for the two methods. Note that there are significant differences in the standard deviations near resonance ( $w=1.5$ ). The Monte Carlo method predicts that the standard deviation of the predictions is significantly more (about 20% more) than that predicted by the sensitivity method near the resonance peak. The Monte Carlo method is generally considered to be one of the most dependable methods (assuming a sufficient number of samples are taken) for the analysis of nonlinear problems. Assuming that this is the case here, the nonlinear effect associated with the resonance peak is apparently too large to be well modeled by the sensitivity analysis.



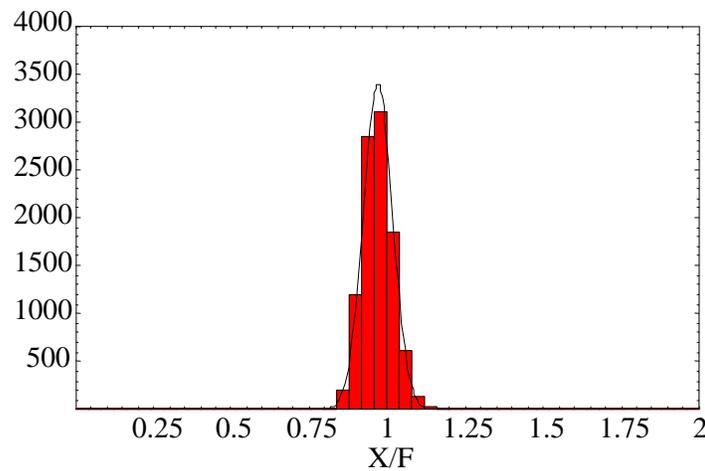
**Figure 2.9. Predicted  $X/F$ : Solid line based on means of model parameters; points are based on means of Monte Carlo predictions.**

We can investigate this nonlinear effect further by evaluating histograms of the Monte Carlo predictions for the ratio  $X/F$ . For reference, we compare these histograms to scaled normal probability density functions obtained from the sensitivity analysis. The predicted means used to estimate an approximate scaled normally distributed PDF were obtained using the mean values of the parameters in the predictive model (Eq. 2.31). The corresponding variances were obtained from the sensitivity analysis (Eq. 2.38). These results are shown in Figures 2.11 and 2.12 for two different frequencies. The resulting probability density functions have been scaled such that the area under the scaled density functions are equal to the corresponding areas under the histograms. The results shown in Figure 2.11 illustrate that the scaled normal PDFs do a reasonable job of modeling the characteristics of the Monte Carlo population of predicted  $X/F$  for  $w=1$ . In contrast, Figure 2.12 shows less agreement for frequencies closer to resonance. The Monte Carlo

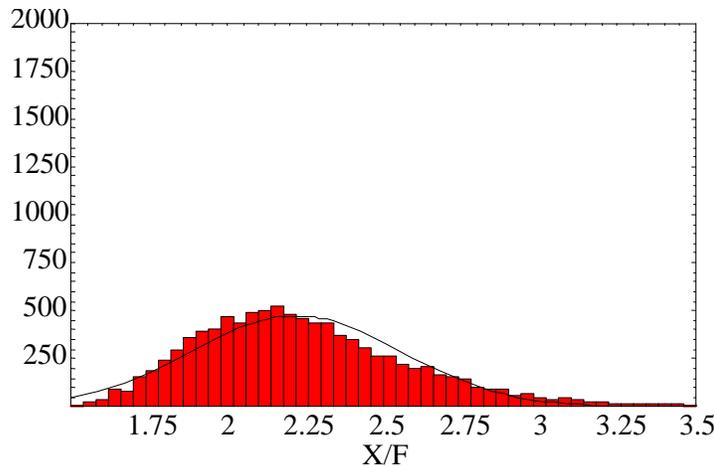


**Figure 2.10. Predicted Standard Deviation for  $X/F$ : Sensitivity analysis results are represented by the solid line; Monte Carlo results are represented by the points.**

population shows a skewness that cannot be modeled by a normal distribution. These results indicate that even though the model input parameters are distributed with normal PDFs, the resulting prediction uncertainty are not necessarily well modeled by a normal PDFs. This observation is especially true for highly nonlinear models.



**Figure 2.11. Predicted Populations for  $X/F$ :  $w=1.0$ ; Monte Carlo predictions are shown by the histogram; scaled Normal PDF based on sensitivity analysis statistics are shown by the solid curve.**

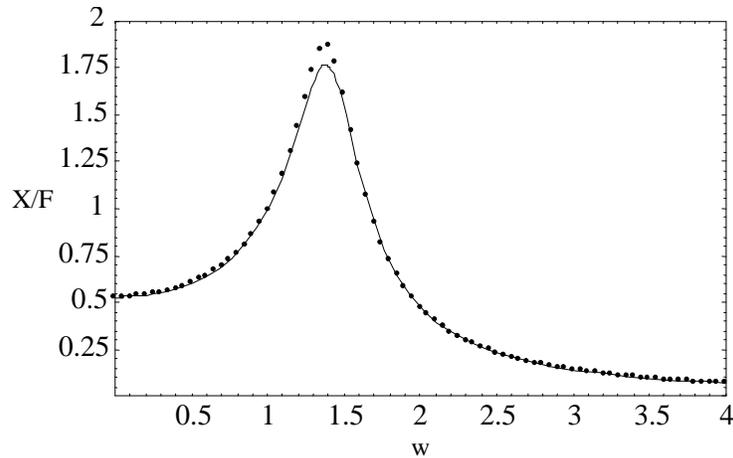


**Figure 2.12. Predicted Populations of  $X/F$ :  $w=1.5$ ; Monte Carlo predictions are shown by the histogram; scaled Normal PDF based on sensitivity analysis statistics are shown by the solid curve.**

We can easily use the resulting set of Monte Carlo predictions to evaluate confidence levels for the predicted ratio  $X/F$ . For example, let's say we are interested in bounding the maximum amplitude of the oscillation for some  $w$  at the 99% confidence level. This amplitude can be estimated from the Monte Carlo simulations for  $X/F$  simply by sorting the resulting 10000 Monte Carlo predictions from small to large and taking the prediction at the 99% location through the list ( $0.99 \cdot 10000 = 9900^{\text{th}}$  prediction in the sorted list). Figure 2.13 shows a comparison of the 99% confidence level using the Monte Carlo method and that obtained from the means and standard deviations estimated from the sensitivity analysis, assuming a normal distribution. For such a distribution, the one sided 99% confidence level is  $2.326 \sigma$  from the mean. Note that the 99% confidence levels are well approximated by the sensitivity analysis statistics except near resonance. At resonance, the sensitivity analysis under-predicts the 99% confidence maximum  $X/F$  by approximately 7%. This discrepancy is due to both the incorrect estimation of the prediction standard deviation and due to the fact that the uncertainty of the predictions are not normally distributed near resonance as was assumed when using the sensitivity analysis based statistics.

Even though the model for the spring/mass/damper system is quite simple, this model demonstrates several important features of the propagation of uncertainty analysis:

1. A model which is linear in the dependent variable may not be linear in the model parameters.



**Figure 2.13. Predicted Maximum  $X/F$  at 99% Confidence Level: Monte Carlo results represented by points; sensitivity analysis results represented by the solid curve.**

2. Model parameter uncertainty which can be modeled by normal probability distributions do not necessarily lead to prediction uncertainty which can be well modeled by a normal probability distribution for models nonlinear in these input parameters.
3. The differences in the predictions of mean behavior, using the mean of the input model parameters and using the average of the Monte Carlo predictions, were very small for the examples presented. This is not always the case.
4. The differences between the Monte Carlo based results and the results using the sensitivity analysis statistics for both the predicted standard deviations and 99% confidence level were larger near resonance for the large uncertainty associated with the input parameters (i.e., the damping coefficient). While gradient based methods, such as the sensitivity method, can do a good job at predicting mean behavior of a system, such methods can give poor results away from the mean for models which are very nonlinear in the parameters. For cases for which the input parameter uncertainty is not as large, such as would be the case for carefully controlled validation experiments, the sensitivity-based methods can give good results, even for problems that are very nonlinear.

## 2.6 Diffusion Equation

In our next example, we consider the one-dimensional transient diffusion equation

$$\frac{\partial T}{\partial t} = a \frac{\partial^2 T}{\partial x^2} \quad (2.39)$$

with the initial and boundary conditions

$$T(x,0) = 0 \quad (2.40)$$

$$T(0,t) = 1 \quad (2.41)$$

$$T(1,t) = 0 \quad (2.42)$$

Here we consider two prediction quantities. The first is the centerline temperature,  $T(0.5,t)$  as a function of time. The second is the location of the  $T = 0.5$  isotherm as a function of time. The thermal diffusivity  $\mathbf{a}$  is the only model parameter with uncertainty and is assumed to be well modeled by a normally distributed random variable with mean of 1 and a standard deviation of 0.1. This standard deviation represents a significant level of uncertainty.

The analytical solution to the above system of equations can be easily derived in terms of a Fourier series. However, we chose to solve the problem numerically to show some of the difficulties caused by the approximate nature of a numerical method. We used an explicit finite difference approximation with  $\Delta x = 0.05$  and  $\Delta t = 0.00025$ . All calculations were performed in double precision (Real\*8). The Monte Carlo simulations used 5000 samples of the model parameter  $\mathbf{a}$ .

The sensitivity method requires that we know the derivative of the prediction variable with respect to the random input variable. Two approaches have been traditionally used to estimate this derivative. The first is to develop a partial differential equation governing the sensitivity (i.e., derivative) of the prediction variable with respect to the input parameter. For prediction variables that are the same as the dependent variables, such as the temperature for this case, this method results in differential equations for the sensitivity that can be similar in form to the original differential equation. This often allows the resulting sensitivity differential equations to be solved using the same numerical algorithm that was used for the original differential equation. Since the sensitivity equations are linear, they can often be solved faster than the original equation. When the prediction variable is different from the dependent variable, such as the location of the  $T = 0.5$  isotherm, the development of the resulting differential equations for sensitivity is more problematic.

The second method is simply to solve the problem using the standard numerical algorithm with the input parameter perturbed from the mean value so that a finite difference approximation for the sensitivity derivatives can be developed. This approach is used here. We use central finite differences to approximate the derivatives;

$$\frac{dp(t, \mathbf{a}_{mean})}{d\mathbf{a}} \cong \frac{p(t, \mathbf{a}_{mean} + \Delta\mathbf{a}) - p(t, \mathbf{a}_{mean} - \Delta\mathbf{a})}{2\Delta\mathbf{a}} \quad (2.43)$$

where  $p$  is the predicted quantity (center line temperature or the location of the 0.5 isotherm) and  $\Delta\mathbf{a}$  is a perturbation from the mean for the parameter  $\alpha$ . Here we use  $\Delta\alpha =$

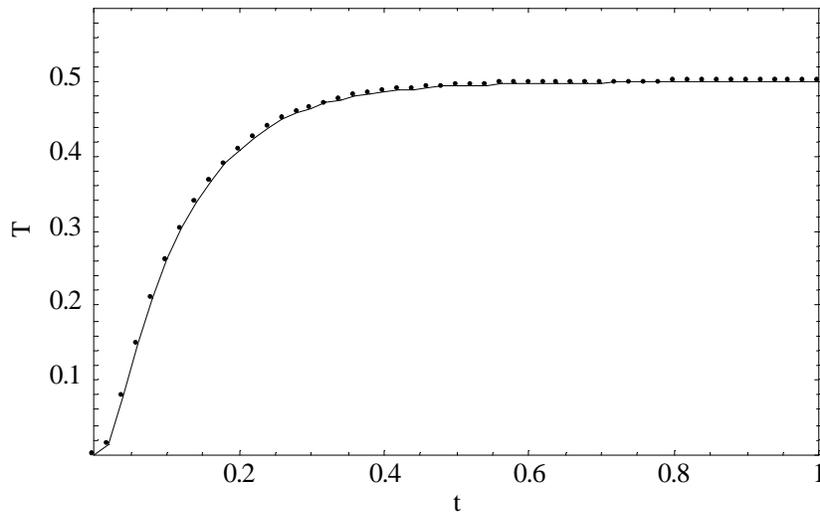
0.0002. Note that since the prediction quantity is a function of time, the sensitivity of this quantity with respect to the input variable will also be a function of time. For the present case of one input variable, Eq. (2.26) gives

$$\text{var}(p(t)) \cong \left[ \frac{p(t, \mathbf{a}_{mean} + \Delta \mathbf{a}) - p(t, \mathbf{a}_{mean} - \Delta \mathbf{a})}{2\Delta \mathbf{a}} \right]^2 \text{var}(\mathbf{a}) \quad (2.44)$$

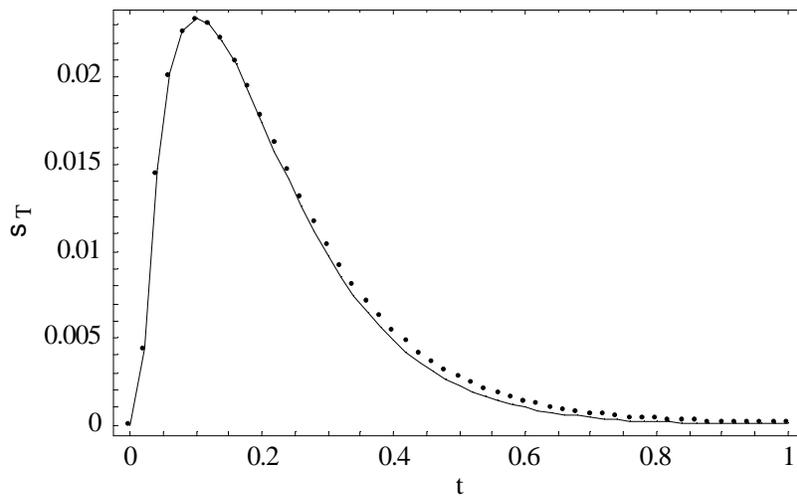
As before, we apply the sensitivity method and the Monte Carlo method to investigate the propagation of uncertainty.

The resulting estimates for the mean behavior of the centerline temperature and the standard deviation of the predicted centerline temperature as a function of time are shown in Figures 2.14 and 2.15. We also show Monte Carlo generated histograms and the scaled (scaled such that the area under the histograms and sensitivity plots are the same) normal PDFs using the sensitivity analysis statistics (as discussed earlier) for several times in Figure 2.16. Note that the standard deviation is a strong function of time and reaches a maximum near  $t = 0.1$ . Also note that the centerline temperature is bounded by the initial condition value ( $T = 0$ ) and the steady state value ( $T = 0.5$ ). Because of this, the prediction population will necessarily be skew at the initial time and for large times. Careful inspection of histograms also indicates that there is a slight skewing in the predicted population for intermediate times. Overall, the results presented in Figures 2.14 through 2.16 show reasonable agreement between the results from the Monte Carlo method and the results based on the sensitivity analysis.

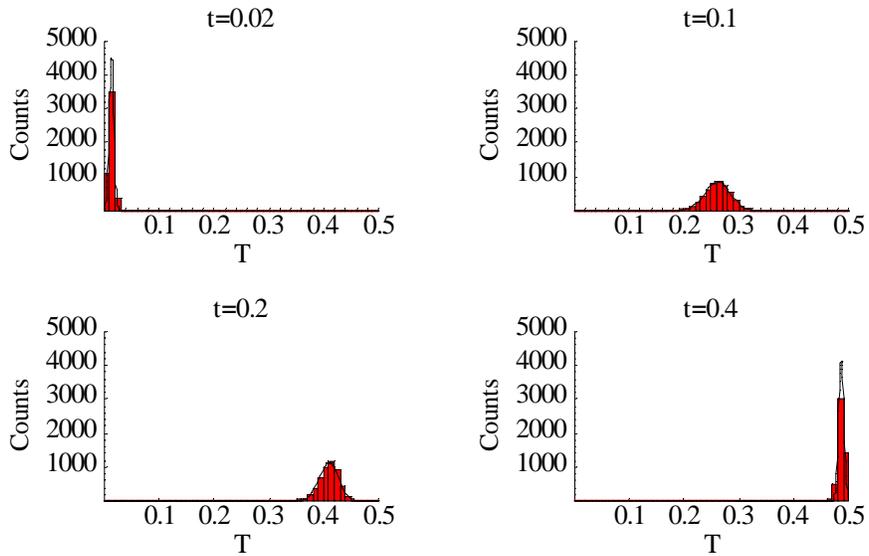
Figures 2.17 through 2.19 illustrate the results obtained for the predicted location  $x$  of the  $T = 0.5$  isotherm as a function of time. Figure 2.18 indicates that the uncertainty in the 0.5 isotherm location reaches a maximum at  $t = 0.15$ , which is later than the peak time for the centerline temperature. At this point, the standard deviation is approximately 0.017. Other than the magnitudes of uncertainty, the behavior of the uncertainty over time for the centerline temperature and for the location of the isotherm is remarkably similar. The sensitivity analysis results shown in Figure 2.18 exhibit some erratic behavior around  $t = 0.15$  which was not apparent for the centerline temperature results shown in Figure 2.15. This illustrates one of the problems associated with using finite differences to approximate the sensitivity derivatives (Eq. (2.44)). The procedure used here to estimate the location of the  $T = 0.5$  isotherm was to first find the temperatures and location of the two finite difference nodes on either side of this isotherm, then use linear interpolation to find the intermediate value for  $x$ . This is a common procedure and is inherently noisy. When finite differences are taken using small changes in the model parameter  $\mathbf{a}$ , the effect of this noise is amplified. Larger changes in the model parameter  $\mathbf{a}$  can be used to reduce the sensitivity to noise, but this leads to less accurate estimates of the sensitivity derivative. The Monte Carlo method does not suffer from this problem since a sensitivity derivative is not required.



