SANDIA REPORT

SAND2014-1839
Unlimited Release
Printed March 2014

Dakota Uncertainty Quantification Methods Applied to the NEK-5000 SAHEX Model

V. Gregory Weirs

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

SAND2014-1839
Unlimited Release
Printed March 2014

Dakota Uncertainty Quantification Methods Applied to the NEK-5000 SAHEX Model

V. G. Weirs
Multiphysics Applications
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1320
vgweirs@sandia.gov

Abstract

This report summarizes the results of a NEAMS project focused on the use of uncertainty and sensitivity analysis methods within the NEK-5000 and Dakota software framework for assessing failure probabilities as part of probabilistic risk assessment. NEK-5000 is a software tool under development at Argonne National Laboratory to perform computational fluid dynamics calculations for applications such as thermohydraulics of nuclear reactor cores. Dakota is a software tool developed at Sandia National Laboratories containing optimization, sensitivity analysis, and uncertainty quantification algorithms. The goal of this work is to demonstrate the use of uncertainty quantification methods in Dakota with NEK-5000.

Acknowledgments

This work was funded by the Nuclear Energy Advanced Modeling and Simulation (NEAMS) Program under the Advanced Modeling and Simulation Office (AMSO) in the Nuclear Energy Office in the U.S. Department of Energy. The authors specifically acknowledge the Program Manager of the Reactors Product Line, Dr. David Pointer (ORNL), and Dr. Keith Bradley, National Technical Director of NEAMS, for their support of this work. Finally, we thank the NEK-5000 and Dakota development teams.

Contents

1 Introduction **7**

1.1 The NEK-5000 SAHEX Model and Dakota 7

1.2 Input Random Variables and Response Quantities 9

1.3 Sensitivity Analysis and Uncertainty Quantification Approach 10

1.3.1 Sampling Methods 11

1.3.2 Polynomial Chaos Expansions 11

1.3.3 Reliability Methods 12

2 Sensitivity Analysis Results **15**

2.1 Scatter Plots 15

2.2 Correlation Coefficients 15

2.3 Sobol' Sensitivity Indices 20

2.4 Summary 22

3 Uncertainty Quantification Results **23**

3.1 Moments 23

3.2 Cumulative Distribution Function 24

3.3 Probability of Failure 24

4 Concluding Remarks **27**

References **28**

Chapter 1

Introduction

Nuclear reactor simulation and modeling is complex. A wide range of physical phenomena must be considered, and each of these must be accurately modeled. Uncertainties in these models must be assessed in any analysis because they can have a significant effect on the conclusions and how they are used by decision makers. If the uncertainties of a model's inputs can be characterized, for example by probability distributions, then the uncertainties can be propagated through the model to estimate uncertainties on the model's outputs.

This report is structured as an analyst might proceed through a probability of failure analysis. The engineering problem involves the heat transfer from hot fuel rods to liquid sodium in a nuclear reactor. The analyst is asked to determine the probability that the liquid sodium exceeds a threshold temperature; beyond this temperature the sodium can no longer absorb heat and the fuel rods are no longer cooled. While the details of the heat transfer problem are greatly simplified in this example, it is sufficient to demonstrate the steps in the analysis and how Dakota and NEK-5000 might be used to address a practical problem in nuclear reactor analysis.

1.1 The NEK-5000 SAHEX Model and Dakota

The simplified 6-pin hexagonal assembly (SAHEX) is used as our demonstration problem [9]. The geometry consists of six fuel rods around a steel control rod, all in a steel hexagonal "can". The fuel rods are also clad with steel. The can is open at the top and bottom, and represents a section of a much longer assembly. Liquid sodium absorbs heat from the hot fuel rods as it flows down through the can. Figure 1.1 shows the spatial relationship between the fuel rods, control rod, and the can.

In a real reactor, the fuel rods may deform because of thermal stresses and turbulent flow, and each fuel rod may generate a different amount of heat for a variety of reasons. Across arrays of hexagonal assemblies, the thermal and fluid flow conditions can vary significantly. Simulations to model these various phenomena often involve several different physics codes operating in combination: for example in [9], NEK-5000 models the the fluid flow and heat transfer, Diablo models the stress field and deformation of the fuel rods, and PROTEUS models the the nuclear reactions and heat generation within the fuel rods themselves.

