Uncertainty quantification in simple linear and non-linear problems

By T. Chantrasmi†, P. Constantine, N. Etemadi‡, G. Iaccarino AND Q. Wang

1. Motivation and objectives

Despite the considerable success of computer simulation technology in science and engineering, it remains difficult to provide objective confidence measures of the numerical predictions. This difficulty arises from the uncertainties associated with the inputs of any computation attempting to model a physical system. Uncertainties are typically classified as *aleatory* and *epistemic*. Aleatory uncertainty (also called variability) arises naturally from randomness in the system, and is studied using probabilistic approaches. The determination of material properties or operating conditions of a physical system typically leads to aleatory uncertainties; additional experimental characterization of such quantities might provide more conclusive evidence and characterization of their variability, but in practical situations it is not possible reduce this type of uncertainty completely. On the other hand, epistemic uncertainty is typically due to incomplete knowledge (or ignorance). This can arise from assumptions introduced in the derivation of the mathematical model or simplifications related to the correlation or dependence between physical processes. It is possible to *reduce* the epistemic uncertainty by using, for example, a combination of calibration, inference from experimental observations, and improvement of the physical models.

In this report, we consider two test problems to explore various numerical techniques to identify, propagate, and quantify uncertainties. The first problem involves transient heat conduction (a problem governed by a linear PDE) and corresponds to a test problem proposed by Sandia National Laboratory (Dowding *et al.* 2006). In this case, both aleatory and epistemic uncertainties are considered. The second problem is related to transient non-linear advection (Burgers' equation), and only variability in the initial conditions is considered.

2. Uncertainty analysis for a linear problem: heat conduction

The design and analysis of complex engineering systems is challenging not only because of the physical processes involved but also as a result of the limited amount of precise characterization of the overall conditions. This is in stark contrast to the assumptions in a computational model that typically requires well-defined (deterministic) inputs. For example, the material property of a certain component might be a function of the manufacturing process and exhibit a strong random variability. In a validation workshop organized by Sandia National Laboratory, three nominally simple problems were proposed to identify procedures and techniques to determine confidence bounds on the predictions. We describe the first of these problems in the following.

† Authors are listed in alphabetical order.
‡ University of Illinois at Chicago

Data set	Samples	k $[W/mC]$	ρC_p $\times 10^-5[J/m^3C]$
Low data	6	[0.0496:0.0796]	[3.38:4.52]
Medium data	20	[0.0455:0.0796]	[3.38:4.69]
High data	30	[0.0455:0.0811]	[3.37:4.69]

TABLE 1. Measured material properties.

2.1. *Description of the problem: the Sandia thermal challenge*

Consider a 1-D slab with a known heat flux applied at one end and an adiabatic boundary condition at the other. The thermal conductivity (k) and specific heat coefficient (ρC_p) of the slab material are known with a degree of variability; these are the only two sources of aleatory uncertainty. The objective is to determine the probability that after a certain amount of time (1000 seconds) the temperature in the slab is above a critical value ($T_c = 900^oC$); this is referred to as the regulatory assessment. The source of epistemic uncertainty is the choice of the numerical model to represent the physical system. Instead of considering the PDE describing the 1-D heat conduction in the slab, the challenge is built around an *exact* solution to the PDE, expressed as a series:

$$T(x,t) = T_{init} + \frac{qL}{k}\left[\frac{(k/\rho C_p)t}{L^2} + \frac{1}{3} - \frac{x}{L} + \frac{1}{2}\left(\frac{x}{L}\right)^2\right] - \frac{2qL}{\pi^2 k}\sum_{n=1}^{6}\frac{1}{n^2}\, e^{-n^2\pi^2\frac{(k/\rho C_p)t}{L^2}}\cos\left(n\pi\frac{x}{L}\right), \tag{2.1}$$

where $T(x,t)$ is the temperature within the slab, T_{init} the initial condition, L the slab length and q the heat flux applied at one end of the slab ($x = 0$). Note that both q and L are assumed known without uncertainty. Experimental data are provided for the material properties as a collection of samples. In addition, measured temperature distributions for five choices of q and L are given. These measurements are repeated few times and, therefore, several temperature dataset are available for the same nominal conditions (q and L), but without a reference to the material characterization experiments. In other words, the material properties of the slab used for the temperature measurements are unknown. Moreover, the conditions used in these experiments are different from the conditions specified for the regulatory assessment, and therefore, a direct validation of the predictions is not possible.

2.2. *Characterization of the aleatory uncertainty*

The first step of any uncertainty quantification (UQ) study is to characterize the sources of variability. Thirty measured values are provided for k and ρC_p. The samples are organized in three groups corresponding to *low*, *medium*, or *high* volume of data with the first set including only 6 samples, the other two including 20 and 30 samples, respectively. The objective is to identify the effect of a better characterization of the input parameters on the confidence bounds for the predictions. Note that the last set includes the first two sets plus 10 new samples.

Initially we analyzed the three data sets to verify their *consistency*. Using the three datasets sample, and assuming that k and ρC_p are random variables independent of

Data set	Samples	k $[W/mC]$	ρC_p $\times 10^-5[J/m^3C]$
Low data	6	[0.0127:0.1073]	[2.047:6.063]
Medium data	20	[0.0416:0.0833]	[3.153:4.892]
High data	30	[0.0423:0.0822]	[3.189:4.689]

TABLE 2. Material characterization for the thermal problem, assuming a Gaussian distribution of the variables within the intervals reported in Table 1

everything else, we can compute the 99.5% confidence interval from a student-t distribution. The obtained intervals are reported in Table 2. The addition of more samples allows to reduce the variation interval for both quantities: the three sets of samples are consistent.