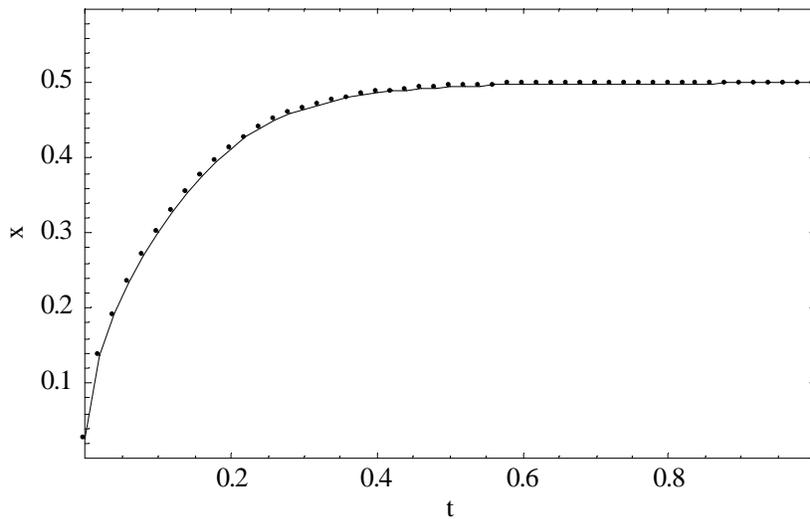
**Figure 2.14. Predicted Mean Centerline Temperature: Monte Carlo results represented by points; sensitivity analysis results represented by the solid curve.**



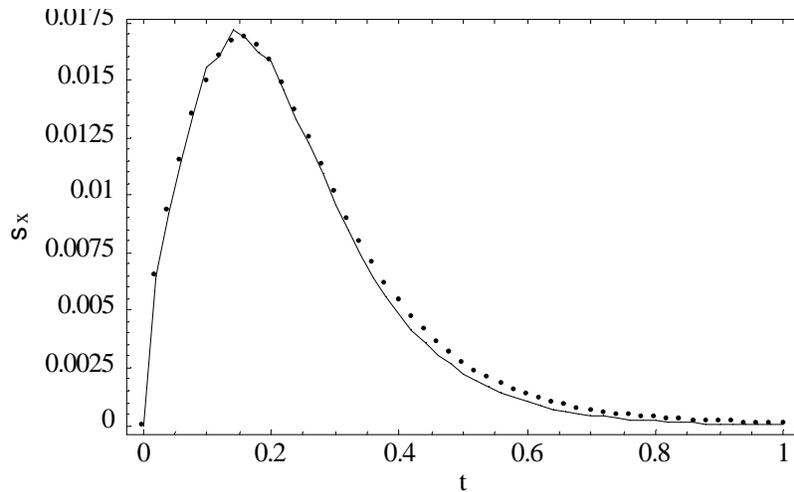
**Figure 2.15. Predicted Standard Deviation for Centerline Temperature: Monte Carlo results represented by points; sensitivity analysis results represented by the solid curve.**



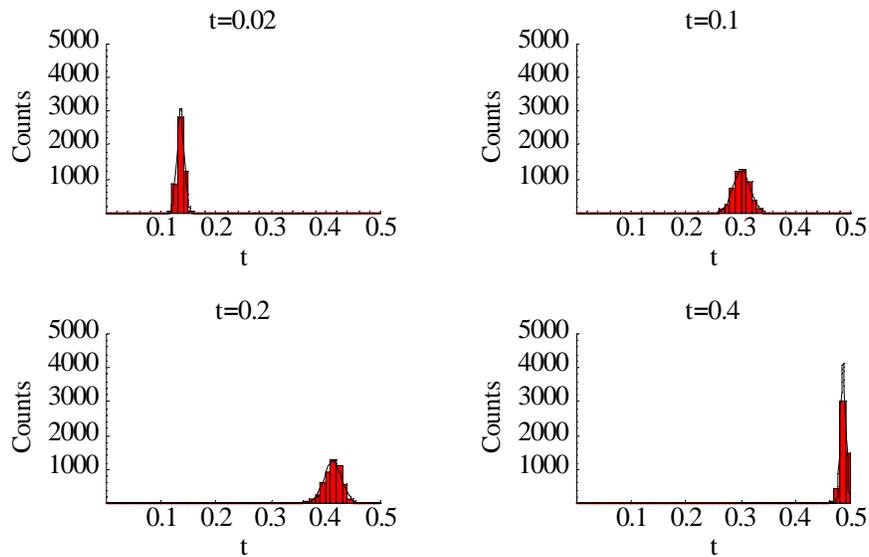
**Figure 2.16. Predicted Populations for Centerline Temperature: Monte Carlo results represented by histograms; scaled sensitivity analysis results represented by solid curve.**



**Figure 2.17. Predicted Mean  $T = 0.5$  Isotherm Location: Monte Carlo results represented by points; sensitivity analysis results represented by the solid curve.**



**Figure 2.18. Predicted Standard Deviation for  $T = 0.5$  Isotherm Location: Monte Carlo results represented by points; scaled sensitivity analysis results represented by solid curve.**



**Figure 2.19. Predicted Populations for  $T = 0.5$  Isotherm Location: Monte Carlo predictions -represented by histograms; scaled sensitivity analysis represented by solid curve.**

Overall, this example problem shows that the propagation of uncertainty analysis based on the sensitivity method works quite well. The uncertainty is a strong function of time. Since the initial temperature distribution and the steady state temperature distribution are not functions of thermal diffusivity, these quantities are known exactly. The standard

deviation of the centerline temperature and of the location of the 0.5 isotherm thus vary from zero at  $t = 0$  to some maximum value, and then decrease to zero as the solution reaches steady state. Plotting the sensitivity coefficients as a function of time would show a similar behavior. The sensitivity coefficients will be zero at time zero and at steady-state. However, if the initial and boundary conditions contain uncertain parameters, the time zero and steady state results will have non-zero standard deviations and non-zero sensitivity coefficients.

## 2.7 Hyperbolic Equation

Our last example for the uncertainty analysis is based on a modified form of Burgers' equation. This equation is nonlinear and can lead to expansion and compression waves. The form we use here is

$$\frac{\partial u}{\partial t} + \frac{\partial (cu^2)}{\partial x} + a \frac{\partial^2 u}{\partial x^2} = 0 \quad (2.45)$$

where  $c$  and  $a$  are parameters. We take  $c$  and  $a$  to be spatially distributed random variables (i.e., both are functions of  $x$ ). We assume that  $c$  and  $a$  are independent of each other but assume that there is spatial correlation among the various spatial values of  $a$  and among the various spatial values of  $c$ .

Equation (2.45) is solved numerically using an operator splitting technique. This equation was split into a strictly advective equation and a diffusive equation for each time step (Hills et al., 1994a). The resulting advective equation is solved using the second order TVD scheme of Roe and Sweby combined with a Superbee limiter (Roe, 1985, 1986 and Sweby, 1984). Comparison of this method with other shock-capturing methods, as applied to Burger's inviscid equation, is presented by Yang and Przekwas (1992). The diffusive equation was solved using simple explicit finite differences. The following initial conditions are used:

$$u(x,0) = \begin{cases} 0.5, & 0 \leq x < 1 \\ x - 0.5, & 1 \leq x < 2 \\ 1.5, & 2 \leq x < 3 \\ 4.5 - x, & 3 \leq x < 4 \\ 0.5, & 4 \leq x < 10 \end{cases} \quad (2.46)$$

The computational domain is  $0 \leq x \leq 10$ , with a  $\Delta x = 0.1$  (i.e., 100 cells) and  $\Delta t = 0.002$ . Cyclic boundary conditions are applied. The initial conditions result in a pulse, with a maximum initial velocity of 1.5, being advected in the positive direction with compression occurring at the leading edge and expansion occurring on the trailing edge.

Some diffusion of the pulse takes place due to the last term in Eq. (2.45). We define the prediction variable to be the location of the leading edge of the pulse. More specifically, we define this location to be the  $x$  location where the value of  $u$  drops to half the difference between the maximum and minimum value of  $u$  over the domain at the time of interest.

The model parameter statistics used here are listed in Table 2.7. Normal distributions were assumed. Note that we choose high values for the relative uncertainty for purposes of demonstration.

**Table 2.7. Statistics for the Burger Model Parameters**

Parameter	Mean	Standard Deviation	Standard Deviation / Mean
$c$	0.4	0.04	0.1
$a$	0.1	0.01	0.1

As mentioned earlier, we assume that the spatial variation in each model parameter is spatially correlated. If the value for the parameter is high at one location, then it is more likely to be high at adjacent locations. Spatial correlation is common in naturally occurring porous media. This type of model can also be appropriate for materials with complex chemical compositions that change due to aging effects in the presence of temperature gradients. There are many models for spatial correlation and the choice of the model depends on the application (much like the choice of the probability density function depends on the application). Here we use one of the more common models, that of exponential correlation. This model is given by

$$\text{cov}(c(x) c(x + \Delta)) = \mathbf{s}_c^2 \exp(-3\Delta / l) \quad (2.47)$$

$$\text{cov}(a(x) a(x + \Delta)) = \mathbf{s}_a^2 \exp(-3\Delta / l) \quad (2.48)$$

where  $l$  is a correlation length parameter and  $\Delta$  is the distance between the two points of interest. Note that we assume the same spatial correlation structure for each parameter. As the correlation length  $l$  increases, the correlation between neighboring values for the parameters increases. Thus neighbors of a location where  $c$  is large are more likely to have large values of  $c$ .

Since  $c$  and  $a$  are spatially dependent and uncertain, the value of  $c$  and  $a$  at each finite difference node is considered to be a random variable with the statistics given in Table 2.7. The sensitivity analysis requires we know the sensitivity of the prediction variable to each of these nodal values for  $c$  and  $a$ . Since we are modeling the computational domain with a 100 cells, each possessing their own value for  $c$  and  $a$ , we effectively have 200 parameters. We must evaluate the sensitivity of the prediction variable to each of these 200 parameters. The resulting sensitivity vectors are given by

$$\mathbf{X}_c(t)^T = \begin{bmatrix} \mathcal{J}p(t) & \mathcal{J}p(t) & \dots & \mathcal{J}p(t) \\ \mathcal{J}c_1 & \mathcal{J}c_2 & \dots & \mathcal{J}c_{100} \end{bmatrix} \quad (2.49)$$

$$\mathbf{X}_a(t)^T = \begin{bmatrix} \mathcal{J}p(t) & \mathcal{J}p(t) & \dots & \mathcal{J}p(t) \\ \mathcal{J}a_1 & \mathcal{J}a_2 & \dots & \mathcal{J}a_{100} \end{bmatrix} \quad (2.50)$$

The resulting covariance matrices are given by

$$\mathbf{V}_c = \begin{bmatrix} \mathbf{s}_c^2 e^{-\frac{3(x_1-x_1)}{l}} & \mathbf{s}_c^2 e^{-\frac{3(x_2-x_1)}{l}} & \dots & \mathbf{s}_c^2 e^{-\frac{3(x_{100}-x_1)}{l}} \\ \mathbf{s}_c^2 e^{-\frac{3(x_2-x_1)}{l}} & \mathbf{s}_c^2 e^{-\frac{3(x_2-x_2)}{l}} & \dots & \mathbf{s}_c^2 e^{-\frac{3(x_{100}-x_2)}{l}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{s}_c^2 e^{-\frac{3(x_n-x_1)}{l}} & \mathbf{s}_c^2 e^{-\frac{3(x_n-x_2)}{l}} & \dots & \mathbf{s}_c^2 e^{-\frac{3(x_{100}-x_{100})}{l}} \end{bmatrix} \quad (2.51)$$

$$\mathbf{V}_a = \begin{bmatrix} \mathbf{s}_a^2 e^{-\frac{3(x_1-x_1)}{l}} & \mathbf{s}_a^2 e^{-\frac{3(x_2-x_1)}{l}} & \dots & \mathbf{s}_a^2 e^{-\frac{3(x_{100}-x_1)}{l}} \\ \mathbf{s}_a^2 e^{-\frac{3(x_2-x_1)}{l}} & \mathbf{s}_a^2 e^{-\frac{3(x_2-x_2)}{l}} & \dots & \mathbf{s}_a^2 e^{-\frac{3(x_{100}-x_2)}{l}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{s}_a^2 e^{-\frac{3(x_n-x_1)}{l}} & \mathbf{s}_a^2 e^{-\frac{3(x_n-x_2)}{l}} & \dots & \mathbf{s}_a^2 e^{-\frac{3(x_{100}-x_{100})}{l}} \end{bmatrix} \quad (2.52)$$

Using Eq. (2.26) and noting that the correlation between  $a$  and  $c$  are zero, the variance of the predicted variable is given by

$$\text{var}(p(t)) = \mathbf{X}_c(t)^T \mathbf{V}_c \mathbf{X}_c(t) + \mathbf{X}_a(t)^T \mathbf{V}_a \mathbf{X}_a(t) \quad (2.53)$$

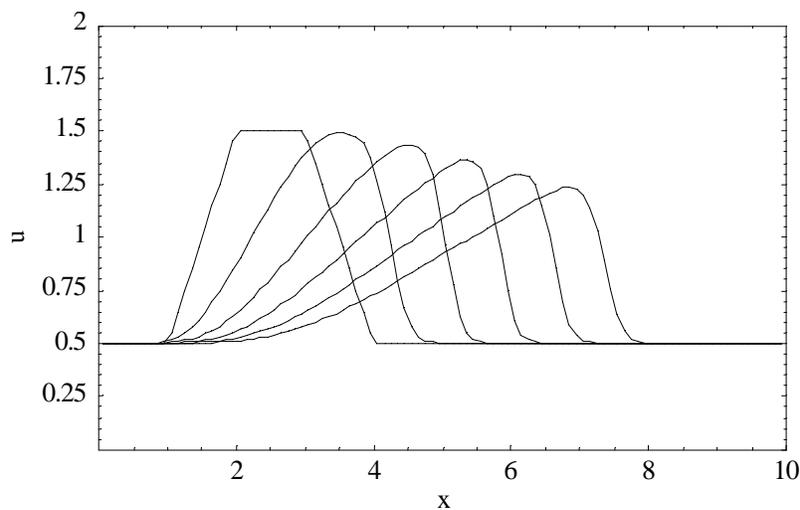
As in previous cases, the sensitivity matrix derivatives are estimated using central finite differences for each time. Each derivative as a function of time requires two solutions of the governing equations. The total number of solutions of Burgers' equation required to estimate the sensitivity vectors is thus 400 (200 for  $a$  and 200 for  $c$ ) plus 1 for the nominal values for  $a$  and  $c$ . We set  $\Delta a$  and  $\Delta c$  to be 1% of the corresponding mean parameter value for the center difference approximation to the sensitivity derivatives.

The resulting evolution of the pulse is shown in Figures 2.20 and 2.21 for a correlation length of  $l = 5$ . Figure 2.20 shows the predicted  $u$  using the mean values of the parameters (i.e.,  $c = 0.4$  and  $a = 0.1$  over the entire domain) and Figure 2.21 shows the results for one of the random realizations of the parameters. Note that the pulse widens due to the expansion at the trailing edge and some compression occurs at the leading edge. Also note that the pulse shows the random spatial dependence of the model parameters through the bumps and valleys in the profile (Figure 2.21).

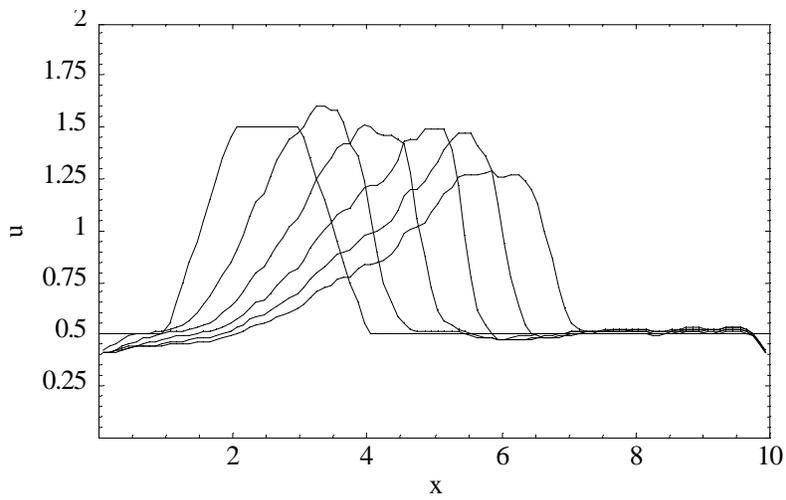
Figures 2.22 and 2.23 show the results of the uncertainty analysis for the location of the leading edge of the pulse as a function of time. Note that there is very little difference between the predicted location of the leading edge using the mean value of the parameters (sensitivity analysis) and the location of the leading edge which results from the average of the Monte Carlo simulations.

The Monte Carlo results for the standard deviation of the front location as a function of time are shown in Figure 2.23 for 8192 simulations and for 16384 simulations. The results are very similar, suggesting that little is gained by running the extra 8192 simulations. This illustrates one of the features of Monte Carlo analysis. The number of simulations required to adequately resolve statistical distributions for the prediction uncertainty is not a strong function of the number of uncertain input variables. For this case, we used approximately the same number of simulations as was used for previous cases, even though we randomly specified spatial distributions for  $a$  and  $c$  in terms of 100 correlated nodal values for  $a$  and 100 correlated nodal values for  $c$ . If, on the other hand, we desire to predict the correlation structure of two prediction variables, the number of simulations required to adequately resolve this correlation structure would increase dramatically.

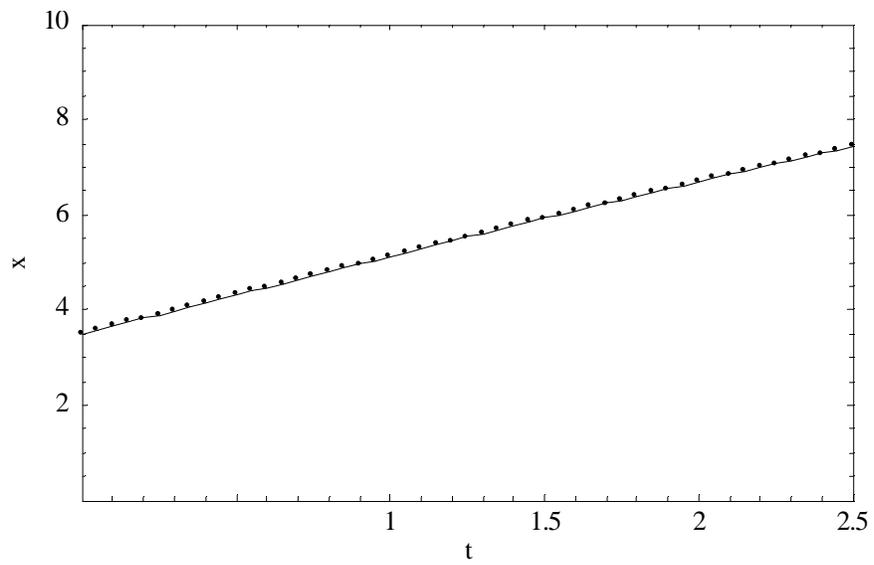
There are significant differences between the Monte Carlo results in Figure 2.23 and those from the sensitivity analysis, also shown in that figure. First, the sensitivity analysis demonstrates a somewhat irregular behavior due to the finite difference approximation. This irregular behavior is caused by the same effect that was demonstrated in the previous example for the diffusion equation. However, the effect is amplified by the limiters in the advective algorithm and by the heterogeneous parameter fields  $a(x)$  and  $c(x)$ . Another difference is that the sensitivity analysis predicts a slightly larger standard deviation than does the Monte Carlo method for large times.