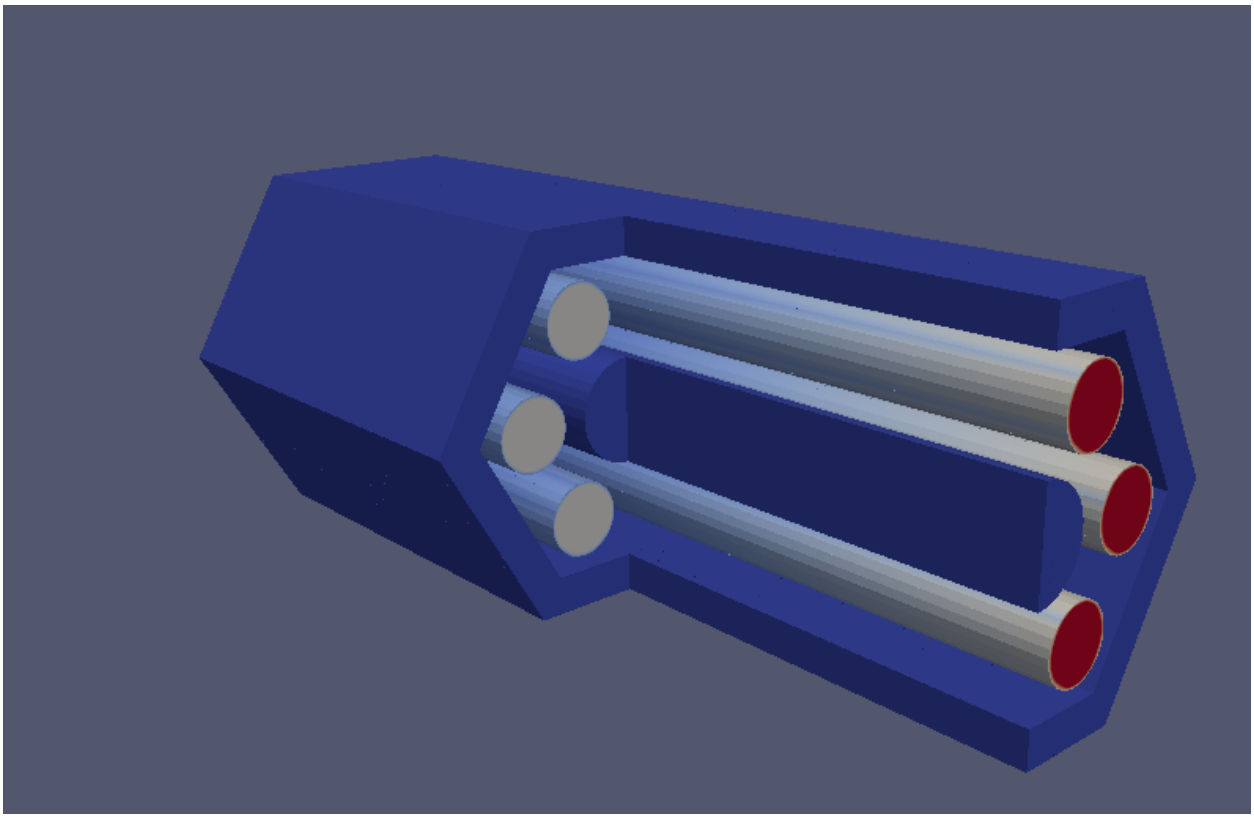

Figure 1.1. SAHEX model geometry with cutaway showing fuel rods, control rod, and can.

In this report we consider a simplified version of the problem. The fuel rods are assumed to be rigid and the heat generation rate is assumed to be constant but uncertain. Only the fluid flow and heat transfer are simulated. The flow velocity entering the top of the can is constant, and no-slip conditions are applied at the rod and can surfaces. Flow is assumed to be laminar and incompressible. The fluid temperature at the inlet is $600K$ and the can boundaries are adiabatic.

We use NEK-5000 [5] to compute approximate, steady-state solutions to this problem. NEK-5000 solves the incompressible Navier-Stokes equations describing fluid flow, in conjunction with the energy equation to describe heat transfer in the fluid as well as solids. The computational domain is discretized by the spectral element method, and several time advancement methods are available for transient or steady-state solutions. NEK-5000 is designed to run on massively parallel supercomputers and has demonstrated excellent scaling on the largest computing platforms in the world.

Dakota [1] is a software tool that includes methods for parameter studies, sensitivity analysis, uncertainty quantification, and optimization. While each of these has its own set of challenges and solution algorithms, they share a common structure: all involve a number of function evaluations, and each function evaluation provides a response vector corresponding to a specific input parameter vector. For an optimization problem, the input parameters are the design variables and the response vector is the objective function. For an uncertainty quantification problem, the input parameters have characterized uncertainties and the uncertainties of the responses are computed.

Dakota was designed to handle a function evaluation that involves the approximate solution of a system of partial differential equations as implemented in a massively parallel, multiphysics simulation code, such as NEK-5000. It has a file-based, script-driven interface to the simulation code (each simulation corresponding to a function evaluation) to provide maximum flexibility to the user. For driving NEK-5000 as the simulation code, the function evaluation includes pre-processing the NEK-5000 input file to set the input parameter values that Dakota specifies, and postprocessing the simulation results to extract the response quantities to return to Dakota.

For the studies in this report, each SAHEX simulation took about 15 minutes running on 128 cores (16 nodes at 8 cores/node.) In this case, the function evaluation also included submission to the queue system of massively parallel computing platform. It is also possible to run Dakota as a parallel job, and let Dakota manage the execution of each simulation on the nodes the queue system provides, although this can be complicated depending on the platform and queue system.

1.2 Input Random Variables and Response Quantities

For this study our hypothetical analyst has been asked to investigate the temperature of the liquid sodium. If the sodium gets too hot, above $660.0K$, it will undergo a phase change and will no longer carry heat away from the fuel rods. The analyst is asked to calculate the probability that the sodium will remain below this temperature, while accounting for uncertainties in other aspects of the SAHEX model. The maximum sodium temperature in the flow field, $T_{Na,max}$, can be easily

determined from the NEK-5000 solution.