2.3. *Predicting confidence intervals using direct sampling*

The next step is to determine the confidence intervals of the predicted temperature at $x = 0$ and $t = 1000sec$ from the formula (2.1). Noting that T decreases monotonicly with respect to ρC_p, we can calculate the confidence interval by evaluating Eq. (2.1) using the two ends of the 99.5% confidence interval of ρC_p, and sampling inside the 99.5% confidence interval of k. This produces a 99% confidence interval of temperature T. The predictions obtained in this way are compared to the experimental measurements for all the cases available. In Fig. 1 the results obtained for one of the configurations (corresponding to $L = 0.019m$ and $q = 3000W/m^2$) are reported. The experimental data fall within the predicted bounds, and moreover, as we include more information, the confidence interval tightens.

Applying the same procedure to the regulatory assessment conditions, we obtain the predictions reported in Fig. 1. Repeating the predictions for the three sets of material properties previously identified yields the following predicted values and confidence bounds: 1306 ± 682, 906.0 ± 220, and 900.5 ± 217.5 respectively. Note that although the confidence interval gets tighter as the material characterization data increase, the 99% confidence interval is not entirely below $900C$.

To identify the effect of the assumed stochastic distribution of the material properties, we also considered the case of k and ρC_p uniformly distributed within the minimum and maximum in the experimental samples (see Table 1. We evaluated Eq. (2.1) using the same statistical sampling discussed earlier and obtained the results reported in Fig. 2. The bounds are somewhat reduced in this case (due to the elimination of the tails in the assumed distribution of the material properties), but still the system has a large probability of achieving temperatures above the critical temperature indicated in the Sandia challenge.

Therefore, we cannot safely conclude that the system satisfies the regulatory requirement. More material properties' measurements might be required to improve the characterization of the distributions of k or ρC_p, or a different model must be developed to identify the epistemic uncertainty. In the following section, we consider the heat conduction PDE and apply a probabilistic approach to propagate the distributions of the material properties.

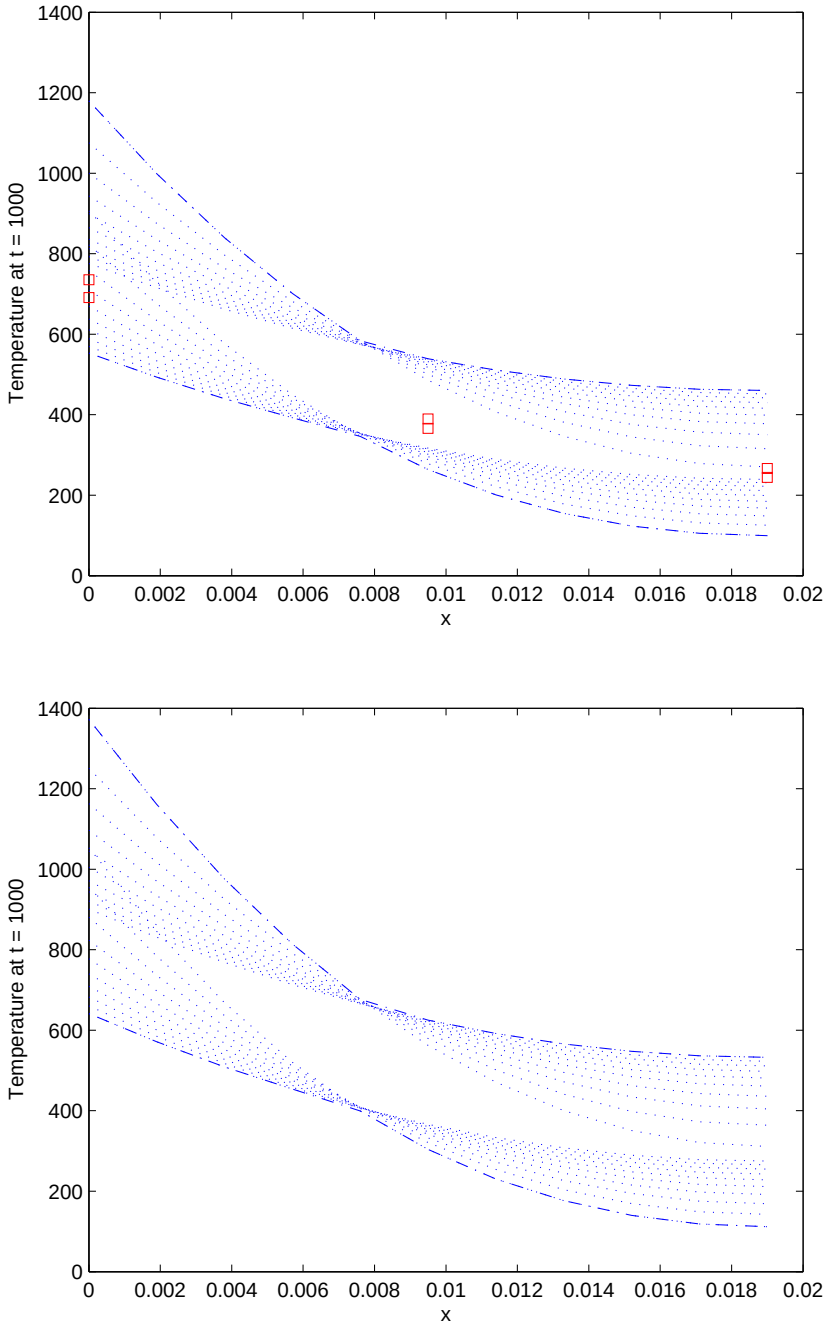

FIGURE 1. 99% confidence interval of T assuming Gaussian distribution for k and ρC_p. The conditions correspond to $L = 0.019m$ and $q = 3000W/m^2$ (top) and to the regulatory assessment $L = 0.019m$ and $q = 3500W/m^2$ (bottom). □ : experimental data - ·····: Eq. 2.1 - —·—·—: computed confidence bounds.