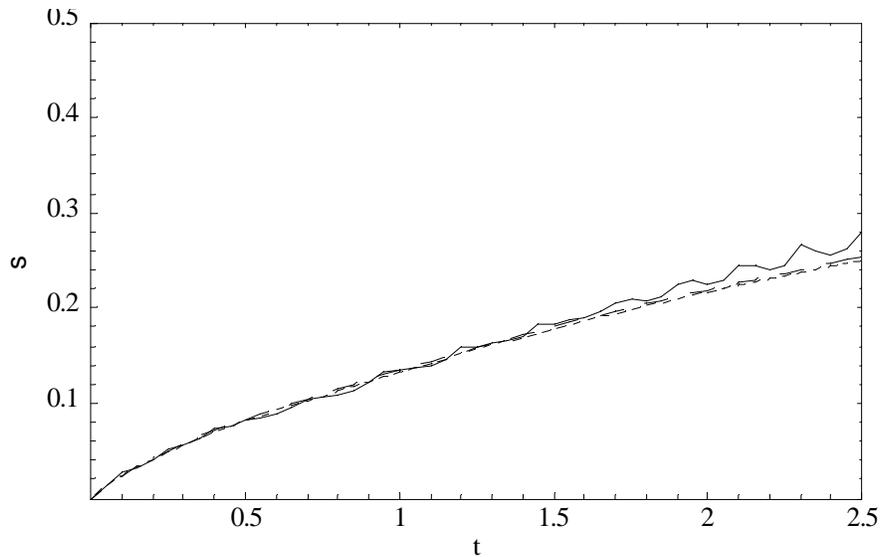


**Figure 2.20. Mean Velocity as a function of  $x$ :  $t = 0.0, 0.5, 1.0, 1.5, 2.0, 2.5$**



**Figure 2.21. Velocity as a function of  $x$ : One Monte Carlo sample,  $t = 0.0, 0.5, 1.0, 1.5, 2.0, 2.5$**





**Figure 2.23 Standard Deviation for Front location as a Function of Time: Monte Carlo results with 16384 simulations represented by long dashes, 8192 simulations represented by short dashes; sensitivity analysis results represented by the solid curve.**

Even though this example is much more complex than previous ones, the methodology used is a direct extension of that used earlier for both the Monte Carlo method and the sensitivity analysis method. The most significant differences are the large number of input parameters (200) and the spatial correlation in the model parameters.

## 2.8 Summary

The basic concepts in estimating uncertainty of model predictions given uncertainty in the model parameters were introduced in this chapter. Several different examples were considered ranging from a simple linear algebraic model to one associated with nonlinear advection with linear diffusion. While these models are significantly different, the approach to estimating the prediction uncertainty was surprisingly similar. The results show that models which are linear in the dependent variables may be nonlinear in the parameters (damped spring mass system). The resulting relation between prediction uncertainty and the uncertainty in the input parameters can be complex and be a strong function of other variables (such as the strong frequency dependence of the spring mass system).

Despite this complexity, the result of the previous analysis indicates that the sensitivity analysis provides a good tool to estimate the variance of the predicted result given the covariance matrix of the input parameters. The sensitivity analysis requires much less computation than does the Monte Carlo method, but does depend on, to some extent, the

linearity of the model. This dependence is demonstrated by the damped, spring-mass system near resonance where the sensitivity method fails to adequately predict the standard deviation. The Monte Carlo method is robust in the sense that it will provide good estimates (assuming a sufficient number of simulations are performed) of uncertainty in the predicted parameters, whether or not the model is highly nonlinear. The Monte Carlo method also provides estimates of the shape of the probability density functions for the prediction uncertainty. Such knowledge is useful when one is interested in the behavior of the distributions near their tails. For example, we must know the behavior near the tails to define the location of statistical bounds, and to perform statistical tests such as tests for model validity. The principle downside to the Monte Carlo method is it requires a very large number of simulations to obtain good estimates for the uncertainty of the predicted quantities.

As this tutorial illustrates, the uncertainty analysis allows one to estimate the uncertainty in the predictions due to uncertainty in the model parameters. However, one must keep in mind that the predicted uncertainty in the prediction variables depends directly on the models used for the uncertainty in the input parameters. There are many cases for which there just isn't enough data to resolve the statistical models for the model parameter uncertainty. In such cases, it may be more appropriate to use analysis of this type illustrated here to scope out the sensitivities of the model predictions to the various model parameters, which allows one to rank order the relative importance of the input parameters. This information can be used to guide future efforts (test programs, data gathering, etc.) to reach the overall goal of a program.

### 3.0 Model Validation

The concepts of model verification and model validation have received much attention in recent years. As we grow more dependent on computer models and less on experimentation, the importance of model verification and validation increases. Generally, we consider model verification to be confirmation that predictions of a numerical model approximate those of the mathematical model to some acceptable level of accuracy. In contrast, we generally consider model validation to be the validation that the predictions of a numerical model approximate the underlying physics of the application being modeled to some acceptable degree of accuracy. A valid model must not only utilize a numerical algorithm that well approximates the solution to the underlying mathematical model, but the underlying mathematical model must also well approximate the underlying physics. This adds a significant complication since we require experimental data to test the validity of a model.

While verification is well established, rigorous model validation is not. In the present tutorial, we demonstrate some of the underlying concepts of statistical model testing that are applicable to complex engineering and physics models. A review of the literature will be presented in a later chapter.

We restrict our attention to validation of the spring-mass-damper model. We choose this model because it is already familiar to the readers of the previous chapter, and at first glance appears simple. However as will be shown, this example does contain enough content to demonstrate the major features of statistical model validation for complex systems.

The model for the ratio of the amplitude of the oscillation of a mass to the amplitude of the harmonic forcing function for a damped spring-mass system is given by Eq. (2.31). We repeat this equation here for convenience.

$$\frac{X}{F} = f(\mathbf{w}, m, k, c) = \frac{1}{\sqrt{(k - m\mathbf{w}^2)^2 + (c\mathbf{w})^2}} \quad (3.1)$$

$k$ ,  $m$ , and  $c$  are the model parameters for the spring constant, mass, and damping coefficient;  $\mathbf{w}$  is the frequency of the forcing function; and  $X/F$  is the ratio of amplitudes. We will test the validity of this model by comparing the predictions of the model against experimental data. We start our discussion with a few questions.

**Question 1:** What are we actually testing when we compare model predictions to experimental observations?

The most we can say is we are evaluating whether model predictions are consistent with experimental observations for that particular validation experiment, or set of validation experiments. We are thus only testing the model for a particular set of conditions, range of input parameters, and range of independent and dependent variables. If we wish to apply a model for a different set of conditions, we must accept the possibility that the model may not be valid under this different set of conditions.

**Question 2:** What is the effect of experimental uncertainty?

Models contain parameters that cannot be measured exactly. For the present example, the values of mass,  $m$ , the spring constant,  $k$ , and the damping coefficient  $c$ , will all have some uncertainty associated with them. The mass can generally be measured with the least uncertainty, the spring constant will have more uncertainty since it requires both a force and a displacement measurement, and the damping coefficient will generally have the most uncertainty since it requires both a force and a displacement with time measurement. In addition, there will be uncertainty in the actual frequencies at which the experiment is being performed and uncertainty in the measured outputs of the experiment. Because of these various components of uncertainty, the chance of exact agreement between the model predictions and the experimental observations is very remote, even if the model is valid.

**Question 3:** What can we actually test given all of these sources of uncertainty?

The most we can test is whether the model predictions are consistent with the experimental observation, given a defined level of uncertainty, for that particular set of experiments.

**Question 4:** What model is actually being tested?

The acknowledgment that uncertainty exists in any validation experiment effectively changes the model that we are testing. Since we are testing whether the model predictions are consistent with the experimental observations within the uncertainty of the experiment, we must include the models for the uncertainty in the predictions and in the experimental observations. Thus the model being tested is actually the underlying physical model and the models for the uncertainty in the model parameters, initial and boundary conditions, and the experimental observations.

**Question 5:** Can two different models give results that are both consistent with the experimental observations?

Yes, two models can give results that are both consistent with the experimental observations, even though one model may not have the correct physics incorporated into

it. This happens when the validation experiments are not specifically designed to detect differences in the models. It is not unusual for one model to contain a component or term that is simply incorrect or coded incorrectly. However, if the effect of this component is not significant over the range of variables for which the experiment is performed, the effect of this component will not be adequately tested.

Unfortunately, while both models provide predictions that are consistent with the experimental observations within the experimental uncertainty, the random nature of this uncertainty can result in the predictions of an invalid model being closer to the experimental observations than those for a valid model. We should not reject one model relative to another simply because it produces a sum of squares of residuals greater than some competing model. We should reject a model only if we have a high level of confidence that it could not predict the experimental observations, given the level of uncertainty in the model parameters and in the experimental observations.

**Question 6:** How do we design the experiments to minimize the possibility that two competing models produce predictions that are both consistent with the experimental observations?

We should design the experiment by using the two predictive models along with the propagation of uncertainty analysis discussed in the previous chapter. This will allow us to evaluate whether the two models will produce statistically different results, given the experimental design. Specifically, the experiment should be designed so that the overlap in the probability density functions for the predictions for the two models is very small, and such that the accuracy of the measured quantities is sufficient to resolve these differences in predictions. This requires a high degree of collaboration between the modeler and the experimentalist during the design process. The traditional environment in which the modeler and experimentalist operate more or less independently greatly increases the chances of nebulous results.

### **3.1 Two Approaches to Validation**

As mentioned above, the chances that complex model predictions will agree *exactly* with the experimental results is very remote. There are always errors in experimental measurements and there is always uncertainty associated with the model parameters. We need some estimate of the overall uncertainty so that we can evaluate the quality of agreement between the experimental measurements and the model predictions.

The standard method in statistics to estimate the overall uncertainty is to perform the experiment, independently, multiple times. If each performance of the experiments is truly independent, then the resulting scatter in the differences between model prediction and the experimental observation can be used to make estimates about the statistics of the uncertainty. We can then evaluate some measure of whether the model predictions and the experimental observations agree within the scatter of the data.

Unfortunately, such multiple, independent runs of the model validation experiments are simply not practical for many of the models of interest. Often, these experiments can be performed only once. We must estimate the prediction uncertainty through analysis. Fortunately, we can do this if we have estimates of the probability density functions for those model parameters whose uncertainty significantly affects the model predictions. Given these estimates, we can perform propagation of uncertainty analysis as shown in the previous chapter to estimate uncertainty in the model predictions. We can then test whether the model predictions are statistically consistent with the experimental observations. The method of propagating the uncertainty in the model input parameters is appropriate if the cost of characterizing the uncertainty in the appropriate model parameters, and the cost of propagating this uncertainty through the model, is less than the cost of repeating the validation experiment a sufficient number of independent times. We use the term cost somewhat generically for purposes of this discussion. It can include the cost in dollars, the required time, the environmental cost, the public perception cost, and any other cost which may occur from repeated execution of the experiments.

There is an added benefit to using the propagation of uncertainty method. The requirement to define the uncertainty in the model parameters forces us to fully understand the sources of uncertainty in our experiments and the uncertainty in the model predictions. In contrast, the standard statistical method of performing repeated experiments to generate enough samples to characterize the uncertainty helps define the level of uncertainty, but does not require that we fully understand the sources of this uncertainty.

### ***3.2 Model Validation for the Damped Spring-Mass System***

Here we assume that our model validation experiment is performed only one time using only a single mass, spring, and damper. We assume the uncertainty in each of the values for mass, spring constant, and damping coefficient are well approximated by normal probability distributions and these values are not correlated.

The second form of uncertainty in our validation experiment is measurement error. We should be able to characterize measurement uncertainty from our experience with the instruments and the measurement environment. For the purposes of this example, we take the measurement error to be well modeled by a normal distribution with a zero mean and a standard deviation of 0.015. The statistics for the various contributions to uncertainty is summarized in Table 3.1.

**Table 3.1. Statistics for the Oscillator Parameters and Measurement Errors: All quantities are assumed to be normally distributed.**

Parameter	Mean	Standard Deviation
$m$	1	0.001
$k$	2	0.05
$c$	0.25	0.05
$X/F$ measurement error	0.0	0.015

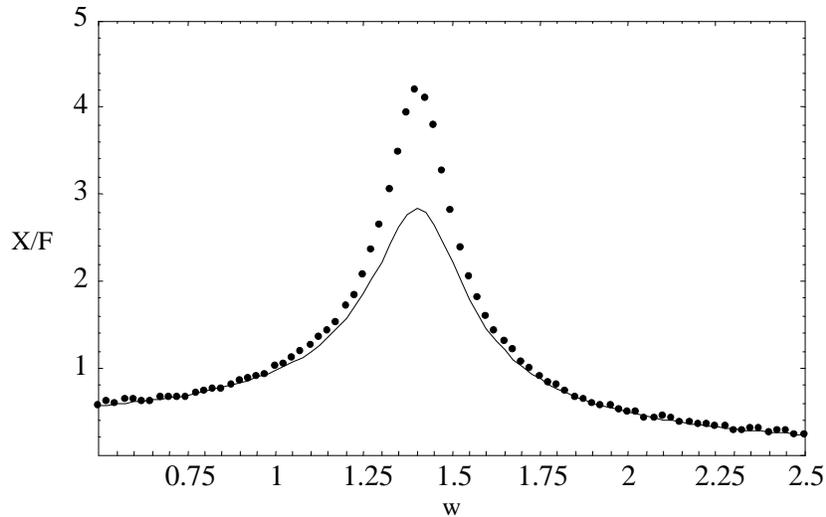
For purposes of demonstration, we simulate the experimental measurement by using the model defined by Eq. (3.1) for a mass, spring constant, and damping coefficient chosen at random from the population of these quantities. We add a random error to the resulting predictions to simulate the effect of the measurement error. Since we are using the same model to simulate the measurements as the model we are testing, we should find that this model is valid!

We begin with simple comparisons of the simulated experimental results to model predictions and move on to more appropriate comparisons.

### 3.3 Simple Comparisons

Figure 3.1 shows a comparison between the simulated measurements and the model predictions using the mean values of the model parameters. Clearly, there are significant differences in the trend of the model predictions relative to the experimental results near the resonance peak. Based on the results of this figure alone, one should not have much confidence that the model is valid. The first question one might ask to more fairly test the model is - do the model predictions of the validation experiment measurements and the actual experimental measurements agree within the measurement error?

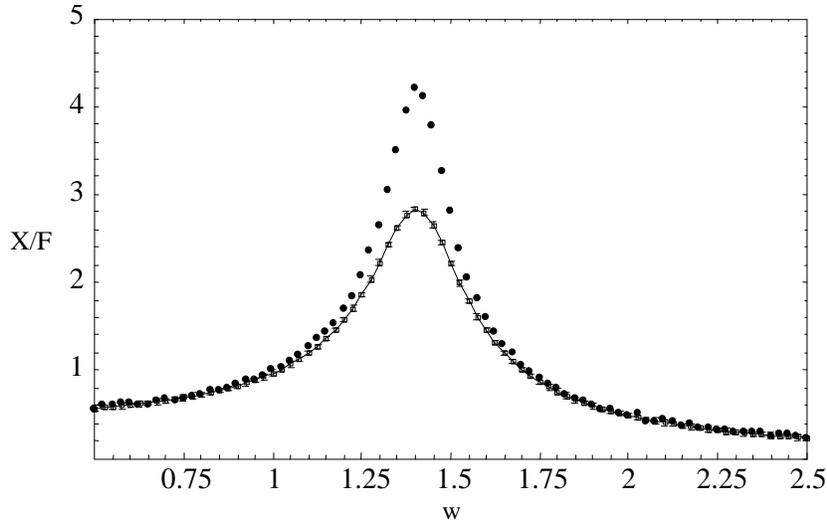
To include the effect of measurement error in  $X/F$ , we re-plotted Figure 3.1 as Figure 3.2 with error bounds added to represent the uncertainty in the experimental measurements. In contrast to the traditional approach, we add these uncertainty bounds to the model predictions rather than to the experimental observations. We do this for two reasons. First, we consider the model for the uncertainty in the experimental measurements as part of the overall model of the validation experiment. Secondly, this approach will be more convenient when we extend this example to incorporate model parameter uncertainty. This approach will result in only one set of error bounds on the predictions that represent both the effects of parameter uncertainty and the effects of measurement error.



**Figure 3.1. Measured and Predicted Spring/Mass/Damper Response: Predicted response represented by the solid line, measured response represented by the solid points.**

The total height of each error bar in Figure 3.2 represents the measurement uncertainty at the 95% confidence level. This height is defined such that if the mean model parameters provide the true value for the  $X/F$  at particular  $w$ , then 95% of the area under the probability density function for the measurement at that particular  $w$  will lie within the range represented by the error bar. In other words, 95% of the measurements at this  $w$ , would lie inside the error bounds if we were to take repeated, but independent measurements. For a normal distribution, the 95% range is given by  $\pm 1.96\sigma$  for the measurements.

Inspection of Figure 3.2 indicates that the error bounds associated with the measurement uncertainty do not contain the experimental observations around the resonance peak. We may suspect that this lack of agreement is due to the additional uncertainty associated with the model parameters and suggests that we should account for this uncertainty in the model. How do we do this? As discussed earlier, we have two choices. We can either repeat the experiment many independent times so that we can estimate the overall effect of uncertainty in the model parameters, or we can use knowledge of the uncertainty of the model parameters, along with a propagation of uncertainty analysis, to estimate the corresponding uncertainty in the model predictions. Here we use the second approach.



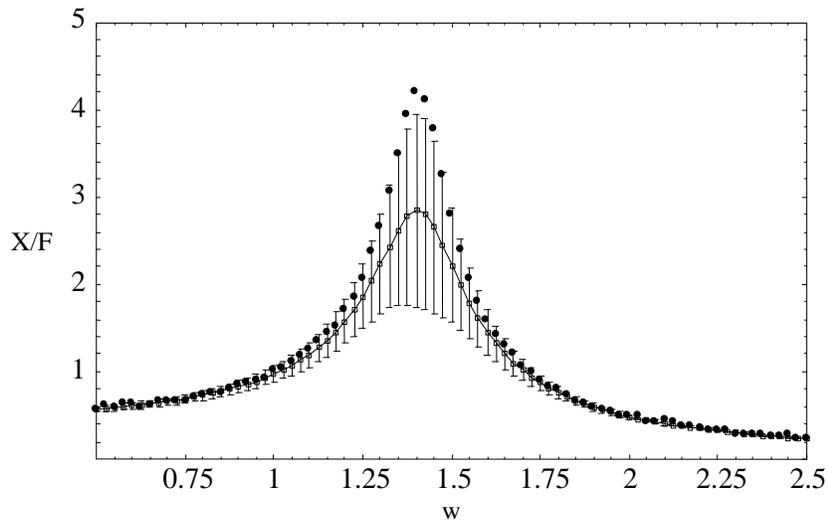
**Figure 3.2. Predicted Spring/Mass/Damper Response including Measurement Errors: Predicted response represented by solid line and the error bars, measured response represented by the solid points.**

Figure 3.3 shows the corresponding 95% confidence level error bars on the predictions due to model parameter uncertainty only (measurement uncertainty is not included). These error bars are based on normally distributed probability density functions and the sensitivity analysis results presented in the previous chapter. Note that the error bars on the predictions do overlap most of the measurements. There are some measurements outside the error bars at large  $w$  and near resonance. We can now add measurement uncertainty to prediction uncertainty as follows.