Six inputs were considered to be uncertain. Four are associated with the material properties of liquid sodium: the density (ρ_{Na}), the dynamic viscosity coefficient (μ_{Na}), the heat capacity per unit volume (expressed as $(\rho C_p)_{Na}$), and the thermal conductivity (k_{Na}). The volumetric heat generation rate of the fuel (q_{vol}) and the fluid inflow velocity (v_z) complete the set. Table 1.1 lists these inputs, nominal values, and the distributions used to characterize their uncertainties.

Parameter	Nominal Value	Probability Distribution
ρ_{Na}	$0.85\ g/cm^3$	Normal, $\sigma = 0.03$
μ_{Na}	$0.26\ g/(cm \cdot s)$	Normal, $\sigma = 0.03$
$(\rho C_p)_{Na}$	$1.08110\ J/(cm^3 \cdot K)$	Normal, $\sigma = 0.05$
k_{Na}	$0.7\ W/(cm \cdot K)$	Normal, $\sigma = 0.04$
q_{vol}	$100.0\ W/cm^3$	Lognormal, $\sigma = 15.0$
v_z	$65.3896\ cm/s$	Uniform, $[45.3896, 85.3896]$

Table 1.1. Uncertain input parameters and their characterization.

The nominal values listed in Tab. 1.1 are estimates but may not be representative of liquid sodium; in particular the viscosity coefficient appears to be quite high. The probability distributions for the sodium parameters were chosen arbitrarily, with the goal of providing detectable uncertainties in the flow field temperatures. The heat generation rate and flow velocity were chosen with the expectation that they would have a strong effect on the temperatures. The distribution of q_{vol} was intended to produce high temperatures but infrequently. The velocity distribution was chosen to represent, perhaps, a variety of different locations in a large array, and through its bounds, to have a strong influence on the temperature. No specific reactor designs informed these choices. Finally, the choice of $660K$ as the threshold temperature was made after an initial parameter study, and has no significance for true sodium.

1.3 Sensitivity Analysis and Uncertainty Quantification Approach

Three classes of methods were used for the SAHEX analysis. Sampling methods are the most common, and many methods for sample design have been proposed to achieve a variety of objectives. Sampling methods are robust but often require a large number of function evaluations to achieve accurate results. Stochastic expansion methods build a high-order polynomial model of the response, and can be more efficient than sampling methods if the response functions have a high degree of regularity (are smooth in many derivatives). In contrast to sampling and stochastic expansion methods, reliability methods do not seek a high fidelity model of the entire response surface; instead they rely on strong assumptions about the response behavior, or focus on just a small region or a point of the response surface.

Sensitivity analysis (SA) and uncertainty quantification (UQ) are closely related. In uncertainty quantification, input uncertainties are propagated through the computational model to determine the uncertainty of the responses. When inputs are modeled with probability distributions, the goal of an uncertainty quantification study is to obtain a probability distribution of the responses. In a sensitivity analysis, the input variables that affect the response the most are identified. It is not necessary to have distribution information on inputs to perform sensitivity analysis: the goal is to identify the most significant inputs and usually only bounds on the inputs are specified.

It is often convenient to do both UQ and SA studies, even when only one is the main objective. Sensitivity analysis is most often used in the exploratory stage of an analysis, to consider a large number of uncertain input variables and identify the most important ones as a precursor to a more accurate uncertainty quantification analysis with a smaller number of inputs. For the SAHEX model, sampling methods and stochastic expansions were used for both sensitivity analysis and uncertainty propagation. Reliability methods were used for uncertainty quantification.

1.3.1 Sampling Methods

In a sampling method, each input random variable is characterized by a probability distribution. A sample point is taken from the input distributions (one value per variable) and the function is evaluated from this sample point, producing a response. This process is repeated a number of times (say, N times) to build up a discrete representation of the distribution of response values.

Random sampling (while still respecting the distributions of each input variable) requires a large number of sample points to accurately determine the output distribution. In particular, resolving the tails of the output distribution takes a very large number of points. Nevertheless, random sampling is easy to implement, easy to understand, and produces unbiased estimates of descriptive statistics.

Latin Hypercube Sampling (LHS) is a stratified sampling. [8, 7] Each input distribution is divided into N strata (or bins) of equal probability, and each sample point gets a value from a unique bin for each variable. In this way, the entirety of each input distribution is sampled uniformly and the clustering of sample points observed in random sampling is avoided. As a consequence, LHS requires fewer sample points to achieve the same accuracy as random sampling, and for some problems the convergence rate (reduction of error as N increases) is higher. Another advantage is that for correlated input variables, LHS can generate a sample design that satisfies the specified correlation structure.

1.3.2 Polynomial Chaos Expansions

Stochastic expansion methods approximate the outputs of the uncertain system through series expansions of standard random variables. [11, 2] Polynomial Chaos Expansions are a particular class of stochastic expansions that model the response by an orthogonal polynomial basis of the input

random variables. [4] Function evaluations are used to determine the coefficients of these polynomials. Once the coefficients of the polynomials are determined, the polynomial model can be analyzed or sampled for much less cost than the simulation code.