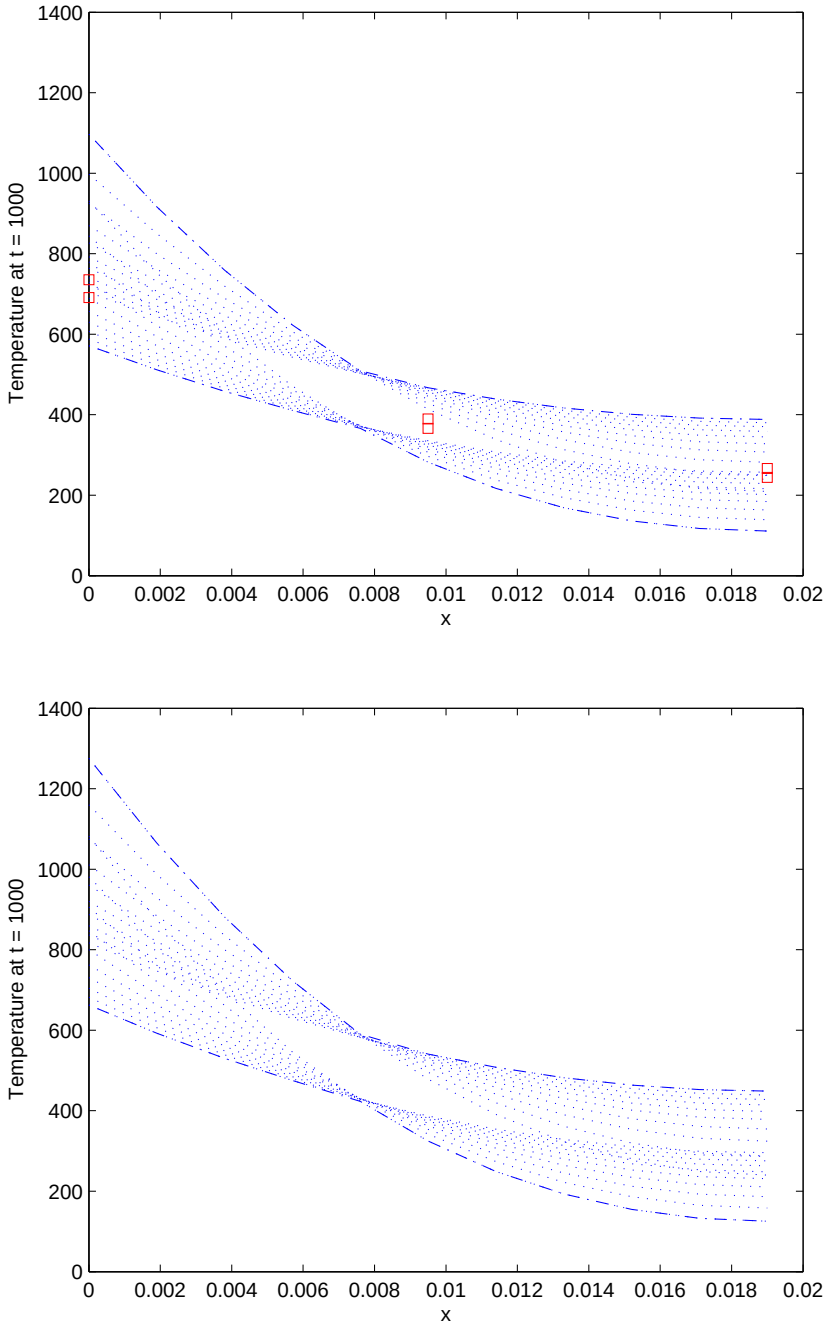

FIGURE 2. 99% confidence interval of T assuming uniform distribution for k and ρp_c. The conditions correspond to $L = 0.019m$ and $q = 3000W/m^2$ (top) and to the regulatory assessment $L = 0.019m$ and $q = 3500W/m^2$ (bottom). ▫ : experimental data - ·····: Eq. 2.1 - —·—·—: computed confidence bounds.

2.4. *Predicting probability bounds using a polynomial chaos approach*

In this section, we illustrate the application of the polynomial chaos expansion (PCE) to solve the present 1-D heat conduction problem. The PCE approach has its foundation in the work of Wiener (1938), who represented a Gaussian process as an infinite series of Hermite polynomials that take a vector of random variables as arguments. Ghanem and Spanos (1991) used this representation to develop the stochastic finite element method. Xiu and Karniadakis (2002) extended the theoretical framework to non-Gaussian process by employing different polynomial basis functions. This generalized polynomial chaos approach was used to address the problem of heat transfer with random material properties by Wan *et al.* 2004.

The PCE is a representation of a random variable, more generally a stochastic process, with an infinite series of orthogonal polynomials that take a vector of independent and identically distributed random variables as arguments. Denote the set of orthogonal polynomials by $\{\Phi_i(\boldsymbol{\xi}(\omega))\}$, $i \geq 0$, where $\boldsymbol{\xi}(\omega)$ is a vector of i.i.d. random variables. Then $\{\Phi_i(\boldsymbol{\xi}(\omega))\}$ have the following properties:

$$\langle \Phi_0 \rangle = 1, \quad \langle \Phi_i \rangle = 0 \ \text{ for } i > 0, \quad \langle \Phi_j \Phi_k \rangle = \delta_{jk} \ \text{ for } j, k \geq 0,$$