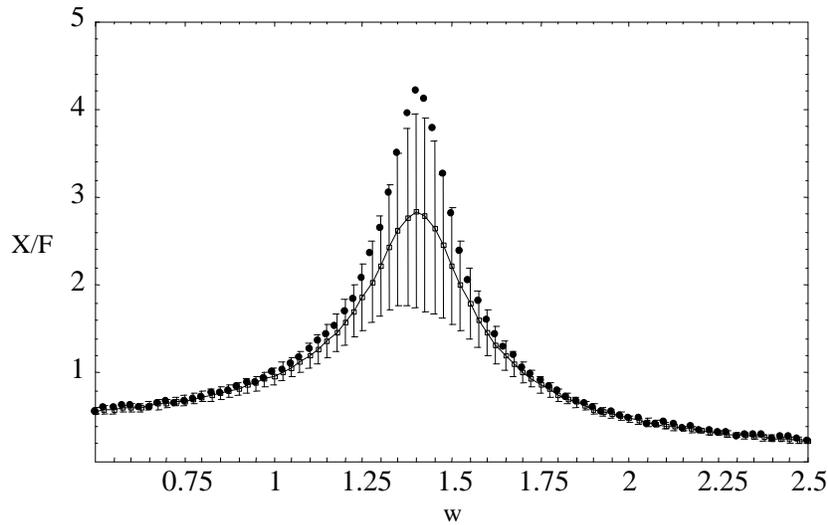
The sum of two uncorrelated random variables has a total variance equal to the sum of the variances of each random variable (see Eqs. (2.10) through (2.12)). The sum of two normally distributed random variables is a normally distributed random variable. The assumption that the experimental observations are independent of the estimates for the model parameters is appropriate if the characterization of the model parameters was performed independently of the model validation experiment. For our case,

$$\mathbf{s}_{total}^2 = \mathbf{s}_{meas}^2 + \mathbf{s}_{pred}^2 \quad (3.2)$$

Using the resulting standard deviations from the sensitivity analysis and the standard deviation for the measurement error (Table 3.1), and assuming that both the prediction uncertainty and the measurement uncertainty are normally distributed, we can evaluate the 95% confidence bars. The 95% confidence levels for the resulting normal distribution occur at  $\pm 1.96\mathbf{s}_{total}$  from the mean. These revised error bars are shown in Figure 3.4.



**Figure 3.3. Measured and Predicted Spring/Mass/Damper Response including Parameter Uncertainty: Predicted response represented by solid line and error bars, measured response represented by the solid points.**



**Figure 3.4. Measured and Predicted Spring/Mass/Damper Response including Parameter Uncertainty and Measurement Error: Predicted response represented by the solid line and error bars, measured response represented by the solid points.**

What can we say about the validity of Eq. (3.1) given these last results? Note that at least 5 out of 81 data points do not lie within the 95% error bounds. Since this represents 7% of the measurements, does this mean the model is not valid since 95% of the measurements do not lie inside the error bars? The answer to this question is a resounding no and the reasons are fairly involved. To develop an understanding of how to test models using multiple measurements, we will first look at the comparison of one measurement to the corresponding model prediction in detail. The analysis will then be extended to two and more measurements.

### **3.4 Monte Carlo Model Testing using One Measurement**

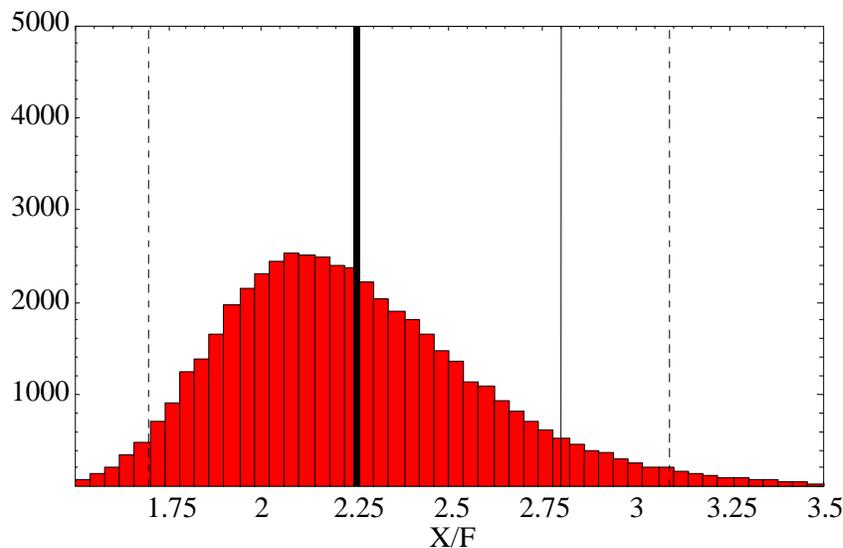
First, we will define some sort of acceptance region for differences between the model predictions and the experimental observations for a single measurement. For this example, we take the measured  $X/F$  at  $w = 1.5$  as our single test measurement. In the previous chapter, we showed that the prediction uncertainty of this ratio near this frequency possessed a skewed probability density function. Here we intentionally choose this measurement frequency to illustrate the effect of this skewness. Because of the skewness, we use the Monte Carlo method rather than the sensitivity method to evaluate the total uncertainty. First, the Monte Carlo method is used to generate a population of possible predictions given samples from the model parameter populations (as we did in the previous chapter). We then add randomly generated measurement error to these Monte Carlo predictions of  $X/F$  to generate a set of model predictions which include the uncertainty due to both the model parameters and the measurements. Note that we are again considering the overall model to include our models for the uncertainty in both the parameters and the measurements.

Figure 3.5 shows the corresponding histogram for the resulting Monte Carlo analysis using 50000 simulations. Also shown are the corresponding simulated measurement (shown as a thin vertical line) taken from the previous example (see Figure 3.1 for  $w = 1.5$ ), and the mean (shown as a thick vertical line) of the resulting Monte Carlo model predictions. We also show the 95% confidence bounds (shown as dashed vertical lines) on the total model predictions. These bounds are easily estimated from the population of the Monte Carlo predictions such that 2.5% of the predictions are equal or above the upper bound and 2.5% are equal or below the lower bound.

Inspection of Figure 3.5 indicates that the 95% confidence bounds are different distances from the mean. This is expected since the distribution is skewed. The figure also shows that the experimental measurement of  $X/F$  at  $w = 1.5$  is clearly within the 95% confidence bounds. This indicates that the model is consistent with the experimental results, given the uncertainty in the model predictions. If the measurements were outside these bounds, we would reject the model at the 95% confidence level. In statistical terms, if we repeated the experiment many independent times and if the model is correct (including the models for parameter and measurement uncertainty), then 95% of the experiments would produce an experimental observation of  $X/F$  for  $w = 1.5$  within the bounds shown in Figure 3.5.

Conversely, 5% of the experiments would produce measurements outside these bounds. There is thus a 5% chance of declaring a model invalid, when in-fact, the model is valid. For the case of the present measurement, we say that the model is valid at the 95% confidence level.

This points out one of the features of using standard statistical inference that is often misunderstood by those not familiar with statistical methods. When the measurement lies within the 95% confidence bounds, we are not saying that we have 95% confidence that the model is valid. We are saying that we have a 5% chance of rejecting a valid model using these bounds. We could use higher confidence levels, say 99%, which would widen our acceptance region. While this would reduce our chances of rejecting a valid model, it would also increase our chance of accepting an invalid model.



**Figure 3.5. Histogram for Monte Carlo Predictions at  $w=1.5$ : Predicted mean given by the thick line, experimental measurement given by the thin line, 95% confidence bounds given by the dashed lines**

The above procedure represents a statistical test by which we can declare a model invalid, given one observation of  $X/F$ . We can perform a more rigorous test if we use experimental observations at multiple values of the frequency  $w$ .

### **3.5 Model Testing for Multiple Measurements**

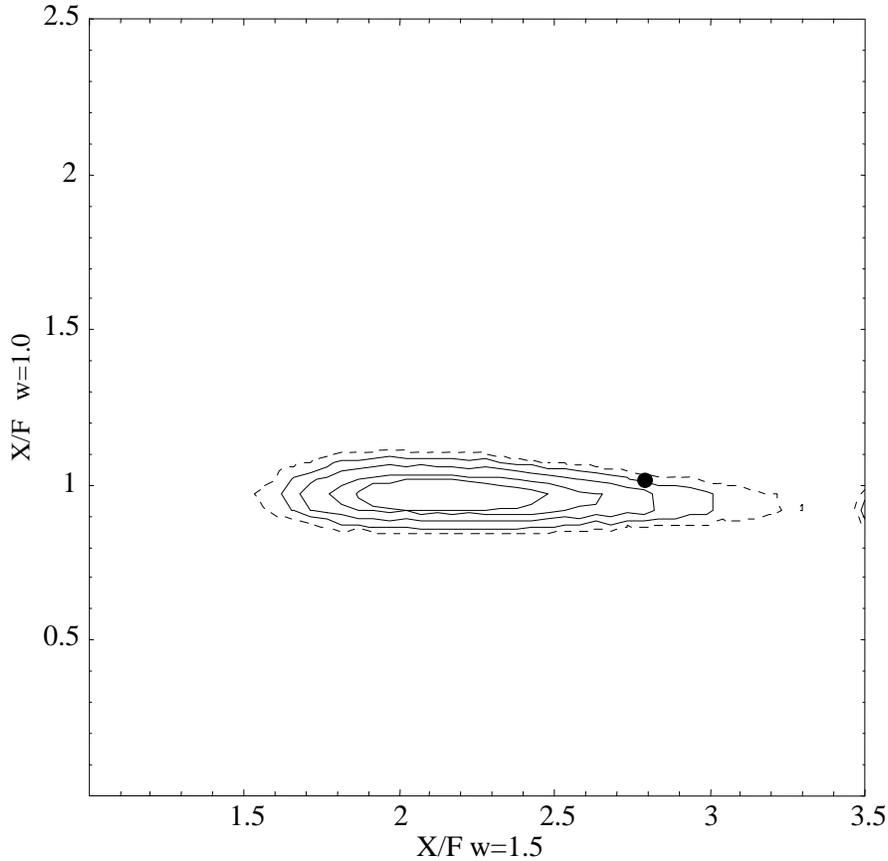
We now look at the use of more than a single measurement for model testing. We first consider the case of two measurements at  $w = 1.0$  and  $1.5$ . The corresponding Monte Carlo generated two-dimensional histogram using 50,000 samples (including the effect of measurement error on both measurements) is shown as a contour plot in Figure 3.6. The 95% confidence region is also shown along with the corresponding measurements for the

two frequencies. The boundary of this confidence region is defined by the iso-count contour that contains 95% of the Monte Carlo simulations. Note that the acceptance region is somewhat elliptical in shape. The roughness of the contours is due to the difficulty associated with resolving multidimensional contours at low probability levels using Monte Carlo analysis. Increasing the number of prediction variables (i.e., the number of  $\mathbf{w}$  at which the model is tested at) will increase the number of Monte Carlo simulations required to adequately resolve the multidimensional contours. While the 10000 simulations were adequate to resolve the shape to the population for one prediction frequency (see the results presented Chapter 2, in particular Section 2.7), 50000 simulations were not adequate to give smooth contours for the two dimensional results of Figure 3.6. However, one can obtain a good feel for the general shape of the acceptance region.

Inspection of Figure 3.6 shows that the measurements are in the 95% confidence region. Therefore we do not have sufficient statistical evidence, at the 95% confidence level, to state that this model is invalid. Figure 3.6 also shows that the acceptance region for the measurement at  $\mathbf{w} = 1$  depends on the value of the measurement at  $\mathbf{w} = 1.5$ , and vice versa. The acceptance ranges of these two measurements are clearly interrelated. This interrelation is totally ignored by the results presented in the previous section and must be included to evaluate confidence levels for more than one measurement.

The extension of this method to more measurements is conceptually straightforward if not practical. If we have  $m$  measurement locations (i.e., frequencies in this case), we could conceptually generate similar  $m-1$  dimensional contour surfaces, including a surface bounding the 95% confidence region. If our set of  $m$  measurements, which is represented by a point in the  $m$  dimensional space, lies outside this confidence region, we reject the model at the 95% confidence level. Of course, it is not possible to plot such contours for more than a few dimensions. In addition, as the previous results illustrated, increasing the number of measurement locations requires evaluating the corresponding  $n$ -dimensional histogram and the corresponding  $n-1$  dimensional contour surfaces. The number of Monte Carlo simulations required to resolve these histograms goes up rapidly with the number of measurements. As a result, the computational requirements of such a procedure rapidly becomes excessive as  $n$  increases. More efficient methods are desired.

We now look at alternatives to the method discussed above. We begin with the special case of normally distributed model prediction uncertainty.



**Figure 3.6. Contour Plot for Histogram Counts for Monte Carlo Predictions: Solid contour lines are for Counts = 125, 250, 500, 750. The dashed line represents the 95% confidence region. The heavy point represents the measurement pair  $(X/F(1.5), X/F(1.0))$ . The Monte Carlo Predictions were counted into a 50 x 50 grid of bins.**

### **3.6 Normally Distributed PDF**

The prediction uncertainty was well modeled by normally distributed probability density functions for many of the examples presented in the previous chapter. When such is the case, an alternative approach can be used that does not require that we generate  $n-1$  dimensional contour surfaces.

Assume that the PDF for the total model uncertainty of  $X/F$  (including the measurement uncertainty) is well modeled by jointly distributed normal probability density functions. While uncertainty in the model prediction at  $w = 1.5$  is clearly not normally distributed, we apply this assumption for this test case to illustrate the effect of this assumption. We initially restrict our attention to the two measurements of  $X/F$  at  $w = 1$  and at  $w = 1.5$ . Since we are assuming normal distributions for the prediction variables, we could use a

sensitivity analysis to estimate prediction uncertainty without loss of generality. However, to perform a statistical test, we must possess an estimate of the correlation between the uncertainty in different predicted measurements. While this can be done using an extension of the sensitivity analysis approach shown in the previous chapter, it is somewhat beyond the scope of this tutorial. Here we estimate the correlation using the 50000 Monte Carlo simulations and Eq. (2.7). The resulting means and the resulting covariance matrix, including measurement uncertainty, were found to be

$$\left[ \langle X / F_{mean} (1.5) \rangle \quad \langle X / F_{mean} (1.0) \rangle \right] = \left[ \langle p_{mean1} \rangle \quad \langle p_{mean2} \rangle \right] = [2.2518 \quad 0.9709] \quad (3.3)$$

$$\langle \mathbf{V} \rangle = \begin{bmatrix} 0.128277 & -0.003228 \\ -0.003228 & 0.002471 \end{bmatrix} \quad (3.4)$$

Note that we show the quantities in Eqs. (3.3) and (3.4) as estimates since they were evaluated from a finite sample from the total population of model parameters. It can be shown that for multivariate distributions symmetric about their means, contours of constant probability are given by the following ellipses (Beck and Arnold, p. 294) where  $r$  is constant for iso-probability curves.

$$r^2 = \begin{bmatrix} p_1 - p_{mean1} & p_2 - p_{mean2} & \cdots & p_n - p_{mean n} \end{bmatrix} \mathbf{V}^{-1} \begin{bmatrix} p_1 - p_{mean1} \\ p_2 - p_{mean2} \\ \vdots \\ p_n - p_{mean n} \end{bmatrix} \quad (3.5)$$

Each  $p_i$  represents a possible prediction of  $X/F$  for the measurement frequency  $w_i$ . Since the distributions are assumed to be normally distributed,  $r^2$  can be related to the probability through the  $\chi^2$  distribution (Beck and Arnold, p. 294):

$$r^2 = l_{1-\alpha}^2(n) \quad (3.6)$$

where the value of  $l$  is the value associated with the  $100(1-\alpha)$  % confidence region for  $n$  parameters in the model. These values are tabulated in many statistical textbooks (Brownlee, 1965). Figure 3.7 shows the constant probability curves for the normal distribution defined by Eqs. (3.3) and (3.4). For the case of 2 variables at the 95 confidence level, the critical value for  $r^2$  is

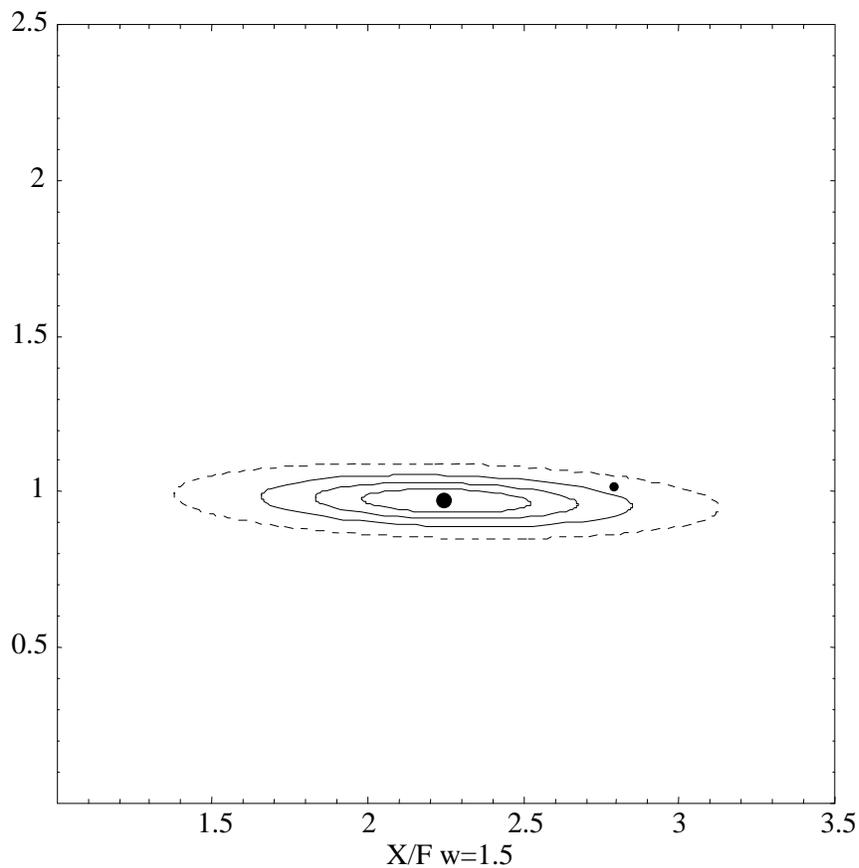
$$r_{critical}^2 = l_{0.95}^2(2) = 5.99 \quad (3.7)$$