In polynomial chaos expansions (PCE), the polynomials are usually chosen according to the Weiner-Askey scheme that provides an orthogonal basis with respect to the probability density function for the input random variables. The expansion is based on Hermite polynomials which are functions of the Gaussian random variables. The goal of a PCE analysis is to determine the unknown coefficients of the Hermite polynomials in the series expansion. Usually, these coefficients can be calculated from a limited number of model simulations. There are a variety of approaches for generating the sample points, but typically tensor product grids or sparse grids are used. Structured sampling schemes typically result in more accurate estimates of the coefficients of the polynomials than random sampling does, but this depends on the order of the polynomials needed to resolve the uncertainty in the response functions. Two PCE studies were carried out for this report. The first used a 2-level sparse grid design, and the second used a 3-level sparse grid with "dimension preference." Dimension preference allows the user to increase the polynomial accuracy for some variables and reduce it for others.

An important distinguishing feature of the PCE methodology is that the solution is expressed as a random process, not merely as statistics as is the case in many nondeterministic methodologies. PCE can be most efficient when there is some additional information from the physics of the problem about what type of approximating polynomials are most appropriate. One advantage of stochastic expansion methods is that the moments of the expansion (e.g., mean or variance of the response) can be written analytically, along with analytic formulations of the derivatives of these moments with respect to the uncertain variables.

1.3.3 Reliability Methods

Reliability methods are probabilistic algorithms for quantifying the effect of uncertainties on response metrics of interest. They compute approximate response function distribution statistics based on the specified probability distributions for input random variables. These response statistics include mean, standard deviation, and cumulative or complementary cumulative distribution function (CDF/CCDF) response level, and probability/reliability level pairings. Reliability methods are often more efficient at computing statistics in the tails of the response distributions (events with low probability) than sampling-based approaches, because the number of samples required to resolve a low probability can be prohibitive. Consequently these methods are often used to assess the probability of failure of a system when confronted with an uncertain environment.

The Mean Value method [6] and the family of most probable point (MPP) search methods are local reliability methods; they are gradient-based and employ local approximations or incorporate a local optimization method. They require few function evaluations, but are only effective when the limit state function is smooth and does not have multiple MPPs. The efficient global reliability analysis (EGRA) method [3] is designed to address these drawbacks by first building a Gaussian process model from a small number of samples, then adaptively adding function evaluations to

pinpoint the limit state. That is, EGRA uses function evaluations to accurately determine a specific point of the response CDF, while the objective of the sampling and PCE methods is to represent the entire response distribution.

Chapter 2

Sensitivity Analysis Results

For this report, sensitivity analysis and uncertainty quantification were performed on the same set of data for the sampling and uncertainty quantification studies. Three LHS studies were performed, with 60, 300, and 600 samples, and correlation coefficients were examined. Correlation coefficients measure the degree to which the response is linearly related to each input. Two PCE studies were performed, and Sobol' sensitivity indices were calculated. Sobol' sensitivity indices estimate how much of the output uncertainty can be attributed to each input. Sobol' sensitivity indices can also be computed for LHS designs, but require significantly more function evaluations to do so.

2.1 Scatter Plots

First consider scatter plots of the response QOI, $T_{Na,max}$. Figure 2.1 shows scatter plots of $T_{Na,max}$ vs. each of the uncertain input variables. Each scatter plot shows the results of all the simulations in the LHS60 study. There is one point from each NEK-5000 simulation. Scatter plots give a visual indication of the relationship between the response quantity and the uncertain input. For example, there appears to be a broadly linear relationship between $T_{Na,max}$ and the heat generation rate, q_{vol}, and also between $T_{Na,max}$ and v_z. There does not seem to be a strong relationship with any of the other inputs. Figures 2.2 and 2.3 show scatter plots for the LHS300 and LHS600 studies, respectively. While many more points are shown, the basic relationships between inputs and outputs are the same as for the LHS60 study.

Scatter plots provide a qualitative view of the sampling results, but can be misleading for a number of reasons. Too few or too many samples can make a scatter plot difficult to read. For a large number of uncertain inputs, each scatter plot shows the influence of all the inputs, but in a way that they cannot be distinguished. Interactions between inputs are also hard to detect.

2.2 Correlation Coefficients

The Pearson correlation coefficient is a statistical measure of the degree to which the relationship between an input variable and a response quantity is linear. Dakota computes Pearson correlation