where $\langle \cdot \rangle$ is the expectation and δ_{jk} is the Kronecker delta. By the Cameron-Martin theorem (1947), the PCE converges in the L_2 sense. So, for example, if we represent $T = \sum_{i=0}^{\infty} T_i(x,t)\Phi_i(\boldsymbol{\xi}(\omega))$, where $\{T_i(x,t)\}$ are the PCE coefficients, then

$$\left\langle \left(T - \sum_{i=0}^{N} T_i \Phi_i(\boldsymbol{\xi}(\omega)) \right)^2 \right\rangle \to 0$$

as $N \to \infty$. The L_2 convergence of this expansion justifies an approximation of the random quantities by a truncated finite series. For the example above, we approximate

$$T \approx \sum_{i=0}^{P} T_i(x,t)\Phi_i(\boldsymbol{\xi}(\omega)),$$

where P is the order of the truncated PCE. The value of P is determined by the number d of random dimensions - the length of $\boldsymbol{\xi}(\omega)$ - and the highest degree n of polynomial employed. In particular, we have the formula

$$(P+1) = \frac{(n+d)!}{n!d!}.$$

In the heat conduction problem considered above, we assume that $\rho C_p(\omega_1)$ is a uniformly distributed random variable over the interval $[(\rho C_p)_a, (\rho C_p)_b]$ and let $k(\omega_2)$ be a uniformly distributed random variable over $[k_a, k_b]$ such that ρC_p and k are independent; we use the arguments ω_1 and ω_2 to emphasize that ρC_p and k are random quantities. Then we consider the following differential equation:

$$\rho C_p \frac{\partial T}{\partial t} = k \frac{\partial^2 T}{\partial x^2}, \tag{2.2}$$

where

$$k\frac{\partial T}{\partial x}|_{x=0} = q, \qquad k\frac{\partial T}{\partial x}|_{x=L} = 0, \qquad T(x,0) = T_{init}.$$

over the domain $x \in [0, L]$, $t \in [0, t_f]$, where T is in general a function of time, space, and the two random variables ω_1 and ω_2.

By choosing a uniform distribution to model k and ρC_p, we ensure that their support is bounded. If we had chosen some other distribution, say Gaussian, that has support in the negative real line, then we risk attempting to solve the backward heat equation, which is unstable.

Given (2.2), the objective is to compute the mean and variance of T, both of which are functions of x and t using the PCE approach. Since ρC_p and k are both uniform random variables, we can achieve exponential convergence of the PCE coefficients of T by choosing $\{\Phi_i(\xi_1(\omega_1), \xi_2(\omega_2))\}$ to be the 2-D Legendre polynomials with each ξ_i, $i = 1, 2$ distributed uniformly over the interval $[-1, 1]$, as described in Xiu and Karniadakis (2002). For reference, the first few 2-D Legendre polynomials are given by

$$\begin{aligned}
\Phi_0 &= 1 \\
\Phi_1 &= \xi_1(\omega_1) \\
\Phi_2 &= \xi_2(\omega_2) \\
\Phi_3 &= \frac{1}{2}(3\xi_1(\omega_1)^2 - 1) \\
\Phi_4 &= \xi_1(\omega_1)\xi_2(\omega_2) \\
\Phi_5 &= \frac{1}{2}(3\xi_2(\omega_2)^2 - 1).
\end{aligned} \tag{2.3}$$

Equipped with these basis polynomials, we now represent

$$\begin{aligned}
\rho C_p &\approx \sum_{i=0}^{P} (\rho C_p)_i \Phi_i(\xi_1(\omega_1), \xi_2(\omega_2)), \\
k &\approx \sum_{i=0}^{P} k_i \Phi_i(\xi_1(\omega_1), \xi_2(\omega_2)), \\
T &\approx \sum_{i=0}^{P} T_i \Phi_i(\xi_1(\omega_1), \xi_2(\omega_2)).
\end{aligned}$$

In general, since the distributions of ρC_p and k are known, we can compute their PCE coefficients by taking advantage of the orthogonality of $\{\Phi_i\}$. In particular, we have the following formulas:

$$(\rho C_p)_i = \frac{\langle \rho \Phi_i \rangle}{\langle \Phi_i^2 \rangle}, \qquad k_i = \frac{\langle k \Phi_i \rangle}{\langle \Phi_i^2 \rangle}.$$

In this simple case, however, we note that the PCEs of ρC_p and k are simply scalings of their respective ranges, i.e.,

$$\rho = \frac{\rho_b + \rho_a}{2} + \frac{\rho_b - \rho_a}{2}\xi_1(\omega_1), \qquad k = \frac{k_b + k_a}{2} + \frac{k_b - k_a}{2}\xi_2(\omega_2).$$

Then the PCE coeffcients are

$$\rho_0 = \frac{\rho_b + \rho_a}{2}, \qquad \rho_1 = \frac{\rho_b - \rho - a}{2}, \qquad \rho_2 = 0, \qquad \rho_i = 0,\ i > 2$$

and

$$k_0 = \frac{k_b + k_a}{2}, \qquad k_1 = 0, \qquad k_2 = \frac{k_b - k_a}{2}, \qquad k_i = 0,\ i > 2.$$