Note that there are similarities between the shape of the curves in Figures 3.6 and 3.7. We plot the 95% confidence region from both methods in Figure 3.8. Comparison of the two curves indicates that they are similar. There does appear to be some skewing of the Monte Carlo based confidence region to the right. This is expected since the Monte Carlo

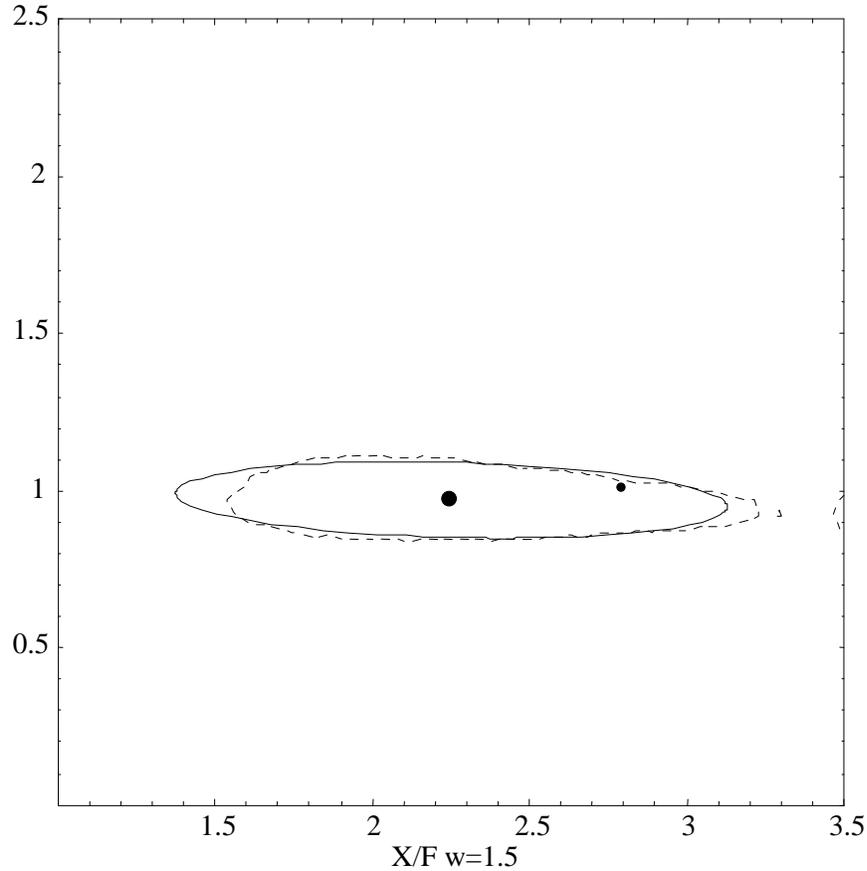
analysis of the  $w = 1.5$  prediction is skewed (see Figure 3.5). If the prediction uncertainty were normally distributed, then we would expect to obtain good agreement between the two curves of Figure 3.8. The results of Figure 3.8 illustrate that the measurement point clearly lies inside the 95% confidence regions for both methods. As a result, we do not possess sufficient statistical evidence, at the 95% confidence level, to declare the model invalid.

We can easily extend this method to all 81 measurements since we are assuming that the joint probability distribution is normally distributed. The critical value for 81 measurements at the 95% confidence level is (Brownlee, 1965)

$$r_{critical}^2 = l_{0.95}^2(81) = 102.9 \quad (3.8)$$



**Figure 3.7. Equal Probability Contours for the Multivariate Normal Distribution: Solid contour lines are for probabilities = 25%, 50%, 75%. The dashed line represents the 95% confidence region. The light point represents the measurement pair  $(X/F(1.5), X/F(1.0))$ . The heavy point represents the mean prediction.**



**Figure 3.8. Comparison of the 95% Confidence Regions: Monte Carlo results represented by the dashed curve. Normal approximation to the represented by the solid curve. The light point represents the measurement pair  $(X/F(1.5), X/F(1.0))$ . The heavy point represents the mean prediction.**

Again, we evaluated the 81 means and the elements in the 81 by 81 covariance matrix using the 50000 Monte Carlo simulations. Given these, we evaluate the corresponding  $r^2$  value from Eq. (3.5).

$$r^2 = 95.7 \quad (3.9)$$

This value is less than the critical value of 102.9. Thus we do not possess statistically significant evidence to declare our model invalid. One can think of  $r^2$  as a square of a weighted distance from the mean where the weighted distance is measured to equal probability surfaces. Since the weighted distance from the mean to the measurement point is less than the weighted distance to the 95% confidence surface, we have no statistically significant evidence to declare the model invalid. Such measures are appropriate when the

joint PDFs of the model uncertainty (including the measurements) are in the form of ellipses around the means.

### **3.7 The Problem with the Sum of Squares of Residuals**

A common approach to measure the agreement between model predictions and experimental observations is to take the sum of the squares of the residuals (SSR) between the measurements and the corresponding model predictions. Mathematically, this measure is given by

$$SSR = \sum_{i=1}^n (m_i - p_i)^2 \quad (3.10)$$

where  $m_i$  and  $p_i$  represent the measurement and the prediction for measurement location  $i$  (i.e., the measurement at a particular frequency in this case). If we take the prediction to be based on the mean of the model parameters, then curves of constant SSR will be circles centered on the prediction. If the probability density function were the same for each measurement location (i.e., for each frequency for the present example), and if each probability density function was symmetric about the mean with the same standard deviation, then these circles would represent curves of constant probability.

Unfortunately, the standard deviation was found to be a strong function of  $\mathbf{w}$  for the damped spring-mass system. In fact, the standard deviation was found to be a function of some variable for all of the examples presented in the previous chapter. Because of this, curves of constant SSR do not represent curves of constant probability for the problems illustrated here. For such cases, SSR is not an appropriate measure of model validity if one wishes to attach statistical significance to the results.

### **3.8 An Alternative Approach**

We end our tutorial of statistical model validation with the presentation of a non Monte Carlo based approach which does not require that the prediction uncertainty be normally distributed. It does require that well defined statistical models can be used for the input parameters and for the measurement uncertainty. In the previous examples, we used normally distributed (possibly correlated) probability models for input parameters and for the measurement. We can take advantage of this knowledge by defining our confidence regions in terms of these quantities directly rather than performing a Monte Carlo analysis to estimate prediction uncertainty. To demonstrate this, let us return to the example of one measurement at  $\mathbf{w} = 1.5$  for our damped spring-mass system. We will also restrict the uncertainty in the model parameters to the damping coefficient only. The other parameters are assumed known exactly. Both of these assumptions will be relaxed later.

We begin by plotting the uncertainty in the measurement and in the model parameter  $c$  on the same plot. We use the actual measurement as an estimate for the mean of the measurement population (this assumption is discussed in the following chapter that

surveys the literature). We use the mean of the input model parameter and the standard deviation of this parameter to describe the uncertainty for this model parameter. Since the uncertainty in the measurement comes from a different source than the uncertainty in the model parameter, these uncertainties will not be correlated. The resulting confidence region will be an ellipse whose major axis is horizontal or vertical. The 95% confidence region is shown in Figure 3.9 for the two variables and is defined using Eq. (3.5).

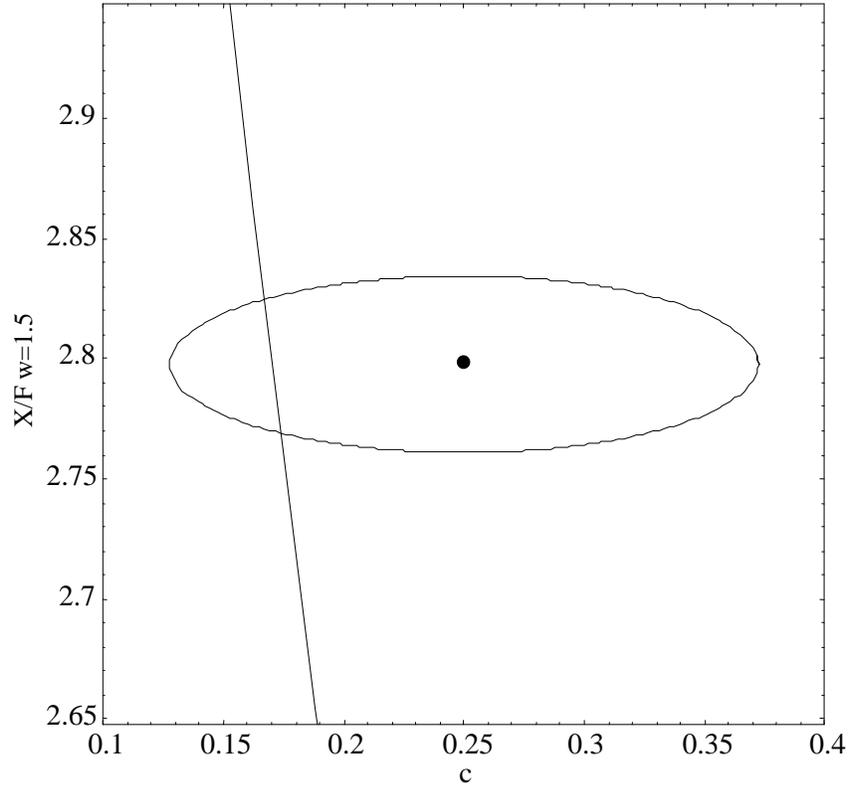
We also graph the model prediction (the straight line) as a function of the model parameter on the same plot (Figure 3.9). Assuming that our measurement provides a correct estimate of the mean measurement, a valid model will pass through the confidence region shown in Figure 3.9 for some value of the model parameter. In contrast, for a model to be declared invalid at the 95% confidence level, there can be no value for the parameter  $c$  that results in the model prediction passing through the confidence region. This provides an alternative test of model validity. An advantage of this approach is we can use optimization techniques rather than the Monte Carlo analysis presented earlier to evaluate whether there is a value for the parameter  $c$  that results in the model prediction lying within the ellipse of Figure 3.9. Optimization techniques (even those based on Monte Carlo techniques such as thermal annealing) do not require as many simulations as does a full Monte Carlo analysis of model uncertainty. If there is a value for  $c$  which lies in the ellipse then there is no statistically significant evidence that the model is invalid. If there is no such value, then we reject the model with 95% confidence.

To apply an optimization technique, we must first define an objective function or distance measure. Since we are presumably dealing with simple distributions, we can define this measure in terms of equal probability curves directly from these distributions. For example, the equation of the ellipse shown in Figure 3.9 resulting from normal probability distributions is given by (see Eqs. (3.5)-(3.7))

$$\left( \frac{c - c_{mean}}{\mathbf{s}_c} \right)^2 + \left( \frac{X / F - \text{Measured} X / F}{\mathbf{s}_p} \right)^2 = l_{0.95}^2(2) = r_{crit}^2 \quad (3.11)$$

where  $l$  is a number based on the level of confidence and the number of parameters in Eq. (3.11) (in this case, there are two parameters,  $c$  and  $p$ , the measurement at  $\mathbf{w} = 1.5$ ). For normal probability distributions, we can use the  $\chi^2$  statistic to evaluate  $l$  as discussed for Eq. (3.6). For the present case, the value of  $l = r_{crit}$  for the 95% confidence with  $n = 2$  is  $r_{crit} = 2.447$ . The corresponding distance measure is

$$\left( \frac{c - c_{mean}}{\mathbf{s}_c} \right)^2 + \left( \frac{X / F - \text{Measured} X / F}{\mathbf{s}_p} \right)^2 = r^2 \quad (3.12)$$



**Figure 3.9. 95% Joint Confidence Region and Model Predictions: The point is at  $(c_{mean}, \text{Measured } X/F(1.5))$ .**

The  $r$  for  $(c, X/F(1.5))$  pairs that lie inside the confidence region will be less than or equal to 2.447. Thus, we cannot declare a model invalid if we can find just one value  $c$  for which the predicted  $X/F(1.5)$  gives  $r \leq 2.447$ . Figure 3.10 shows the corresponding  $r$  as a function of  $c$  for the present model. Clearly, values of  $c$  exist for which the model predictions give an  $r$  which lies below  $r = 2.447$ . There is no statistical evidence that the present model is invalid at the 95% confidence level using just one measurement.

The extension of the above method to multiple measurements and multiple model parameters is straightforward. For the case of two measurements and three model parameters, the confidence region will be given by a hyper-ellipse in 5 dimensional space. For this case, the equation for the hyper-ellipses is given by

$$r^2 = \mathbf{d}^T \mathbf{V}^{-1} \mathbf{d} \quad (3.13)$$

where

$$\mathbf{d}^T = [m - m_{mean} \quad k - k_{mean} \quad c - c_{mean} \quad X / F(1.5) - meas(1.5) \quad X / F(1) - meas(1)] \quad (3.14)$$

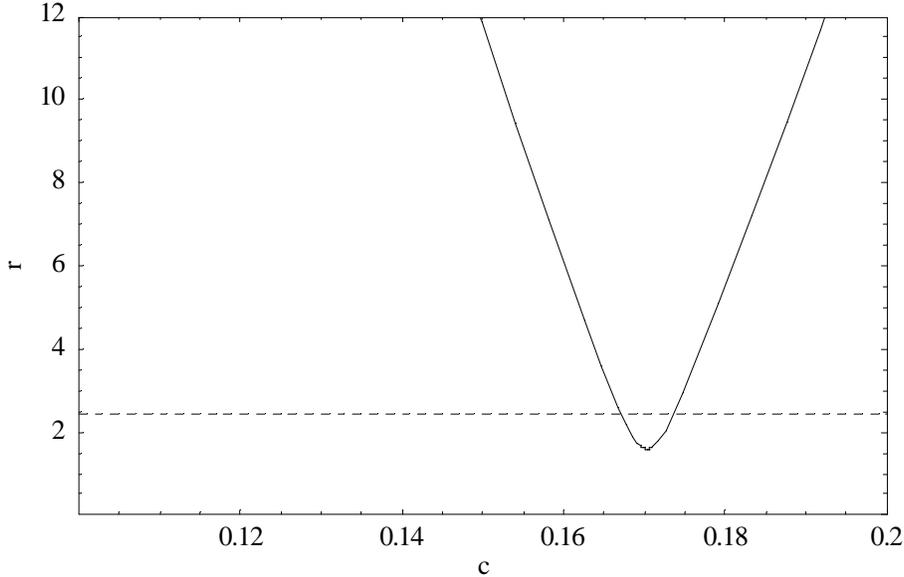
$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{\text{param}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{\text{meas}} \end{bmatrix} \quad (3.15)$$

and  $\mathbf{V}_{\text{param}}$  and  $\mathbf{V}_{\text{meas}}$  are the covariance matrices of the measurements and the parameters. In our case, we have

$$\mathbf{d}^T = [m-1 \quad k-2 \quad c-0.5 \quad X/F(1.5) - \text{meas}(1.5) \quad X/F(1.0) - \text{meas}(1.0)] \quad (3.16)$$

$$V = \begin{bmatrix} 0.001^2 & 0 & 0 & 0 & 0 \\ 0 & 0.05^2 & 0 & 0 & 0 \\ 0 & 0 & 0.05^2 & 0 & 0 \\ 0 & 0 & 0 & 0.015^2 & 0 \\ 0 & 0 & 0 & 0 & 0.015^2 \end{bmatrix} \quad (3.17)$$

For 5 model parameters, each normally distributed, the corresponding critical distance is  $r_{\text{crit}} = 3.33$ . If we can find a set of input parameters  $m, k, c$  such that the model predictions for the measurements  $X/F$  at  $w=1.0, 1.5$  lead to an  $r$  less than  $r_{\text{crit}}$ , then we do not have statistically significant evidence that the model is invalid.



**Figure 3.10. Distance  $r$  as a Function of  $c$ : The critical value of  $r$  is represented by the horizontal dashed line.**

To evaluate whether such a set of parameters exist, we use an optimization algorithm to find the values of the three model parameters that minimize Eq. (3.13). Using the mean values as initial guesses, we found that the parameters that minimize Eq. (3.13) are given by  $m=1.000, k=1.984, c=0.159$  with the resulting distance given by  $r=1.90$ . This value is

less than the critical value of 3.33. Therefore, we have no statistically significant justification to reject the model at the 95% confidence level.

This methodology can also be applied to the full set of 81 measurements shown in Figure 3.1. For 81 measurements and 3 model parameters, the critical distance is  $r_{\text{crit}} = 10.31$ . The minimization procedure yields the parameter set  $m=1.000$ ,  $k=1.996$ ,  $c=0.168$ , which results in  $r=9.77$ . This distance is less than the critical distance. Therefore, the model cannot be declared invalid at the 95% confidence level.

What are we really doing when we apply this method? We are basically using parameter estimation techniques to see if there is a set of parameters that result in model predictions that are consistent with the experimental observations. However, there is a very significant difference between this approach and traditional parameter estimation. Here we account for our knowledge of the natural variability in the model parameters and we evaluate whether the resulting parameter set and the corresponding model predictions are likely at some confidence level. For example, there can be many models with parameters that can be chosen to give good agreement with a particular set of experimental observations. However, the set of these models that utilize reasonable values for the parameters is limited. The method presented here restricts the set of acceptable parameters and corresponding model predictions to those that are reasonable at the 95% confidence level.

### ***3.9 Engineering Validation and Approximate Models***

In the previous examples, we evaluated whether a model's predictions were consistent with the experimental observations to within the uncertainty associated with the parameter uncertainty and measurement error. We intentionally chose very large uncertainties in the model parameters to clearly illustrate the effect of these uncertainties and the effects of nonlinearity in the model predictions.

In the testing of engineering models, we often have the ability to accurately measure the model parameters for the actual model validation test configuration. For example, while there may be significant uncertainty in the damping coefficient from damper to damper for our damped spring mass system, we can often accurately measure the damping coefficient for the actual damper used for the validation test. In this case, the uncertainty in the damping coefficient can be significantly less than that illustrated in the previous examples.

There are many engineering and scientific models for which the intent is to not reproduce the behavior of a physical system exactly, but simply to approximate the behavior to some acceptable level of error. Such models can be declared invalid by the previously defined methodology even though they provide predictions within an error that the user finds acceptable for a particular application. This additional but acceptable error can be due to known approximations in the mathematical models used to expedite computation, or can

be due to known or unknown second or higher order physical effects which are not incorporated into the model. The model validation methodology presented previously can be extended to handle additional acceptable error if a definition for this error can be provided. We refer to this extended methodology as Engineering Model Validation.