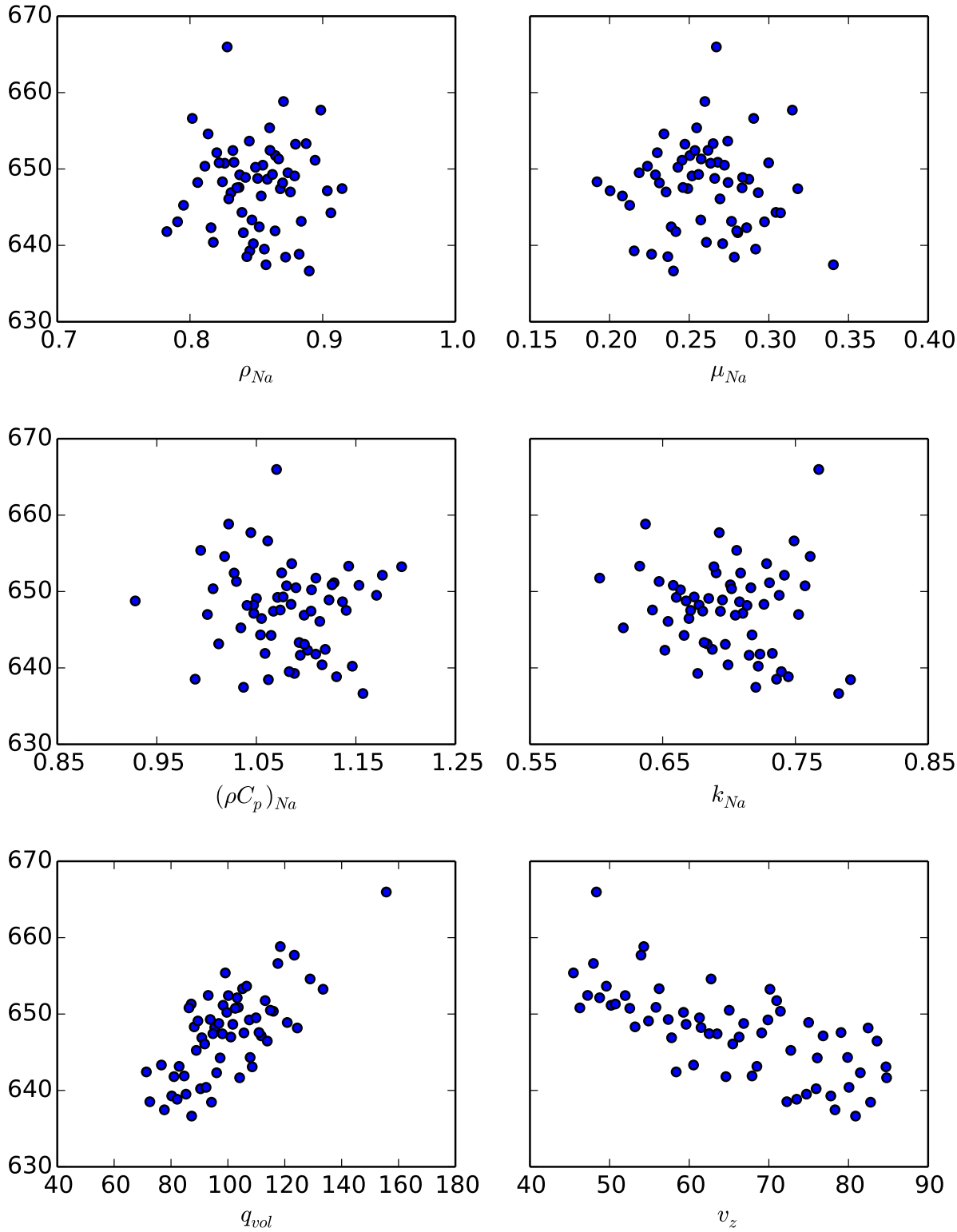

Figure 2.1. Scatter plots of $T_{Na,max}$ vs. uncertain input variables for LHS60.