When we substitute the PCE representations into the differential equation, the problem

becomes

$$[(\rho C_p)_0 + (\rho C_p)_1 \xi_1(\omega_1)] \sum_{i=0}^{P} \frac{\partial T_i(x,t)}{\partial t} \Phi_i(\xi_1(\omega_1), \xi_2(\omega_2)) =$$

$$(k_0 + k_2 \xi_2(\omega_2)) \sum_{i=0}^{P} \frac{\partial^2 T_i(x,t)}{\partial x^2} \Phi_i(\xi_1(\omega_1), \xi_2(\omega_2)) \tag{2.4}$$

with boundary and initial conditions prescribed as:

$$(k_0 + k_2 \xi_2(\omega_2)) \sum_{i=0}^{P} \frac{\partial T_i}{\partial x}|_{x=0} \Phi_i(\xi_1(\omega_1), \xi_2(\omega_2)) = q \tag{2.5}$$

$$(k_0 + k_2 \xi_2(\omega_2)) \sum_{i=0}^{P} \frac{\partial T_i}{\partial x}|_{x=L} \Phi_i(\xi_1(\omega_1), \xi_2(\omega_2)) = 0 \tag{2.6}$$

$$\sum_{i=0}^{P} T_i(x,0) \Phi_i(\xi_1(\omega_1), \xi_2(\omega_2)) = T_{init}. \tag{2.7}$$

To solve this set of equations for the coefficients $T_i(x,t)$, we multiply each equation by Φ_k for $k = 0, \ldots, P$ and take the expectation of both sides. Then, by the orthogonality of $\{\Phi_i\}$, we're left with a system of linear differential equations for the PCE coefficients. By using second-order central differencing in space and the Crank-Nicholson scheme in time, we obtain a linear system to be solved at each time step. Once we have computed these coefficients, we can compute the expectation and variance of T at each grid point j and time instant n with the following formulas:

$$\bar{T}_j^n \equiv \langle T \rangle = (T_0)_j^n, \quad \hat{T}(_j^n) \equiv \langle (T - \langle T \rangle)^2 \rangle = \sum_{i=1}^{P} (T_i)_j^n \langle \Phi_i^2 \rangle.$$

To address the Sandia thermal challenge, we choose the material property intervals obtained from the *high* volume dataset, reported in Table 1. We then represent T with a truncated PCE in two random dimensions with second-degree orthogonal polynomials, so $P = (2+2)!/2!2! = 6$. We discretize the spatial domain with $N = 33$ grid points; the size of the linear system to be solved at each time step is then $(N \times P)^2 = 198^2$. There is no time-step restriction using the Crank-Nicolson scheme, so we choose $\Delta t = 10\Delta x$.

Although the present approach has been compared to all the experimental data provided, we only illustrate the results obtained for the configuration reported earlier, corresponding to $L = 0.019m$ and $q = 3000W/m^2$. As previously mentioned, there are two measured values of T at $x = 0$, $L/2$, and L, and t a multiple of 50 from 0 up to 1,000 seconds. We compare the results of the simulation at time 1000 seconds with the experimental data in Fig. 3.

The results show that the predicted results envelop the measured data, providing confidence that the material property variability is well characterized in these simulations. The application of the same numerical approach to the regulatory assessment is also illustrated in Fig. 3.

Once we have the PCE coefficients at $x = 0$ and $t = 1000$, we can generate samples from a uniform distribution over $[-1, 1]$ and evaluate the PCE representation of $T(0, 1000)$. The percentage of those realizations that are above T_c is approximately:

$$\mathrm{P}(T(0,1000) > T_c) \approx 0.2145$$

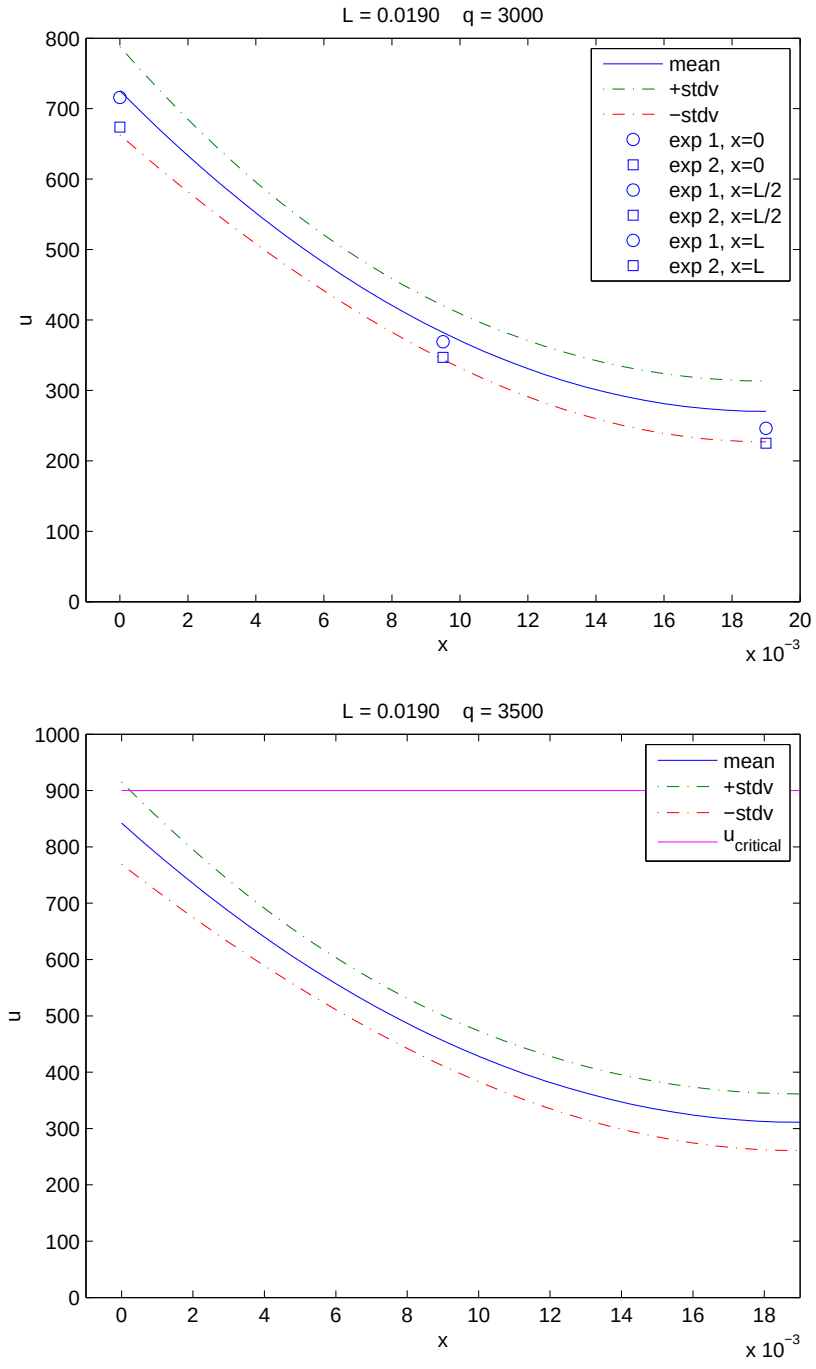

FIGURE 3. Polynomial chaos solution to the 1-D heat conduction problem assuming uniform distribution for k and ρC_p. The conditions correspond to the $L = 0.019m$ and $q = 3000W/m^2$ (top) and to the regulatory assessment $L = 0.019m$ and $q = 3500W/m^2$ (bottom). □ : experimental data - ·····: Eq. 2.1 - —·—·—: computed confidence bounds.

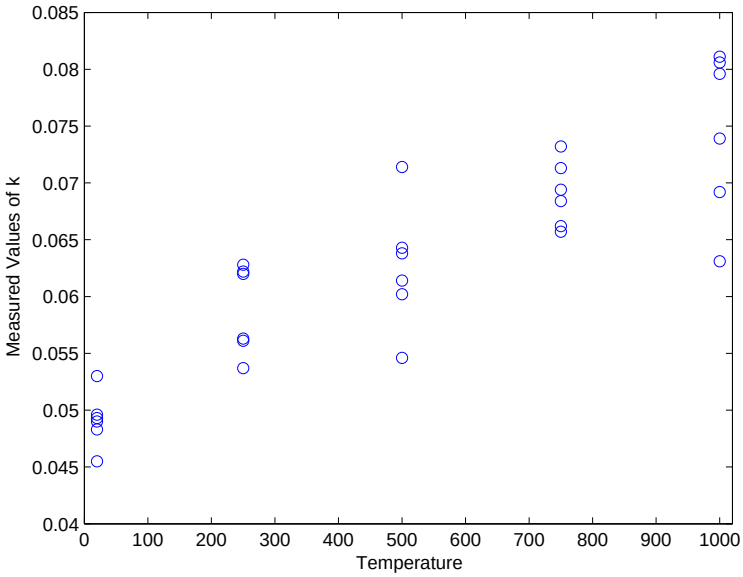

FIGURE 4. Measured thermal conductivity as a function of the experimental temperature

If we follow this approach, we conclude that the device does not meet the regulatory requirement described by the Sandia challenge problem. How confident are we in this assessment? There are, of course, the numerical errors resulting from the spatial and temporal discretization. The PCE representations of ρC_p and k are exact, but error associated with the truncation of the PCE of T is still an area of open research. We compared the PCE solution with a simple Monte Carlo simulation that generated realizations of ρC_p and k, and solved the resulting deterministic problem. We found that the PCE solution was sufficiently close to the Monte Carlo simulation to conclude that the two had converged. Currently this is the only available method for evaluating the accuracy of the truncated PCE.

None of these relatively small errors undermines our confidence in the assessment. The more troubling assumptions are the ones we used to model the uncertain parameters. Supposing that ρC_p and k are uniformly distributed, how do we know that they are actually distributed over the minimum and maximum of their measured values? Their respective ranges could be much larger, and this could affect the final computed probability. For example, running the simulation with a 20% increase in upper bound for k results in $\mathrm{P}(T(0,1000) > T_\mathrm{c}) \approx 0.1480$, a significant decrease from the result above.

In addition, the assumption that k is constant and independent of everything else might be incorrect. A more careful analysis of the conductivity data provided in the Sandia challenge problem is presented in Fig. 4 , where the samples are organized as a function of the experimental temperature. With linear regression, we find that there is *likely* a linear relationship between temperature and *conductivity*, i.e., $k(u) = \alpha + \beta T$. This is not a surprising physical behavior for many materials. Future work on this problem might involve modeling k as a linear function of T with random slope and/or intercept. This improved assumption should decrease $\mathrm{P}(T(0,1000) > T_{\mathrm{critical}})$, but the resulting non-linear equation is much more complicated.

3. Uncertainty propagation for a non linear problem: Burgers' equation

Uncertainty analysis in non-linear problems introduces additional complications. In this section, we apply the polynomial chaos approach introduced earlier to a simple non-linear Burgers' equation. We also introduce a hybrid approach based on the combination of deterministic modeling and stochastic sampling.

3.1. *Description of the problem*

We study the classical problem of non-linear advection in one dimension. The problem is governed by the Burgers equation:

$$u_t + \left(\frac{u^2}{2}\right)_x = 0 \qquad 0 \leq x \leq 1. \tag{3.1}$$

We will assume that the uncertainty is only related to the initial conditions, and in particular, consists of random perturbation of a sinusoidal wave, i.e.,

$$u(x,0) = sin(2\pi x) + \sigma\xi, \tag{3.2}$$

where ξ is a standard Gaussian random variable and σ is the magnitude of perturbation. The simplicity of this problem allows us to find an exact solution for the evolution of the mean and the variance of the solution. In the next section, the polynomial chaos framework is discussed, followed by a hybrid method.