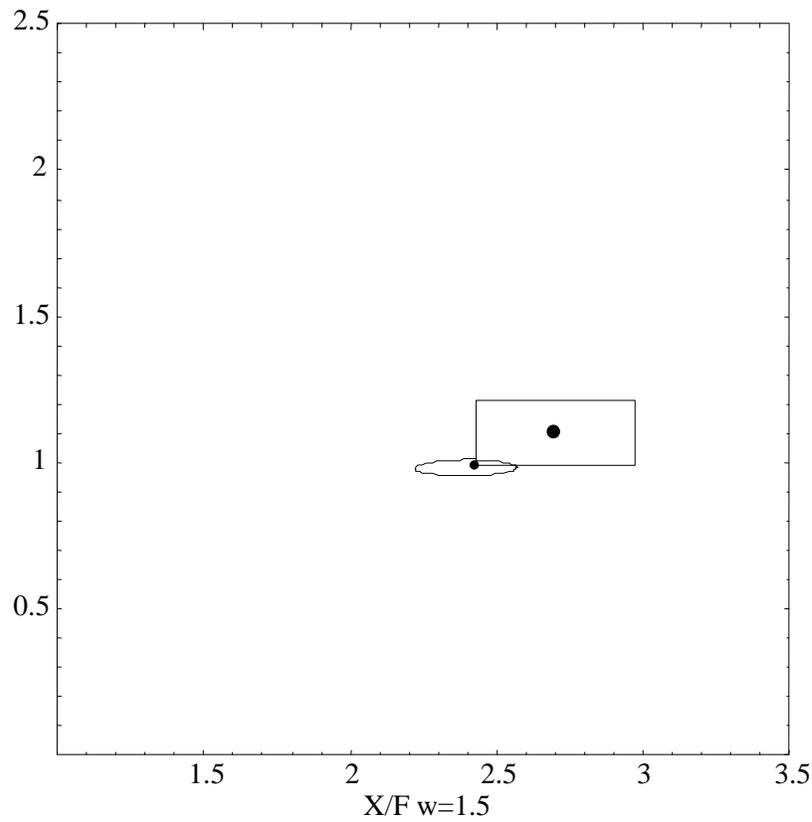
Let us assume that we wish to test whether a model's predictions are consistent with experimental observations. We wish to account for the predicted uncertainty in the model parameters, the experimental error, and an additional acceptable error of  $\pm 10\%$  of the model's mean predictions. Assume that the two-measurement validation experiment discussed in Sections 3.5 and 3.6 returns the measured values for the mass, spring constant, damping coefficient, and the two  $X/F$  measurements shown in Table 3.2. The uncertainties for each of these quantities are represented by their standard deviation and are also shown in the table. Note that we used smaller values for the standard deviations for our measurements of  $k$  and  $c$  than shown in Table 3.1, reflecting the fact that we can measure these quantities directly for the system used for the validation experiment.

**Table 3.2. Oscillator Parameters and Measurement Errors.**

<b>Parameter or Measurement</b>	<b>Measured Value</b>	<b>Standard Deviation</b>
$m$	1.0005	0.001
$k$	1.995	0.005
$c$	0.22	0.01
$X/F(1.0)$	1.10	0.01
$X/F(1.5)$	2.70	0.01

The resulting prediction confidence region, using the normal approximation to the uncertainty analysis as discussed in Section 3.6, is shown in Figure 3.11. Note that the uncertainty of the predicted measurements is considerably less than that shown in Figure 3.8 due to the decreased uncertainty in the model parameters. Also note that the measurement pair lies outside the 95% confidence region for the model predictions. The experimental observations are not consistent with the model predictions at the 95% confidence level and the model would therefore be rejected by the methods of the previous section.

In contrast to the methodology of the previous sections, here we do not declare the model invalid if the experimental observations are within some additional acceptable error of the model predictions. We can represent this additional error by a region about the measurement as shown in Figure 3.11. For this example, the additional error is  $\pm 10\%$  of the model's mean predictions, which results in a rectangular region about the measurements as shown in the figure. We cannot declare this model invalid because there is at least one point in the rectangular region that is also within the 95% confidence region of the model predictions.



**Figure 3.11. Engineering Validation:** The ellipse represent the 95% confidence region of the model prediction (including measurement error). The heavy point represents the measurement pair. The rectangle region represents the acceptable additional error ( $\pm 10\%$  of mean prediction). The light point represents the most probable model prediction lying in the rectangular region.

Extension of this methodology to additional measurements is straight forward in concept. For the present example, extension to all 81 measurements would require searching through the 81 dimensional hyper-prism to evaluate whether there is at least one point that also lies in the 95% confidence region of the model predictions. If there is at least one point, we cannot declare the model invalid.

Other engineering metrics for the allowable error can be defined. For example, we can use an absolute error rather than a percentage error. Or alternatively, we can require that the model be conservative in the sense that we will accept an overprediction of  $X/F$  for a reasonably probable prediction (i.e., those within the 95% confidence region of Figure 3.11), but not an underprediction. In this case, the acceptable error region will include all points to the right and above the measurement point. The model presented here clearly

does not satisfy this condition as all predictions within the 95% confidence region lie below and to the left of the measurement point.

The choice of the metric of this additional acceptable error is somewhat arbitrary and should be based on the particular application of a model. As a result, the appropriate metric will vary from application to application and the resulting engineering validation will be less general than the scientific validation discussed in previous sections.

### ***3.10 What Does This All Mean?***

As the previous discussion illustrates, testing models against experimental data is not simply a matter of comparing plots of model predictions to experimental observations. One has to ask whether the differences between measurements and model predictions are significant. To evaluate the significance of these differences, one must have some sort of measure of the uncertainty in the validation experiments. This requires the experiment to be repeated many independent times, or requires that we have more detailed knowledge of the sources of uncertainty so that we can propagate this uncertainty through the model.

Standard statistical methodology is designed such that the chances of rejecting a valid model or hypothesis are small. A model is considered valid until it is proven otherwise with a high level of confidence. Poor experimental design with significant uncertainty in the model input parameters and the experimental measurements, and high sensitivity to the model parameters will require very large acceptance bounds. This increases the chances of accepting an invalid model. Proper experimental design is required if one wishes to reduce the chances of accepting invalid models. If one wishes to evaluate two competing models, the experiment can be designed using the techniques presented here such that the probability density functions for the two models do not significantly overlap for the anticipated results. This increases the power of the experiment to reject the invalid model while accepting the valid one. Proper design for single and multiple model testing thus requires extensive planning, modeling, and extensive up-front collaboration between the modelers and the experimentalist.

Even with this planning, there are situations when experiments simply cannot be designed such that there is little overlap in the appropriate probability density functions for two models. For example, models often contain additional terms to allow for property dependence on temperature or pressure. The measurement of the parameters in these additional terms is often difficult and can contain significant uncertainty. As a result, the model may be conceptually correct, but now possesses more uncertainty in the predictions due to the added uncertainty resulting from the additional parameters. From a validation perspective, this results in an increased size of the acceptance region which reduces the chances of rejecting the model. However, the acceptance of such a model is not without cost. Larger acceptance regions follow directly from greater uncertainty in the model predictions.

This brings us to the point of model building and model validation. We build models because we wish to understand and predict the behavior of engineered or naturally occurring systems. We desire that these predictions be sufficiently accurate such that we can utilize the predictions to make informed judgments about the behavior of these systems. If the uncertainty in the model input parameters is large, we may not be able to adequately predict the behavior of the system with the required accuracy. Thus models that allow one to define many parameters, such as a full conductivity tensor rather than a simple bi-directional conductivity, may not be more useful than simpler models if the added uncertainty due to these difficult-to-measure additional parameters results in significantly more uncertainty in the predictions.

In the following chapter, we provide a literature review related to the validation of scientific models. The focus will continue to be on statistical model validation methodology.

## 4.0 Literature Review

### 4.1 Introduction

The idea of a *model* takes on different meanings to different researchers. Models of physical phenomena are mathematical or conceptual models developed by humans to describe the behavior of physical systems. Their development can be gradual or can be punctual (Kuhn, 1970). Models that are accepted by one generation may be rejected by another. Depending on one's point of view, models are developed to represent scientific truth or may simply be tools used to approximate the behavior of a system.

The phrase *model validation* also takes on different meanings to different researchers and there are basically two different schools of thought on this subject. There are those, such as Popper (1959) who suggest that we cannot validate a hypothesis, only invalidate it. Others, such as Kuhn (1970) take the view that models are accepted through scientific consensus. In the present work, we focus on a less philosophical aspect of model validation, the quantification of agreement between model prediction and experimental observation through the use of statistical tools. We suggest the development of these tools can be used not only to test models, but can also be used to understand the statistical significance of model validation experiments, used to speed and refine scientific consensus building, and used to improve the design of future validation experiments. Some of the authors who address broader issues of model validation are Davis et al. (1991), Tsang (1991), Konikow and Bredehoeft (1992), Neuman (1992), Rykiel (1996), Beck et al. (1997), Holzbecher (1997), Oberkampf et. al., (1998), Peterson and Kirchner (1998). Aeschliman and Oberkampf (1998) and Oberkampf and Blottner (1998) address issues related to computational fluid dynamics.

### 4.2 Measures

The first step in designing a model building or validation exercise is to define the quantities which will be compared. These quantities can be direct measurements, such as pressure or temperature, or quantities that require post experimental analysis, such as thermal capacity or damping coefficient. These measurements may be point quantities (relative to the scale of the model resolution), such as velocity or temperature, or can be integrated or averaged quantities, such as mean velocity, or total energy. When the time or spatial averaging scale of the measurements is important, the corresponding model predictions should be on the same scales. For example, if one measures water content in a heterogeneous unsaturated porous media using a neutron probe, the water content measurement is actually related to a weighted average of water content over the spatial volume associated with neutron penetration and scattering. The model prediction should be a corresponding weighted average of water content over this volume.

Once the comparison quantities are chosen, some measure of distance between the model predictions and the experimental observations is required. Historically, we have used both subjective and quantitative comparisons. Subjective comparisons are usually through visual inspection of x-y plots, scatter plots, and contour plots of model predictions and experimental observations. These comparisons have the advantage of showing the trend in agreement over space and time which is useful for model building. Unfortunately, visual comparisons are very subjective and dependent on the details of the contouring scheme, the domain and range of the plot axis, and other graphical details. A second approach is through various mathematical measures such as the correlation coefficient (basically a normalized sum of squares of residual as defined in the previous chapter) and other weighted and non-weighted norms. While these measures quantify a ‘distance’ between a set of measurements and predictions, one must still use subjective means to define what magnitude of these measures are acceptable. For example, in the previous chapter we defined our measure in terms of the 95% confidence probability surface. We could have just as easily chosen the 99% confidence surface, which would have increased range of acceptable residuals between model predictions and experimental observations.

### 4.3 Residuals

In an attempt to quantify model validity from a probabilistic perspective, researchers have proposed various statistical inference techniques using residuals between model predictions and experimental observations. Much of this work has been performed in the environmental sciences. Freese (1960) proposed that there are three elements that must be considered when developing new estimation or measuring techniques. These are

- “(a) A statement of the accuracy required, or the inaccuracy that will be tolerated.
- (b) A measure of the accuracy attained in the trials of the new technique.
- (c) A method of deciding whether the accuracy attained is equal to the accuracy required.”

Freese proposed a test to meet these three requirements where the test statistic is defined as follows:

$$c^2(n) = \frac{\sum_{i=1}^n (x_i - m_i)^2}{s^2} \quad (4.1)$$

$x_i$  is the value of the  $i^{\text{th}}$  observation,  $m_i$  is the “true” value of the  $i^{\text{th}}$  observation as measured by the standard (i.e., accepted) technique,  $n$  is the number of observations, and  $s^2$  is the required accuracy (i.e., variance) of the proposed new estimation or measurement technique. The  $c^2$  test is a statistical test for the ratio of the sum of squares

of differences (or variances) and is commonly used to test “goodness of fit” between model predictions and the data used to define the model (Miller and Freund, 1985).

Reynolds (1984) and Gregoire and Reynolds (1988) clarified the assumptions implicit in the  $\chi^2$  test proposed by Freese (1960). This test assumes the population of residuals between the observed and true values is normally distributed with a zero mean, a uniform variance, and also assumes the residuals are independent. Gregoire and Reynolds (1988) point out that independence may not exist for models which predict serial events. For example, a forest stand simulation program that provides successive growth projections may not result in independence among the residuals. Residuals from ordinary least squares models are also generally correlated.

Reynolds (1984) also notes that the  $c^2$  test proposed by Freese (1960) evaluates whether there is significant statistical evidence to reject a model, but does not establish whether a model is sufficiently accurate. Reynolds recast the null hypothesis to test whether there is significant statistical evidence that a model does accurately predict the experimental observations. This test is much stricter and there is significant probability that a good model will be rejected by this test. In a subsequent paper, Gregoire and Reynolds (1988) suggest that while the Freese test may be too liberal in the sense of not requiring reasonable predictive ability, the Reynolds procedure (1984) may be too conservative in that it requires stringent evidence. They propose an alternate procedure, which utilizes a three-way decision rule. If the model is accepted by the Reynolds (1984) test, then the model is declared accurate. If the model is rejected by the Reynolds test but accepted by the Freese test, the accuracy of the model is indeterminate. If the model is rejected by the Freese test, the model is declared inadequate.

As discussed in the tutorial of the present document, the model predictions, and thus the residuals are likely to be correlated and likely to possess a non-uniform variance. This is especially true for predictions that are sequential in nature, such as pressure or front location as a function of time. When this is the case, the methodology proposed by Freese (1960), Reynolds (1984), and Gregoire and Reynolds (1988) should not be applied. The concept of the three-way decision rule of Gregoire and Reynolds does hold some promise to serve as a classification tool of model predictive ability. Cases for which the accuracy of the model is indeterminate may suggest that we reconsider the design of the validation experiment.

Reckhow et al. (1990) discuss a series of statistical tests which they propose be used for model testing. These include the t-test for comparing means for normally distributed, independent random residuals with uniform variances, the Wilcoxon Test (also known as the Mann-Whitney test) for comparing centers of distributions given independent samples (does not require normality), regression analysis (discussed below), and the Kolmogorov-Smirnov test for comparing cumulative distribution functions (not necessarily normally distributed) given independent samples. The Kolmogorov-Smirnov test is often used to establish whether a distribution of residuals between model predictions and experimental observations are normally distributed. The Wilcoxon test is used when the probability

distribution is undefined. For example, Hills (1994b) utilized the Kolmogorov-Smirnov test and the Wilcoxon test to compare predictions from 14 models for water flow and 9 models for solute transport against field scale model validation experimental data. The Kolmogorov-Smirnov test was used to evaluate whether the residuals between model prediction and experimental observations were normally distributed. It was found that for some models, there was sufficient statistical evidence to reject the hypothesis that the residuals were normally distributed. Because of this, the Wilcoxon test was used rather than the t-test to evaluate whether the central tendency of the residuals was zero for each model.

#### **4.4 Regression**

Another approach that has received much attention in the environmental science literature is to perform linear regression on scatter plots of experimental observations and model predictions. The model predictions are usually plotted along the  $x$  axis with the experimental observations plotted along the  $y$  axis. If the model perfectly predicts the experimental data, all of the points in the scatter plot should lie on a line with an intercept of zero and a slope of one. Since model predictions and experimental observations contain error, there will be some scatter about this line. By performing linear regression on the resulting scatter plot, one can estimate a slope and intercept. Statistical tests are then performed to evaluate whether the resulting slope and intercept are significantly different from one and zero. This approach has been promoted by Reckhow et al. (1990). However, recent work by Mitchell (1997) and by Kleijnen et al. (1998) questions the application of this method. In a paper addressing strictly this issue, Mitchell lists several objections to the use of regression to evaluate model validity. One is the ambiguity of the null hypothesis tests. The more scatter in the data, the larger the standard error of the slope. Models with more scatter are more likely to have slopes not significantly different from unity and are more likely to not be rejected. Another objection is the violation of the statistical assumptions. In standard linear regression, it is assumed that there is no error in the values plotted along the  $x$  axis (usually the model axis). This is certainly not true if the model contains uncertainty in the model parameters. In addition, it is assumed that the errors associated with the  $y$  values (the experimental data axis) are independent and normally distributed, with a variance that is not a function of  $x$ . If the data are taken over time or space, these errors may be correlated and will often possess a variance which is not uniform.

#### **4.5 Error Bounds**

A third method, which occasionally is proposed, is to plot both model predictions and experimental observations as a function of important independent variables (Peterson and Kirchner, 1998). Error bands are included on both the model predictions and the experimental observations. Alternatively, error bands which include the effects of both model and measurement uncertainty can be shown on the model predictions as illustrated at the beginning of the previous chapter. By counting the number of points for which the

error bands do not overlap in the first case, or measurements which lie outside the error bands in the second case, one can presumably reach some judgment as to the validity of the model. For example, Rykiel (1996) suggests that a valid model output should fall within the 95% confidence intervals 75% of the time for the most important variables in dynamic models. Unfortunately, the joint probability of 75% of the results lying within the 95% confidence level is very dependent on the number of points, and the correlation between the points as was shown in the previous chapter. Such correlation and non-uniformity of variance should be accounted for in defining the joint probability of the differences between experimental measurement and model prediction, as was illustrated in the tutorial in Chapter 3. Otherwise, the statistical significance of the results is not clear.

#### **4.6 General Methods**

As the above discussion illustrates, the application of the above procedures to transient problems is problematic due to the invalidity of the statistical assumptions. We can expect the residuals between model predictions to be correlated and to possess a non-uniform variance. The residual distribution can also be non-normal. To account for these effects, we must utilize methodology which allows us to estimate the correlation structure and to incorporate this information into our statistical tests.

In contrast to many of the models used in environmental science that are derived largely through approximation (including time series), our models are often derived from basic principles. Our models are generally considered deterministic with uncertainty due mostly to model parameter uncertainty. In addition, our validation experiments can be more carefully controlled. As a result, it should be easier for us to quantify measurement uncertainty and to account for, if not eliminate, correlation between measurement errors. We should be able to model the correlation structure of the resulting model predictions and experimental measurements using propagation of uncertainty analysis. This information can then be used to formulate rigorous statistical model testing methodology that accounts for our knowledge of this more complex statistical structure. The tutorial in Chapters 2 and 3 is designed to serve as a foundation for such rigorous statistical model validation.

#### **4.7 Propagation of Uncertainty**

The propagation of uncertainty in the input parameters (rate constants, transport properties, mechanical properties) through a model to evaluate uncertainty in the output predictions has received much attention at Sandia National Laboratories. Because of this, this topic will only be lightly reviewed here. Extensive literature is available on analytical and numerical methods for stochastic modeling (Soong, 1973, Sabelfeld, 1991, Holden et al., 1996, Rockhold et al., 1996, Romero, 1998, Romero and Bankston, 1998). Analytical methods generally require models to be linear, or require the use of perturbation methods

such as the sensitivity method (Beck and Arnold, 1977, and Beck et. al., 1985) and the methods of Yeh et al. (1985a, b, c), Mantoglou and Gelhar (1987a, b, c), and Polmann (1990). Such methods can give useful information about the first few moments of the prediction probability distributions, such as their means and standard deviations, but are limited when dealing with highly nonlinear models away from these means. Some of these methods can, in theory, be extended to obtain higher moments of the prediction probability distributions, such as skewness and kurtosis, but this extension can be labor intensive.