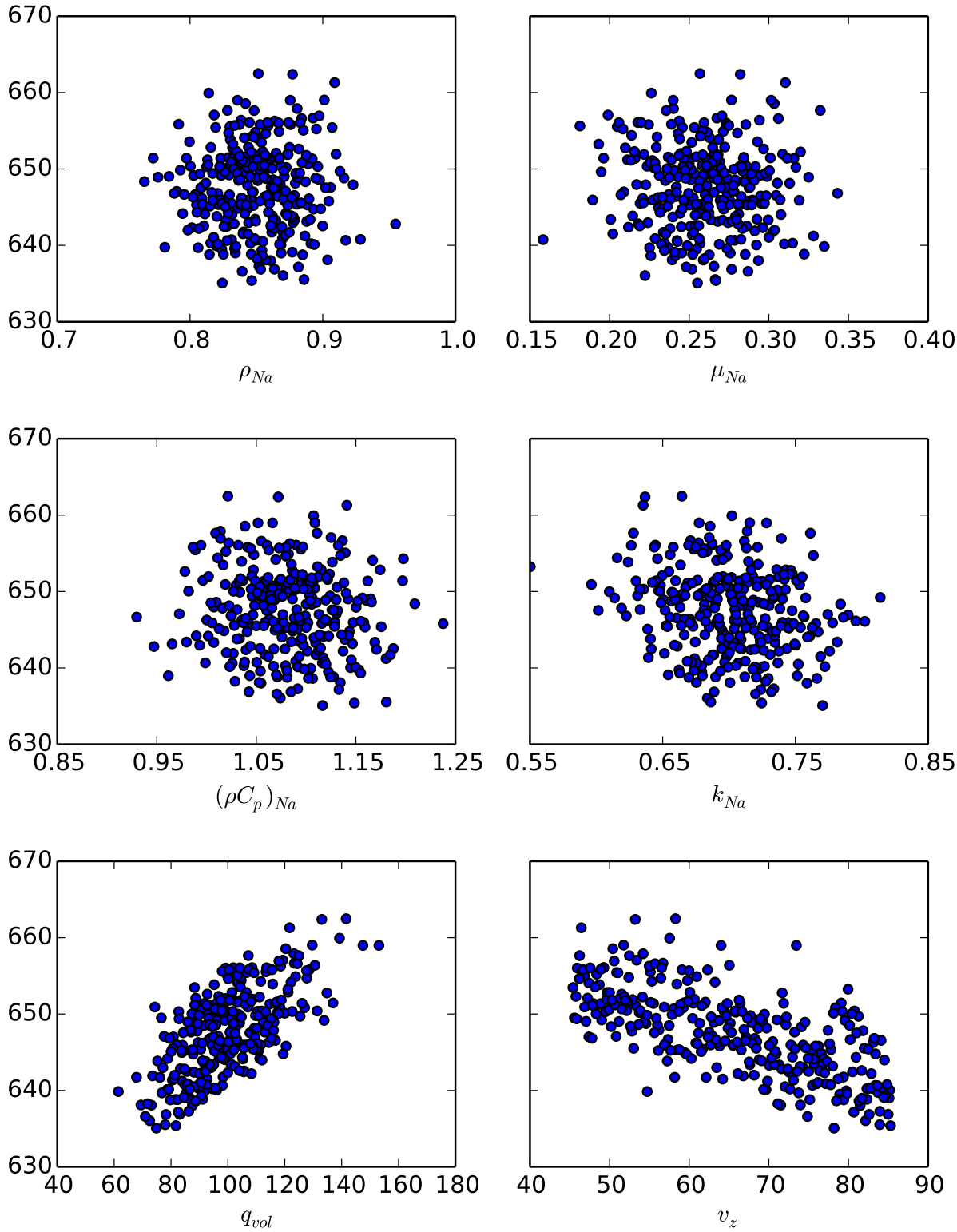

Figure 2.2. Scatter plots of $T_{Na,max}$ vs. uncertain input variables for LHS300.

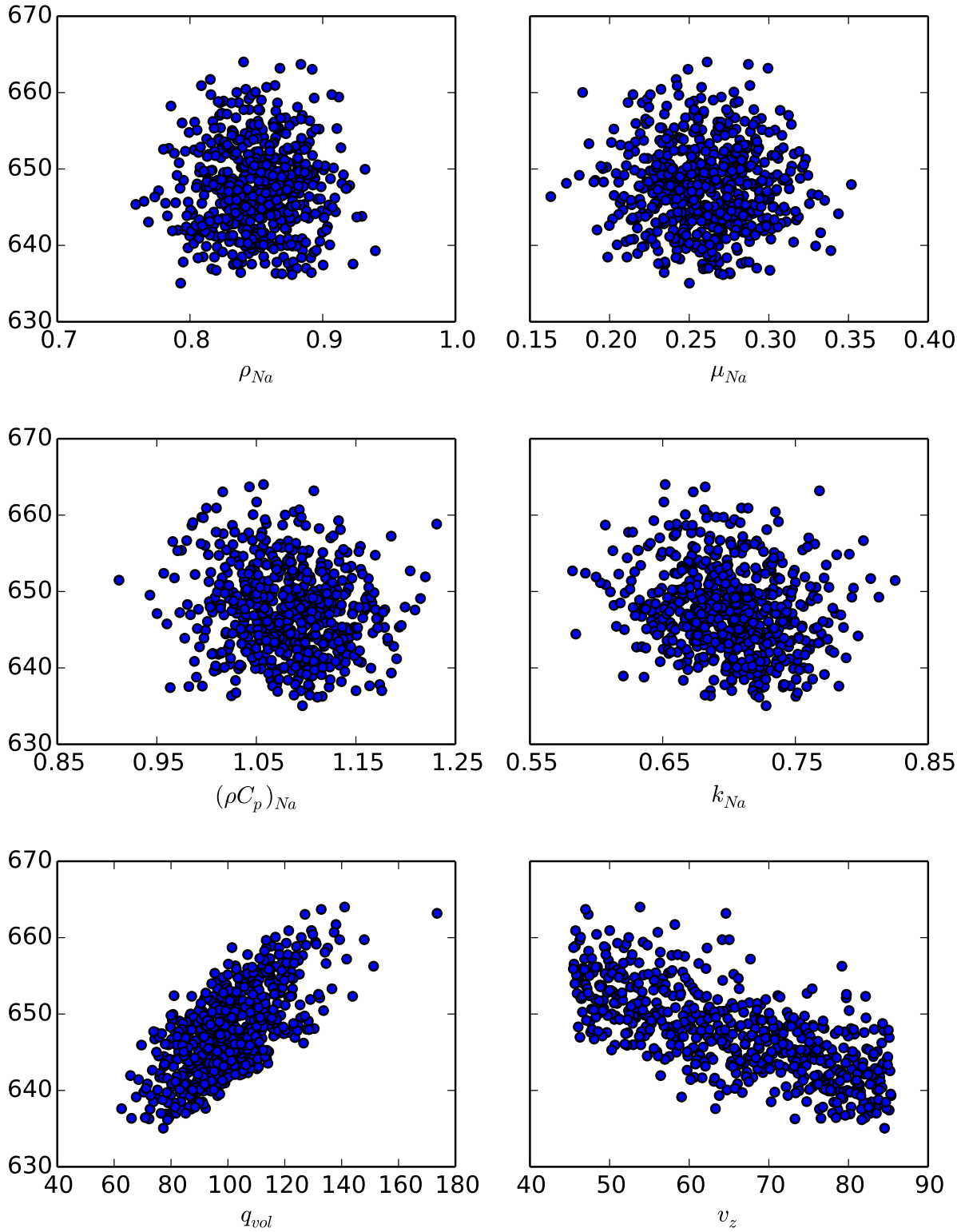

Figure 2.3. Scatter plots of $T_{Na,max}$ vs. uncertain input variables for LHS600.