3.2. *Polynomial chaos approach*

As previously illustrated, within the polynomial chaos framework we represent the solution of the Burgers' equation as

$$\mathbf{u}(x,t) = \sum_{i=0}^{\infty} u_i(x,t)\Phi_i(\xi), \tag{3.3}$$

where Φ_i is the ith-order Hermite polynomial (as mentioned earlier, this choice is motivated by the distribution of the random variable)

$$\Phi_0(\xi) = 1 \tag{3.4}$$

$$\Phi_1(\xi) = 2\xi \tag{3.5}$$

$$\Phi_{n+1}(\xi) = 2\xi\Phi_n(\xi) - 2n\Phi_{n-1}(\xi). \tag{3.6}$$

As before, we truncate the infinite series (3.3) to order P. Inserting the truncated polynomial chaos expansion into the Burgers equation results in

$$\sum_{i=0}^{P} \frac{\partial u_i}{\partial t}\Phi_i(\xi) + \sum_{i=0}^{P}\sum_{j=0}^{P} \frac{\partial(u_i u_j)}{\partial x}\Phi_i(\xi)\Phi_j(\xi) = 0. \tag{3.7}$$

Applying Galerkin projection in the probability space yields:

$$\frac{\partial u_k}{\partial t} + \sum_{i=0}^{P}\sum_{j=0}^{P} C_{ijk}\frac{\partial(u_i u_j)}{\partial x} = 0 \qquad 0 \leq k \leq P, \tag{3.8}$$

where $C_{ijk} = \langle\Phi_i\Phi_j\Phi_k\rangle$.

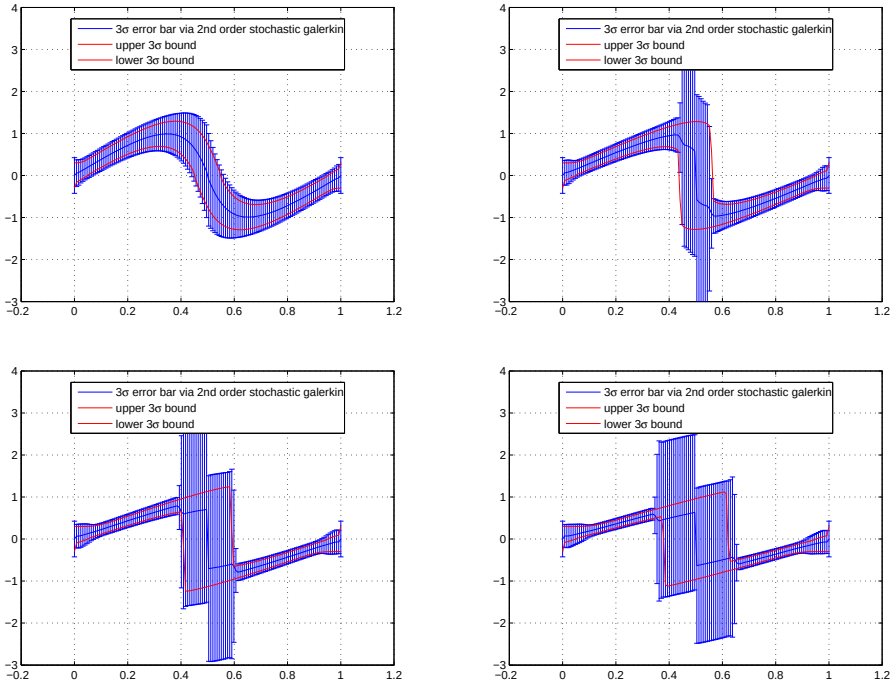

FIGURE 5. Evolution of the computed error bars at various times: t=0.1, 0.2, 0.3, 0.4. Note that the exact solution is also reported in the plots.

Equation (3.8) is again a system of partial differential equations that we can solve numerically. In order to ensure stability of the numerical scheme, the time step must be much smaller than the restriction imposed by the CFL number in the deterministic analogue. The details are not discussed in this report. After we obtain a numerical solution for $u_i(x,t)$, $i = 0, \ldots, P$, we can reconstruct the stochastic solution of (3.1) and its mean and variance. The standard deviation of the solution is given by the formula

$$u_\sigma(x,t) = \left(\sum_{i=1}^{P} u_i^2(x,t) \langle \Phi_i^2 \rangle \right)^{1/2}. \tag{3.9}$$

A 99.7% confidence interval can therefore be constructed as

$$u(x,t) = u_0(x,t) \pm 3u_\sigma(x,t). \tag{3.10}$$

Figure 5 illustrates the confidence interval we constructed using second-order polynomial chaos ($P = 2$) and compared to the exact bounds. It is also interesting to note the PCE coefficients corresponding to the computed solution during the time evolution (Fig. 6).

The comparison with the available exact solution (Fig. 5) shows that the second order polynomial chaos method captures the uncertainty in the location of the shock wave very well.

3.3. *A deterministic/stochastic hybrid approach*

In this section, we report our initial efforts in developing a different approach for uncertainty propagation. As previously mentioned, the uncertainty is introduced into the