Monte Carlo methods use multiple realizations of the input parameters, generated from the probability models for the input parameters. Each realization is used as model input to solve the forward problem using a standard numerical model for the physics. The predictions are accumulated to generate the probability distributions for the resulting predictions. These methods have the advantage that they fully account for the nonlinearity of underlying models, can be used to evaluate the correlation of model prediction error, and do not require output error distributions to be of a common type (i.e., normal, log-normal, Beta, etc.). Unfortunately, even with methods to reduce the number of realizations required, such as various stratified sampling schemes (Iman et. al. 1980, McKay et al., 1979), control variates, and importance sampling (Fishman, 1995, Vose 1996), Monte Carlo methods can be extremely CPU intensive. An alternative approach that shows promise is the use of a response surface to model the numerical model (Romero and Bankston, 1998). This approach utilizes the model to generate an approximated surface (splines, finite elements, etc.) of the model predictions as a function of the important parameters. Monte Carlo analysis is then performed using the response surface model rather than the original numerical model. This approach is more efficient if the cost of generating the response surface and of performing the Monte Carlo analysis using the response surface is less than that of performing Monte Carlo analysis using the original model. The response surface method can be very efficient when the number of model input parameters and the number of prediction variables are not too large.

#### ***4.8 Propagation of Uncertainty for Model Validation***

In the previous chapter, we discussed the use of the Propagation of Uncertainty Analysis for Model Validation. However, we are aware of very little work in this area. Luis and McLaughlin (1992) presented a statistical model validation approach that is applicable to transient transport problems for which a propagation of variance analysis has been performed. They propose several statistical tests to evaluate model validity. The first is to evaluate whether the mean of the residuals is significantly close to zero. They account for non-uniformity in variance across the measurement locations (or times), but assume the measurement locations are sufficiently far apart in space and time such that the residuals are uncorrelated. They assume that residuals from each measurement location are normally distributed. The second test is a mean square test analogous to that performed in Chapter 3. As in Chapter 3, a  $\chi^2$  test statistic is used to evaluate whether the joint probability of the residuals is within the 95% confidence region. The same assumptions

apply as those for the first test. The third test is for model structure and is for the case where measurements are close enough to each other so that there is spatial correlation. To remove correlation, they suggest using a spatial whitening filter (Schweepe, 1973) along one spatial direction (presumably the direction along which the measurements are most correlated). They suggest that the output of this filter should be an uncorrelated series of adjusted measurement residuals if the model is valid. Following Schweepe, they define a test statistic for the correlation of adjacent pairs of filtered measurements. If there is not significant statistical evidence of correlation, and if the model has passed the previous tests, then they consider that no significant evidence was provided to declare the model invalid. Luis and McLaughlin apply their methodology to the data from one of the earlier Las Cruces Trench Experiments (Wierenga, 1986). They use a perturbation approach based on the work of Yeh et al. (1985 a, b, c), Mantoglou and Gelhar (1987 a, b, c), and Polmann (1990) to evaluate the ensemble mean and variance of the predicted saturation levels given a two-dimensional, transient, highly nonlinear model for the movement of water through unsaturated soils. Their model accounts for spatial correlation in the saturated hydraulic conductivity and water retention model parameters. They found that their model passed the validation test for predictions at intermediate times, but failed the validation test for predictions at larger times. The model underpredicted the significant horizontal spreading of the water plume that was observed during the validation experiment.

A common feature of the above application is that the prediction uncertainty was assumed to be normally distributed. Luis and McLaughlin (1992) did allow for non-uniform variance but did not allow for correlation in the prediction residuals. As discussed in the previous chapter, the evaluation of probability levels or acceptance regions for non-normally distributed, correlated predictions is computationally intensive. Methods that allow one to estimate the resulting multivariate PDFs and transform these to normal PDFs are appropriate. An alternate approach may be the use of optimization, if one's goal is to test models rather than define complete PDFs for the uncertainty in model predictions.

#### **4.9 Optimization Methods**

Statistical model validation typically asks the question - are the model predictions consistent with the experimental observation? Statistical inference can be used to evaluate whether there is sufficient statistical evidence to reject a model based on the measurements. One approach to answer this question for highly nonlinear problems is to perform a full Monte Carlo simulation to define a multivariate confidence region for the measurements predicted by the model, and to evaluate whether the experimental observations lie within this region. This method was illustrated in the previous chapter. Unfortunately, the CPU requirements to adequately resolve such confidence regions are excessive when dealing with a large number of measurements. For example, if we have  $n$  validation measurements and we require only 3 model simulations along each side of the corresponding  $n$  dimensional hypercube to resolve the corresponding joint confidence region (see examples presented in previous chapter), we will require  $3^n$  model

simulations. If we have only  $n = 50$  measurements, we will require  $3^{50} \approx 7 \times 10^{23}$  model simulations. Clearly, we must utilize methods which do not require that we fully resolve the  $n$ -dimensional joint confidence region when  $n$  is large.

In model validation, we are generally interested in whether a given set of measurements are likely, given that our models for the physics and the measurement prediction uncertainty are correct. We do not care whether we resolve the whole  $n$ -dimensional confidence region, but must only resolve the value for the probability of a set of measurements given that our model (including our model for uncertainty) is correct. This suggests that the actual validation measurements can be used to somehow simplify our evaluation of the model. One possibility is to use optimization techniques.

Davis et. al. (1992) proposed such a technique in which optimization was used to find the parameters that resulted in the best fit of the model to the data in a simple least squares sense. If a model is valid, they assume that the least squares fit should account for all of the model structure and the residuals should be uncorrelated. If correlation is present, they declare the model invalid. They use a simple semivariogram to define the correlation structure of the residuals. To illustrate the methodology, Davis et. al. applied their technique to the validation of Darcy's law using the data from Darcy (1856) first experiment. The resulting semivariogram showed that the variance increased with lag indicating that the residuals were correlated. As a result, they declared Darcy's law invalid based on the data from this experiment. When their methodology was applied to a data set produced by Stearns (1927), Davis et. al. did not find evidence to declare Darcy's law invalid. In contrast to many of the problems of interest here, the data used to test Darcy's law were not time dependent and each data point was measured using an independent material sample. As a result, we would expect less correlation in the residuals than may be present using data from a single transient experiment.

While not applied to model validation, the inverse methods of Minke (1984) and the optimization method of Hill (1989) may have promise when extended to model validation. Minke's book on discrete inverse theory provides an overview of geophysical techniques for ill-posed inverse problems in the presence of significant uncertainty in model parameters and measurement error. The example at the end of the tutorial in Chapter 3 where the measurement and the model parameters were plotted on the same graph was an extension of some of the ideas presented in Minke (1985). Minke also discusses maximum likelihood estimates of both model parameters in the measurements themselves, given that the model is correct. Extension of this approach to the model validation may be more appropriate than the approach illustrated in Chapter 3 because it relaxes the assumption of using the measurements as estimates of the corresponding measurement population means. One problem with this approach is that it is based on measuring probability in the measurement-parameter space rather than in the prediction space. The mapping between the model parameter space and the prediction space is not necessarily one-to-one. A point in the parameter space that is located on a 95% PDF (probability density function) contour will not necessarily map to a point on the 95% PDF contour for the predicted variables. Hill (1989) presents a method by which one can move

from the confidence region in the model parameter space to an approximate confidence region in the prediction space. Hill states that predictions that lie on the  $(1-\alpha)\%$  confidence boundaries in the prediction space will be bounded by predictions that derive from the  $(1-\alpha)\%$  boundaries in the model parameter space if the sensitivity vector is continuous and at least one component of the sensitivity vector is non zero for all sets of model parameters in the  $(1-\alpha)\%$  joint confidence region for the model parameter space. By using optimization procedures to search along the boundaries of the parameter space confidence region, Hill finds the corresponding minimum and maximum values of the model predictions. Hill claims that the resulting minimum and maximum model predictions bound the  $(1-\alpha)\%$  confidence boundaries for the model predictions in the model prediction space. Such a technique allows one to define model prediction confidence bounds from the model parameter PDFs rather than the prediction PDFs. This reduces the CPU requirements significantly for complex models and may be a potential method for application to model validation.

## 5.0 Conclusions and Recommendations

The application of statistical model validation methodology to complex engineering and scientific models has received little attention in the open literature. Most of these efforts have been in the environmental science areas where uncertainty can be a dominant factor. The uncertainty associated with engineering models is generally far less and can often be quantified through carefully controlled experiments. However, interest in engineering and scientific model validation has changed over the past decade due to our increased ability to model complex systems, the decreased cost of computational simulation, and the increased cost of performing experiments. As a result, computer simulation is replacing experimental work in many areas. Because of these changes, the questions of model validity and model validation methodology are becoming more important.

Model validation for complex models, such as those associated with nonlinear partial differential equations, is not straightforward. This is due to the complex correlation structure and the non-uniformity of the variance, which are associated with the model predictions. Typical statistical model validation experiments, for which the experiment is repeated a sufficient number of independent times such that the statistical characteristics of the measurements can be resolved, are not practical for many of our applications. The experiments are often too expensive or simply too time consuming to repeat the required number of times. We must utilize the model itself to define what we think the structure of the uncertainty in the measurements is, given models for the uncertainty in the model parameters and measurements. This has the advantage over traditional statistical methods in that the resulting validation experiment will not only test our model, but will test our models for the uncertainty.

The difficulty of defining the models for the parameter and measurement uncertainty should not be understated. The examples presented in Chapters 2 and 3 assumed normal distributions for the uncertainty of these quantities. This is a common assumption and is often made when knowledge of the actual distributions is not available. However, such an assumption, without the benefit of supporting data, can lead to misleading results. This is especially true if the validation experimental measurements or the critical model predictions are located near the boundaries of the acceptance region. The location of these boundaries can be very sensitive to not only the distributions used, but to the values of the parameters (such as variance) that characterize these distributions. This problem is compounded by the difficulties associated with identifying the appropriate distributions from limited data. Research is needed to assess the impact of incorrectly identified probability distributions on uncertainty analysis and model validation. This research should include the investigation of non-parametric techniques and the use of “worst case distributions” when knowledge of the actual distributions is lacking. Fortunately, propagation of uncertainty analysis discussed in the previous chapters can still provide useful results, even though the uncertainty models for the model parameters are not adequately defined. The results of such an uncertainty analysis can be used to generate an

approximate rank-order of importance of the model parameters for an application, given our best guesses of the model parameter uncertainties. This information allows programmatic resources to be properly directed to adequately control or characterize the most influential model parameters.

While methodology to estimate prediction uncertainty using model parameter uncertainty is well established, the CPU requirements for many of these techniques can be prohibitive for applications that result in correlation structures in the predicted quantities. Research is needed to reduce these CPU requirements to manageable levels. Included in this research is the need to answer some basic questions about the process of model validation itself.

1. What are the disadvantages of performing model validation using model parameter based, rather than prediction based confidence levels, through the use of optimization methods?
2. What methods are available to establish the statistical significance of the agreement between model predictions and experimental observations without fully resolving the multi-dimensional joint probability density functions for the model predictions?
3. What are the appropriate norms to measure distance between predictions and measurements?
4. Can one define mean behavior norms which can be used for statistical inference for cases for which we are interested only in the mean behavior of a system rather than the detailed behavior of the system?
5. Traditional statistical model validation strategy uses statistical inference to invalidate a model. Can we develop strategies which allow us to validate a model without a significant chance of rejecting a valid model?
6. Some researchers propose to test model structure by characterizing the correlation structure of the residuals between model predictions and experimental observations. What is the significance of this and should we expect no correlation in the residuals of a valid model?

Recommendations for general areas of research are listed below:

1. The CPU requirements of a Monte Carlo analysis are very large for full physics models. Some methods, such as Latin hypercube sampling and importance sampling, exist to reduce the required number of Monte Carlo simulations, but the number of simulations is still excessive. Methods to reduce the number of Monte Carlo simulations required to adequately resolve the prediction uncertainty should continue to be developed. These methods are especially needed for the tails of the distributions which are the most difficult to resolve.

2. Sensitivity analysis based methods are much more efficient than Monte Carlo analysis, but are approximate for nonlinear models, may not give good estimates of the characteristics of the uncertainty in the predictions, and do not provide a complete description of the PDF for the prediction uncertainty. However, in applications for which the uncertainty in the model parameters are small, such as in many carefully designed and built engineered systems, sensitivity based methods can give good results even though the models are nonlinear. Sensitivity based and other methods should continue to be developed.
3. Model parameters are often characterized using various analysis techniques such as nonlinear least squares. This often results in the model parameter probability density functions being poorly approximated by common probability distributions. Methodology should be developed that utilizes the raw data directly to develop these distributions, and then uses these distributions as input for the uncertainty analysis.
4. Many distributions are not normally distributed, but are near normally distributed. Statistical testing techniques based on near-normal distributions should be investigated to evaluate whether these techniques can be used to reduce the CPU requirements for model validation and the propagation of uncertainty analysis.
5. While techniques for the design of experiments to maximize the statistical significance of the results have existed for some time, they are not widely used. A tutorial on statistical experimental design would be useful to many readers. Such a tutorial should also address experimental design to maximize the ability to resolve differences between competing models.
6. Non Monte Carlo model validation methodology, such as the method presented in previous chapters, should be explored further. These methods use optimization techniques, which require significantly fewer model evaluations than do the Monte Carlo techniques.

Finally, we would like to say a few words about the ultimate application of a model and its relationship to the validation experiments. One of the primary focuses of the present work was the development of methodology to test the statistical significance of differences between a set of measurements and model predictions. The techniques presented provide a yes or no answer to the question “does significant statistical evidence exist to declare a model invalid or not useful, based on the differences between the model predictions and experimental observations”. A byproduct of this methodology is a statement of the level of statistical significance of disagreement between the validation experimental results and the model predictions. The ultimate issue, however, is not whether a model is consistent with experimental observations from a validation experiment to some level of confidence, but what level of confidence will we have in the predictions when the model is applied to a specific physical system. This is not an easy question to answer since individual validation experiments are typically designed to test subsets of the physics rather than the full physics anticipated for the ultimate application.

How do we propagate and couple the results from a series of validation experiments to a statement about confidence in the analysis for the ultimate application? Research is needed along these lines.

## 6.0 References

- Aeschliman, D. P., and W. L. Oberkampf, 1998, *Experimental Methodology for Computational Fluid Dynamics Code Validation*, **AIAA Journal**, Vol. 36, No. 5, pp. 733-741.
- Beck, J. V. and K. J. Arnold, 1977, **Parameter Estimation in Engineering and Science**, John Wiley & Sons, New York.
- Beck, J. V., B. Blackwell, and C. R. St. Clair, 1985, **Inverse Heat Conduction: Ill Posed Problems**, John Wiley & Sons, New York.,
- Beck, M. B., J. R. Ravetz, L. A. Mulkey, and T. O. Barnwell, 1997, *On the Problem of Model Validation for Predictive Assessment*, **Stochastic Hydrology and Hydraulics**, Vol. 11, pp. 229-254.
- Brownlee, K. A., 1965, **Statistical Theory and Methodology: In Science and Engineering**, John Wiley & Sons, New York.
- Darcy, H., 1856, *Determination of the Laws of Flow of Water through Sand*, **Physical Hydrogeology**, (Eds. R. A. Freeze and W. Back, 1983), Hutchinson Ross Publishing Co. Stoudsberg, Pennsylvania.
- Davis, P. A., N. E. Olague, M. T. Goodrich, 1991, *Approaches for the Validation of Models Used for Performance Assessment of High-Level Nuclear Waste Repositories*, NUREG/CR-5533 (also SAND90-0575), U.S. Nuclear Regulatory Commission, Washington, DC.
- Davis, P. A., N. E. Olague, M. T. Goodrich, 1992, *Application of a Validation Strategy to Darcy's Experiment*, **Advances in Water Resources AWREDI**, Vol. 15, No. 3, pp. 175-180.
- Fishman, G. S., 1995, **Monte Carlo: Concepts, Algorithms, and Applications**, Springer, New York.
- Freese, F., 1960, *Testing Accuracy*, **Forest Science**, Vol. 6, No. 2, pp. 139-145.
- Gregoire, T. G., and M. R. Reynolds, 1988, *Accuracy Testing and Estimation Alternatives*, **Forest Science**, Vol.. 34, No. 2, pp. 302-320.
- Hill, M. C., 1989, *Analysis of Accuracy of Approximate, Simultaneous, Nonlinear Confidence Intervals on Hydraulic Heads in Analytical and Numerical Test Cases*, **Water Resources Research**, Vol. 25, no. 2, pp. 177-190.

Hills, R. G., K. A. Fisher, M. R. Kirkland, and P. J. Wierenga, 1994a, *Application of Flux-Corrected Transport to the Las Cruces Trench Site*, **Water Resources Research**, Vol. 30, No. 8, pp. 2377-2385.

Hills, R. G., and P. J. Wierenga, 1994b, **INTRAVALE Phase II: Model Testing at the Las Cruces Trench Site**, with contributions from S. Luis, D. McLaughlin, M. Rockhold, J. Xiang, B. Scanlon, and G. Wittmeyer, NUREG/CR-6063, U.S. Nuclear Regulatory Commission, Washington DC.

Holden, H., Øksendal, B., Ubøe, J., and Zhang, T., 1996, **Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach**, Birkhäuser, Boston.

Holzbecher, E., 1997, *Remarks of the Paper by E. J. Rykiel, entitled: 'Testing Ecological Models: The meaning of Validation (Ecol. Model. 90, 229-244, 1996)*, **Ecological Modelling**, Vol. 102, pp. 375-377.

Iman, R. L., and M. J. Shortencarier, 1984, **A FORTRAN 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use With Computer Models**, SAND83-2365, March.

Kleijnen, J. P., B. Bettonvil, and W. Groenendaal, 1998, *Validation of Trace-Driven Simulator Models: A Novel Regression Test*, **Management Science**, Vol. 44, No. 6.

Konikow, L. F., and J. D. Bredehoeft, 1992, *Ground-Water Models cannot be Validated*, **Advances in Water Resources**, Vol. 15, pp. 75-83.

Kuhn, T.S., 1970, **The Structure of Scientific Revolutions**, University of Chicago Press, Chicago Illinois.

Luis, S. J., and D. McLaughlin, 1992, *A Stochastic Approach to Model Validation*, **Advances in Water Resources**, Vol. 15, pp. 15-32.