coefficients (also called simple correlation coefficients) for sampling studies. Table 2.1 presents the simple correlation coefficients for $T_{Na,max}$ with respect to each input. Correlation coefficients range from -1 to 1, with a value of 1 meaning the samples can be perfectly fit by a line with positive slope, i.e., the value of the response increases exactly linearly as the value of the input increases. A value of -1 indicates an inverse relationship, or negative slope. As the correlation tends towards zero, the strength of the linear relationship decreases. For a response and a set of uncertain input variables, the set of correlation coefficients provide a form of sensitivity analysis, with low values indicating insensitivity and values above 0.5 or below -0.5 indicating significant sensitivity. The results in Tab. 2.1 indicate a strong positive correlation between $T_{Na,max}$ and q_{vol} and a strong negative correlation with v_z, while relationships with inputs are much weaker. The results are consistent across the three sample sizes.

	$T_{Na,max}$		
Input Variable	LHS60	LHS300	LHS600
ρ_{Na}	-0.0150409	0.0221388	0.0247335
μ_{Na}	-0.0311482	-0.0413672	-0.0249558
$(\rho C_p)_{Na}$	-0.0895889	-0.100121	-0.112835
k_{Na}	-0.134971	-0.210983	-0.196727
q_{vol}	0.717346	0.679238	0.683053
v_z	-0.704351	-0.687004	-0.691781

Table 2.1. Simple correlation coefficients from sampling studies for $T_{Na,max}$.

Partial correlation coefficients for the sampling studies are shown in Tab. 2.2. Partial correlation coefficients consider the relationships between an input and an output with the effect of the other input random variables removed. As a result, the partial correlation coefficient for a given input expresses the linearity of the relationship with the response that has the effect of the other inputs removed. As shown in Tab. 2.2, the uncertain variables $\rho C_p)_{Na}$, k_{Na}, q_{vol}, and v_z each have very strong partial correlations with the response $T_{Na,max}$. The Pearson correlation coefficients are calculated between an input and an output but the effects of all the other inputs, which are varied simultaneously in each sample, are not filtered out; these effects are filtered out in the partial correlation coefficients.

For the SAHEX model, the correlation coefficients provide insight to the analyst about the relative importance of each input variable, and that the inputs are essentially independent. However, in general, one should interpret correlation coefficients with caution. Correlation coefficients are known to be sensitive to outliers, and to the ranges of the inputs. A strong, but not linear relationship may be obvious in a scatter plot, while the correlation coefficient may be near zero.

	$T_{Na,max}$		
Input Variable	LHS60	LHS300	LHS600
ρ_{Na}	0.214469	0.252528	0.211722
μ_{Na}	-0.540591	-0.651517	-0.623418
$(\rho C_p)_{Na}$	-0.926689	-0.917832	-0.916610
k_{Na}	-0.976045	-0.979460	-0.977381
q_{vol}	0.998216	0.997874	0.997868
v_z	-0.998100	-0.997967	-0.997917

Table 2.2. Partial correlation coefficients from sampling studies for $T_{Na,max}$.

2.3 Sobol' Sensitivity Indices

The Sobol' sensitivity indices are computed from the decomposition of the variance of the response. [10] This decomposition attributes the variance to each input variable and their interactions. The first-order sensitivity index S_i for each uncertain input i is the fraction of the variance of the response that can be attributed to that input alone; the S_i are also called the "main effects" sensitivity indices. The total sensitivity index T_i is the fraction of the variance of the response that can be attributed to that input and all its interactions.

Variation-based sensitivity analysis is a form of global sensitivity analysis; the sensitivity coefficients do not represent partial derivatives of the response with respect to the input variables, but the effects of each of the inputs integrated over the input hypercube. Since polynomial chaos expansions construct a polynomial representation of the response over the input hypercube, it is inexpensive to compute the Sobol' sensitivity indices; in fact, they can be computed analytically from the coefficients of the interpolating polynomials. The sensitivity indices can also be estimated by sampling methods, but require significantly more function evaluations; for a sample N, the cost is $(d+2)N$ function evaluations, where d is the number of input variables.

The main effects indices for the two polynomial chaos expansion studies are shown in Tab. 2.3. The heat generation rate and inflow velocity are clearly identified as the most important, with the sodium thermal conductivity (k_{Na}) a distant third. Table 2.4 presents the total sensitivity indices, and the results are essentially identical to the main effects indices. Since the difference between the main effects and the total indices is the inclusion of interactions between inputs, the conclusion is that there are no significant input interactions for the SAHEX model.