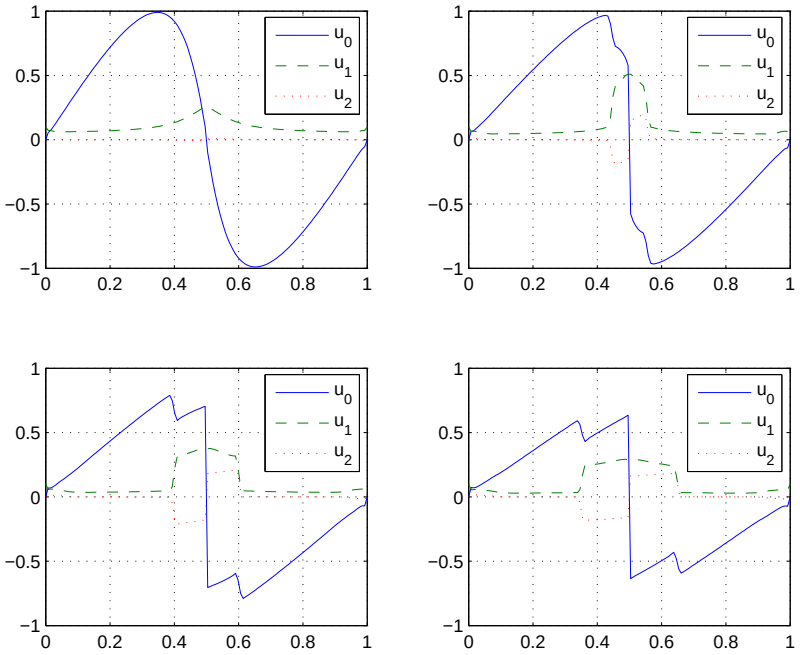

FIGURE 6. Coefficients of polynomial expansion at various instants during the time evolution. t = 0.1, 0.2, 0.3, 0.4

initial condition $u_0 = u_0(x, t = 0, \omega)$, where $\omega \in \Omega$ is the sample space. Thus u is a function of x, t, and ω, and the objective is to compute the statistics of u. In particular, we desire the mean $\bar{u}$ and the variance $\hat{u}$, defined as

$$\bar{u} \equiv \langle u \rangle = \int_{\omega \in \Omega} u dP(\omega), \quad \hat{u} \equiv \mathrm{var}(u) = \frac{1}{2} \int_{\omega \in \Omega} (u - \bar{u})^2 dP(\omega). \tag{3.11}$$

Note that other definitions of the mean are possible. If we decompose the unknown u into the mean part and fluctuation part: $u = \bar{u} + u'$, plug in into (3.1), and integrate over the sample space, we obtain an evolution equation for the mean. Higher-order equations can also be generated by multiplying (3.1) with $(u')^n$ and integrating. Below are the equations for the mean and the variance.

$$\frac{\partial \bar{u}}{\partial t} + \bar{u} \frac{\partial \bar{u}}{\partial x} = - \int_{\omega \in \Omega} u' \frac{\partial u'}{\partial x} dP(\omega), \tag{3.12}$$

$$\frac{\partial \hat{u}}{\partial t} + \bar{u} \frac{\partial \hat{u}}{\partial x} + 2\hat{u} \frac{\partial \bar{u}}{\partial x} = - \int_{\omega \in \Omega} u' u' \frac{\partial u'}{\partial x} dP(\omega). \tag{3.13}$$

Note that these equations describe the evolution of the statistics of the solutions. The terms on the right-hand side are not known (*unclosed*) and must be modeled or approximated. This is similar in many ways to the process of deriving the Reynolds-Averaged Navier-Stokes equations.

In a Monte Carlo approach, one would solve the PDE (3.1) many times with different initial conditions and average over the realizations. In the present hybrid approach, we use Monte Carlo (or other approaches) to approximate the unclosed terms in (3.12) and (3.13), then solve these equations in a deterministic sense.

We proved that, if we simply use Monte Carlo sampling to approximate the unclosed terms, then solve (3.12) and (3.13), the result is identical to using the samples directly to obtain the mean. This conclusion applies even when the process is not converged, namely if we use only a limited number of sample, provided we use the same samples in both methods. However, the former method allows us to incorporate a *model* for the unclosed terms. For instance, we could attempt to model the unclosed term in (3.12) as a diffusion process. One possible model could be written as $\beta t\sigma^2\partial^2\bar{u}/\partial x^2$, where σ is the standard deviation in the initial condition and β is a constant that can be approximated by sampling. Current research is focused on an efficient way to obtain β. We have also investigated the use of bootstrap or a reduced order model (ROM). Only the latter is briefly described here. The idea is to perform Monte Carlo sampling on coarser grids (the simplest ROM in the present problem) and to construct the approximate unclosed terms. To improve the sampling quality we also considered a few realizations obtained directly on the fine grid. For example, for the Burgers equation problem described earlier, about 1000 realizations are needed to obtain converged statistics using the Monte Carlo approach. In the two-grid method, we can achieve the same accuracy with 1000 realizations on the coarse grid (half the grid points) and 100 realizations on fine grid; the overall reduction of the computational cost is approximately 40%. The benefit will be obviously more significant in higher dimensions.

Acknowledgments

This work is supported by the Department of Energy within the Advanced Simulation and Computing Program.

REFERENCES

Cameron R. and Martin W. 1947. The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals *Ann. of Math.* **28**, 285-392.

Dowding, K., Pilch, M., and Hills, R. G. 2006. Model Validation Challenge Problem: Thermal Problem, Sandia Validation Workshop, *Sandia National Laboratory*, May 2005.

Ghanem R. and Spanos P. 1991. *Stochastic Finite Elements: A Spectral Approach* Springer-Verlag, New York.

Wan X., Xiu, D. and Karniadakis, G. E. 2004. Modeling uncertainty in three-dimensional heat transfer problems in Advanced Computational Methods in Heat Transfer VIII, Proceedings of the Eighth International Conference on Advanced Computational Methods in Heat Transfer, Lisbon, Portugal, 2004, B. Sunden, C. Brebbia, and A. Mendes, eds., WIT Press, Southampton, UK.

Wiener, N. 1938. The homogeneous chaos *Amer. J. Math.* **60** 897-936.

Xiu D. and Karniadakis, G. E. 2002. The WienerAskey polynomial chaos for stochastic differential equations *SIAM J. Sci. Comput* **24**, 619-644.