Mantoglou, A., and L. W. Gelhar, 1987a, *Large Scale Models of Transient Unsaturated Flow Systems*, **Water Resources Research**, Vol. 23, No. 1, pp. 37-46.

Mantoglou, A., and L. W. Gelhar, 1987b, *Capillary Tension Head Variance, Mean Soil Moisture Content, and Effective Specific Soil Moisture Capacity of Transient Unsaturated Flow in Stratified Soils*, **Water Resources Research**, Vol. 23, No. 1, pp. 47-56.

Mantoglou, A., and L. W. Gelhar, 1987c, *Effective Hydraulic Conductivities of Transient Unsaturated Flow in Stratified Soils*, **Water Resources Research**, Vol. 23, No. 1, pp. 57-686.

- McKay, M. D., W. J. Conover., and R. J. Beckman, 1979, *A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code*, **Technometrics**, Vol. 21, pp. 239-245.
- Miller, I. and J. Freund, 1985, **Probability and Statistics for Engineers**, Prentice-Hall, Inc. Englewood cliffs, NJ, p265.
- Minke, W., 1984, **Geophysical Data Analysis: Discrete Inverse Theory**, Academic Press, New York.
- Mitchell, P. L., 1997, *Misuse of Regression for Empirical Validation Models*, **Agricultural Systems**, Vol. 54. No. 3, pp. 313-326.
- Neuman, S. P., 1992, *Validation of Safety Assessment Models as a Process of Scientific and Public Confidence Building*, Presented at the High Level Radioactive Waste Management Conference, Las Vegas, Nevada, April 12-16.
- Oberkampf, W. L. and F. G. Blottner, 1998, *Issues in Computational Fluid Dynamics Code Verification and Validation*, **AIAA Journal**, Vol. 36, No. 5, pp. 687-695.
- Oberkampf, W. L., K. V. Diegert, K. F. Alvin, and B. M. Rutherford, 1998, *Variability, Uncertainty, and Error in Computational Simulation*, **ASME Proceedings of the 7<sup>th</sup> AIAA/ASME Joint Thermophysics and Heat Transfer Conference**, HTD-Vol. 357-2, Book No. H1137B.
- Peterson, S. R., and T. B. Kirchner, 1998, *Data Quality and Validation of Radiological Assessment Models*, **Health Physics**, Vol. 74, No. 2.
- Polmann, D. J., 1990, **Application of Stochastic Methods to Transient Flow and Transport in Heterogeneous Unsaturated Soils**, PhD thesis, Department of Civil Engineering, M.I.T., Cambridge MA.
- Popper, K. R, 1959., **The Logic of Scientific Discovery**, Harper and Row, New York.
- Reckhow, K. H., J. T. Clements, and R. C. Dodd, 1990, *Statistical Evaluation of Mechanistic Water Quality Models*, **Journal of Environmental Engineering**, Vol. 116, No. 2, pp. 250-268.
- Reynolds, M. R., Jr., 1984, *Estimating the Error in Model Predictions*, **Forest Science**, Vol.. 30, No. 2, pp. 454-469.
- Rockhold, M.L., R.E. Rossi, and R.G. Hills, 1996, *Application of Similar Media Scaling and Conditional Simulation for Modeling Water Flow and Tritium Transport at the Las Cruces Trench Site*, **Water Resources Research**, Vol. 32, No. 3, pp.595-609.

Roe, P. L., 1985, *Some Contributions to the Modelling of Discontinuous Flows*, **Lectures in Applied Mathematics**, Vol. 22, pp. 163-193.

Roe, P. L., 1986, *Characteristic-Based Schemes for the Euler Equations*, **Annual Review of Fluid Mechanics**, Vol. 18, pp. 337-365.

Romero, V. J., 1998, *Optimization and Nondeterministic Analysis with Large Simulation Models: Issues and Directions*, presented at the 7<sup>th</sup> AIAA/ASME Joint Thermophysics and Heat Transfer Conference, June 15-18, Albuquerque, NM.

Romero, V. J., and S. D. Bankston, 1998, **Finite-Element/Progressive-Lattice-Sampling Response Surface Methodology and Application to Benchmark Probability Quantification Problems**, Sandia Report SAND98-0567, March.

Rykiel, E. J., 1996, *Testing Ecological Models: The meaning of Validation*, **Ecological Modelling**, Vol. 90, pp. 229-244.

Sabelfeld, K.K., 1991, **Monte Carlo Methods in Boundary Value Problems**, Springer-Verlag, New York.

Schweepe, F. C., 1973, **Uncertain Dynamic Systems**, Prentice Hall, New Jersey.

Soong, T. T., 1973, **Random Differential Equations in Science and Engineering**, Academic Press, New York.

Stearns, N. D., 1927, *Laboratory Tests on Physical Properties of Water-Bearing Materials*, **U. S. Geological Survey, Water Supply Paper 596**, pp. 144-159.

Sweby, P. K., 1984, *High Resolution Schemes using Flux Limiters for Hyperbolic Conservation Laws*, **SIAM Journal on Numerical Analysis**, Vol. 21, No. 5, pp. 995-1011.

Tsang, Chin-Fu, 1991, *The Modeling Process and Model Validation*, **Ground Water**, Vol. 29, No. 6.

Vose, D., 1996, **Quantitative Risk Analysis: A Guide to Monte Carlo Simulation Modelling**, John Wiley & Sons, New York.

Wierenga, P. J., L. W. Gelhar, C. S. Simmons, G. W. Gee, T. J. Nicholson, 1986, **Validation of Stochastic Flow and Transport Models for Unsaturated Soils: A Comprehensive Field Study**, NUREG/CR-4622, U. S. Nuclear Regulatory Commission, Washington DC.

Yang, H. Q., and A. J. Przekwas, 1992, A Comparative Study of Advanced Shock-Capturing Schemes Applied to Burgers' Equation, *Journal of Computational Physics*, Vol. 102, pp. 139-159.

Yeh, T. C., L. W. Gelhar, and A. L. Gutjahr, 1985a, *Stochastic Analysis of Unsaturated Flow in Heterogeneous Soils 1. Statistically Isotropic Media*, **Water Resources Research**, Vol. 21, No. 4, pp. 447-456.

Yeh, T. C., L. W. Gelhar, and A. L. Gutjahr, 1985b, *Stochastic Analysis of Unsaturated Flow in Heterogeneous Soils 2. Statistically Anisotropic Media with Variable  $\alpha$* , **Water Resources Research**, Vol. 21, No. 4, pp. 457-464.

Yeh, T. C., L. W. Gelhar, and A. L. Gutjahr, 1985c, *Stochastic Analysis of Unsaturated Flow in Heterogeneous Soils 3. Observations and Applications*, **Water Resources Research**, Vol. 21, No. 4, pp. 465-471.

## Distribution

### EXTERNAL DISTRIBUTION

Adams, M. A.

Jet Propulsion Laboratory  
4800 Oak Grove Drive, MS 97  
Pasadena, CA 91109

Anderson, Charles E.

Southwest Research Institute  
P.O. Drawer 28510  
San Antonio, TX 78284

Anderson, D.

Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 205-45  
Pasadena, CA 91125

Aivasis, M.

Center for Advanced Computing  
Research  
California Institute of Technology  
1200 E. California Blvd./MS 158-79  
Pasadena, CA 91125

Ayyub, B. M

University of Maryland  
Civil Engineering – Rm. 1155  
College Park, MD 20742-3021

Beissel, S.

Alliant Techsystems Inc.  
600 Second St. NE  
Hopkins, MN 55343

Belytschko, T.

Northwestern University  
Dept. of Mechanical Engineering  
2145 Sheridan Road  
Evanston, IL 60108-3111

Casteel, K.

University of Texas – El Paso  
600 W. University Ave  
El Paso, TX 79968-0521

Cho, K.

Stanford University  
Durand 267  
Stanford, CA 94305

Colemans, H. W.

U. Alabama Huntsville  
S236 Technology Hall  
Huntsville, AL 35899

Coker, D.

Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Cosner, R. R.

Senior Technical Fellow  
Boeing – Phantom Work  
P. O. Box 516, MS: S106 - 7126  
St. Louis, MO 63166

Cuniff, P.

U. S. Army Soldier Systems Center  
Kansas Street  
Natick, MA 01750-5019

Glimm, J. G.

Dept. of Applied Math and Statistics  
Math, P138A  
State University of New York at Stony  
Brook  
Stony Brook, NY 11794-3600

Hills, R. G. (25)

Department of Mechanical Engineering  
New Mexico State University  
Las Cruces, New Mexico

Howes, F. A.  
U. S. Dept. of Energy  
DOE, MICS, ER-31  
19901 Germantown Rd.  
Germantown, MD 20874

Ivy, G.  
Logicon R&D Associates  
P. O. Box 92500  
Los Angeles, CA 90009

Kane, C.  
California Institute of Technology  
1200 E. California Blvd./MS 205-45  
Pasadena, CA 91125

Karniadakis, G.  
Division of Applied Mathematics  
Brown University  
192 George St., Box F  
Providence, RI 02912

Keremes, J.  
The Boeing Company  
Rocketdyne Propulsion & Power  
6633 Canoga Avenue  
Canoga Park, CA 91309-7922

Kimsey, K. D.  
U. S. Army Research Laboratory  
Weapons and Materials Research  
Directorate  
AMSRL-WM-TC 309 120A  
Aberdeen Proving Ground, MD  
21005-5066

Kovac, B. A.  
The Boeing Company  
Rocketdyne Propulsion and Power  
6633 Canoga Avenue  
Canoga Park, CA 91309-7922

Krysl, P.  
Department of Computer Science  
California Institute of Technology  
1200 E. California Blvd./MS 256-80  
Pasadena, CA 91125

Liu, W. K.  
Northwestern University  
Dept. of Mechanical Engineering  
2145 Sheridan Road  
Evanston, IL 60108-3111

Mair, H. U.  
Institute of Defense Analyses  
Operational Evaluation Division  
1801 North Beauregard Street  
Alexandria, VA 22311-1772

McDonald, W.  
Naval Surface Warfare Center  
Code 420  
101 Strauss Avenue  
Indian Head, MD 20640-5035

Mehta, U. B.  
NASA Ames Research Center  
MS: T27 B-1  
Moffett Field, CA 94035-1000

Molinari, J.-F.  
Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Morris, M. D.  
Oak Ridge National Lab  
P. O. Box 2008  
Building 6012  
Oak Ridge, TN 37831-6367

Namburu, R.  
CEWES-SD-R, Bldg. 5014  
USAE Waterways Experiment Station  
3909 Halls Ferry Road  
Vicksburg, MS 39180

Needham, C.  
Applied Research Associates, Inc.  
4300 San Mateo Blvd. Ste A-220  
Albuquerque, NM 87110

Needleman, A.  
Brown University  
Division of Engineering, Box D  
Providence, RI 02912

Omprakash, S.  
Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Ortiz, M. (5)  
Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Pace, D. K.  
The John Hopkins University  
Applied Physics Laboratory  
111000 Johns Hopkins Road  
Laurel, MD 20723-6099

Papoulia, K.  
Inst. Eng. Seismology & Earthquake  
Engineering  
P. O. Box 53, Finikas GR-55105  
Thessaloniki, Greece

Radovitzky, P.  
Graduate Aeronautical Laboratories

California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Rafaniello, W.  
DOW Chemical Company  
1776 Building  
Midland, MI 48674

Ran, H.  
Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Repetto, E.  
Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Rosakis, A. J.  
Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Sevin, E.  
Logicon RDA, Inc.  
1782 Kenton Circle  
Lyndhurst, OH 44124

Shu, Y.-C.  
Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Singleton, R.  
Director, Engineering Sciences  
Directorate  
U. S. Army Research Office  
4300 S. Miami Blvd.  
P. O. Box 1221  
Research Triangle Park, NC 27709-2211

Snowden, W. E.  
DARPA  
7120 Laketree Drive  
Fairfax Station, VA 22039

Stevenson, D. E. (Steve)  
Clemson University  
Computer Science Department  
442 Edwards Hall – Box 341906  
Clemson, SC 29631-1906

Sundaram, S.  
Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Thoutireddy, P.  
Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Voelkl, T.  
Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Xu, L.  
Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Yu, C.  
Graduate Aeronautical Laboratories  
California Institute of Technology  
1200 E. California Blvd./MS 105-50  
Pasadena, CA 91125

Zikry, M. A.  
North Carolina State University  
Mechanical & Aerospace Engineering

2412 Broughton Hall, Box 7910  
Raleigh, NC 27695

Zwissler, J. G.  
Jet Propulsion Laboratory  
4800 Oak Grove Drive – MS 97-8  
Pasadena, CA 91109-8099

Los Alamos National Laboratory  
Mail Station 5000  
P.O. Box 1663  
Los Alamos, NM 87545  
Attn: D. Cagliostro, MS F645  
Attn: D. L. Crane, MS P946  
Attn: J. F. Davis, MS P234  
Attn: K. M. Hanson, MS P940  
Attn: R. Henninger, MS D413  
Attn: B. L. Holian, MS B268  
Attn: K. S. Holian, MS D413  
Attn: D. Holm, MS B284  
Attn: J. M. Hyman, MS B284  
Attn: M. E. Jones, MS B259  
Attn: J. R. Kamm, MS D413  
Attn: L. G. Margolin, MS D413  
Attn: M. McKay, MS F600  
Attn: M. R. Miller, MS D471  
Attn: W. J. Rider, MS D413  
Attn: D. Sharp, MS B213  
Attn: R. N. Silver, MS B262  
Attn: D. Weeks, MS B295  
Attn: M. C. White

University of California  
Lawrence Livermore National  
Laboratory  
7000 East Ave.  
P.O. Box 808  
Livermore, CA 94550  
Attn: T. F. Adams, MS-L098  
Attn: S. F. Ashby, MS-L561  
Attn: D. L. Brown, MS-L561

Attn: P. N. Brown, MS L-561  
Attn: R. Klein, MS L-023  
Attn: S. Lee, MS-L560  
Attn: C. F. MacMillan, MS L-098  
Attn: C. Mailhiot, MS L-055  
Attn: J. F. Mcenerney, MS L-591  
Attn: D. Nikkel, MS L-342  
Attn: P. Raboin, MS L-125

MS 0316 P. F. Chavez, 9204  
MS 0316 E. D. Gorham, 9209  
MS 0318 G. S. Davidson, 9215  
MS 1109 C. F. Diegert, 9215  
MS 0318 P. Heermann, 9215  
MS 1111 S. S. Dosanjh, 9221  
MS 1111 S. Plimpton, 9221  
MS 1111 A. Salinger, 9221  
MS 1111 J. N. Shadid, 9221  
MS 1111 R. Schmidt, 9221  
MS 1111 T. M. Smith, 9221  
MS 1110 D. Womble, 9222  
MS 1110 W. Hart, 9222  
MS 1110 S. Istrail, 9222  
MS 1110 R. Lehoucq, 9222  
MS 1110 S. Minkoff, 9222  
MS 1110 L. A. Romero, 9222  
MS 0321 A. L. Hale, 9224  
MS 0321 J. Ang, 9224  
MS 0321 R. Benner, 9224  
MS 0321 G. M. Pollock, 9224  
MS 0321 J. L. Tompkins, 9224  
MS 1111 G. Heffelfinger, 9225  
MS 1111 R. M. Fye, 9225  
MS 1111 H. Hjalmarson, 9225  
MS 1111 M. D. Rintoul, 9225  
MS 1111 M. Stevens, 9225  
MS 0441 R. Leland, 9226  
MS 1111 B. Hendrickson, 9226  
MS 0819 J. S. Peery, 9231  
MS 0819 E. Boucheron, 9231  
MS 0819 K. G. Budge, 9231  
MS 0819 D. Carroll, 9231  
MS 0819 R. Drake, 9231  
MS 0819 M. Christon, 9231  
MS 0819 E. S. Hertel, 9231  
MS 0819 A. C. Robinson, 9231  
MS 0819 R. Summers, 9231  
MS 0819 T. G. Trucano, 9231 (40)  
MS 0819 R. Weatherby, 9231  
MS 0819 M. Wong, 9231  
MS 0820 P. Yarrington, 9232  
MS 0820 M. Boslough, 9232  
MS 0820 R. Brannon, 9232  
MS 0820 D. Crawford, 9232

#### SANDIA INTERNAL

MS 1202 J. M. McGlaun, 5903  
MS 0746 D. G. Robinson, 5411  
MS 0778 G. Barr, 6851  
MS 9403 M. Baskes, 8712  
MS 9405 R. Jones, 8742  
MS 9405 P. Klein, 8742  
MS 9405 R. A. Regueiro, 8743  
MS 9011 J. C. Meza, 8950  
MS 9011 P. T. Boggs, 8950  
MS 9011 M. L. Koszykowski, 8950  
MS 1110 L. J. Lehoucq, 8950  
MS 9011 C. H. Tong, 8950  
MS 0841 P. Hommert, 9100  
MS 0828 R. K. Thomas, 9104  
MS 0828 J. Moya, 9105  
MS 0828 J. Garcia, 9106  
MS 0834 A. Ratzel, 9112  
MS 0834 M. R. Baer, 9112  
MS 0834 W. M. Trott, 9112  
MS 0835 S. N. Kempka, 9113  
MS 0835 B. F. Blackwell, 9113  
MS 0835 V. J. Romero, 9113  
MS 0835 T. E. Voth, 9113  
MS 0825 W. H. Rutledge, 9115  
MS 0825 R. Baty, 9115  
MS 0825 F. G. Blottner, 9115  
MS 0825 A. R. Lopez, 9115  
MS 0825 M. McWherter-Payne, 9115  
MS 0825 W. L. Oberkampf, 9115  
MS 0825 K. Salari, 9115  
MS 0557 T. Paez, 9119  
MS 0836 J. Stewart, 9121  
MS 0321 W. J. Camp, 9200

MS 0820 E. Fang, 9232  
MS 0820 A. Farnsworth, 9232  
MS 0820 M. E. Kipp, 9232  
MS 0820 S. A. Silling, 9232  
MS 0820 P. A. Taylor, 9232  
MS 0439 D. R. Martinez, 9234  
MS 0439 K. Alvin, 9234  
MS 0439 J. Dohner, 9234  
MS 0439 M. Eldred, 9234  
MS 0439 J. Red-Horse, 9234  
MS 0439 G. Reese, 9234  
MS 0439 H. P. Walter, 9234  
MS 1221 J. S. Rottler, 9800

MS 0419 R. G. Easterling, 9800  
MS 0829 K. V. Diegert, 12323  
MS 0829 M. Abate, 12323  
MS 0829 B. M. Rutherford, 12323  
MS 0337 R. D. Skocypec, 15002  
MS 1179 J. R. Lee, 15340  
MS 0899 Technical Library, 4916 (2)  
MS 0619 Review and Approval Desk,  
00111 (1) for DOE/OSTI  
MS 9018 Central Technical Files,  
8940-2