	$T_{Na,max}$	
Input Variable	PCE SG2	PCE SG3-DP
ρ_{Na}	0.000131	0.000120
μ_{Na}	0.001377	0.001543
$(\rho C_p)_{Na}$	0.010558	0.010506
k_{Na}	0.043282	0.043340
q_{vol}	0.463005	0.462860
v_z	0.480864	0.480992

Table 2.3. Sobol' main effects sensitivity indices from polynomial chaos expansion studies for $T_{Na,max}$.

	$T_{Na,max}$	
Input Variable	PCE SG2	PCE SG3-DP
ρ_{Na}	0.000144	0.000120
μ_{Na}	0.001502	0.001543
$(\rho C_p)_{Na}$	0.010595	0.010543
k_{Na}	0.04.3467	0.043525
q_{vol}	0.463605	0.463448
v_z	0.481470	0.481460

Table 2.4. Sobol' total effects sensitivity indices from polynomial chaos expansion studies for $T_{Na,max}$.

2.4 Summary

From the sampling and polynomial chaos studies of the SAHEX model, we have computed simple and partial correlation coefficients and Sobol' sensitivity indices. The most important inputs are q_{vol} and v_z, as indicated by both the simple correlation coefficients and the sensitivity indices. The negligible differences between the main effects and total sensitivity indices demonstrates that the interactions between the inputs are not important, and this can also be inferred from the simple and partial correlation coefficients. We have computed correlation coefficients from sampling methods and sensitivity indices from polynomial chaos expansions. However, it is possible to perform both analyses from both classes of methods.

Chapter 3

Uncertainty Quantification Results

In this section the focus is on the uncertainty of the response. Histograms and distribution functions show the overall shape and bounds of the response. The probability of failure provides information about a specific part of the response distribution.

3.1 Moments

The moments of the response distribution are statistical measures of the distribution's shape. The first and second moments are the familiar mean and standard deviation, and these are presented for each method in Tab. 3.1. For the SAHEX problem, all the methods exhibit very similar results. The LHS60 study might be less accurate due to the small sample size, and the Mean Value reliability method because of the strong assumptions it makes. The X-Taylor-Mean and EGRA reliability methods do not report the moments.

		$T_{Na,max}$	
Method	Function Evaluations	Mean Value	Standard Deviation
LHS 60	60	647.6426K	5.811488K
LHS 300	300	647.5600K	5.560594K
LHS 600	600	647.5618K	5.614899K
PCE SG2	117	647.5506K	5.590595K
PCE SG3-DP	102	647.5572K	5.585691K
Reliability MV	13	647.2665K	5.513859K
Reliability XTM	14	–	–
Reliability EGRA	78	–	–

Table 3.1. Moments of the maximum sodium temperature as computed by different methods.

Table 3.1 also reports the number of function evaluations each method used, which is a good representation of their computational cost. The mean value method and MPP class of methods

require the fewest. The EGRA method starts with a small LHS sample before applying a global optimization approach, both of which increase the cost. The polynomial chaos methods are more expensive than the reliability methods, but characterize the entire response distribution are probably more accurate. Latin Hypercube Sampling is similar to the PCE methods in the data it provides, and while less efficient than PCE methods, it may be affordable at low resolution when PCE methods are not.

3.2 Cumulative Distribution Function

The cumulative distribution function (CDF) expresses the probability that the response will be less than a particular value. This type of information is particularly useful for decision makers and regulators. For example, a design or regulatory threshold might be expressed as " the temperature will remain below X with 99% certainty." The CDF shows the certainty as a function of the temperature.

If the response obeys a known distribution function, the CDF can be easily calculated. More often, the response distribution is described only by a set of sample values, and the distribution and CDF must be estimated or modeled. The most common way to estimate a distribution is to plot the histogram of the sample data. The overall range of response values $\{Y_n\}$ is divided up into non-overlapping "bins," which usually have the same width, ΔY. The histogram is the number of response values in each bin. When each bin value is replaced by the bin value plus the cumulative sum of the bins to the left, then divided by the total number of sample points, the resulting values approximate the cumulative distribution function. Figure 3.1 plots the approximate CDF for each method. For all the methods, the histogram was computed with a bin width of $1K$. For the LHS studies, the CDFs are comparable for the different sample sizes, although the LHS60 CDF is less smooth than the larger samples. The PCE studies give essentially identical results, and are most similar to the LHS600 CDF. Recall that for the PCE studies, NEK-5000 provides the function evaluations from which the coefficients of the polynomial expansion are determined. To compute the CDFs, Dakota used a 10,000 point sample of the expansion to compute the CDF. The large sample size results in much smoother histograms and CDFs, compared to the LHS studies.

3.3 Probability of Failure

The "probability of failure" refers to the context in which a system or component "fails" if a critical value is exceeded, but below which it operates as designed; the probability of failure (POF) is the probability that the critical value is exceeded. In our hypothetical scenario, the analyst is asked to determine the probability that the temperature of the liquid sodium remains below a temperature of $660.0K$ throughout the flowfield; from a computational perspective this is the same as a probability of failure estimate.

The POF can be graphically determined by reading the probability corresponding to the critical

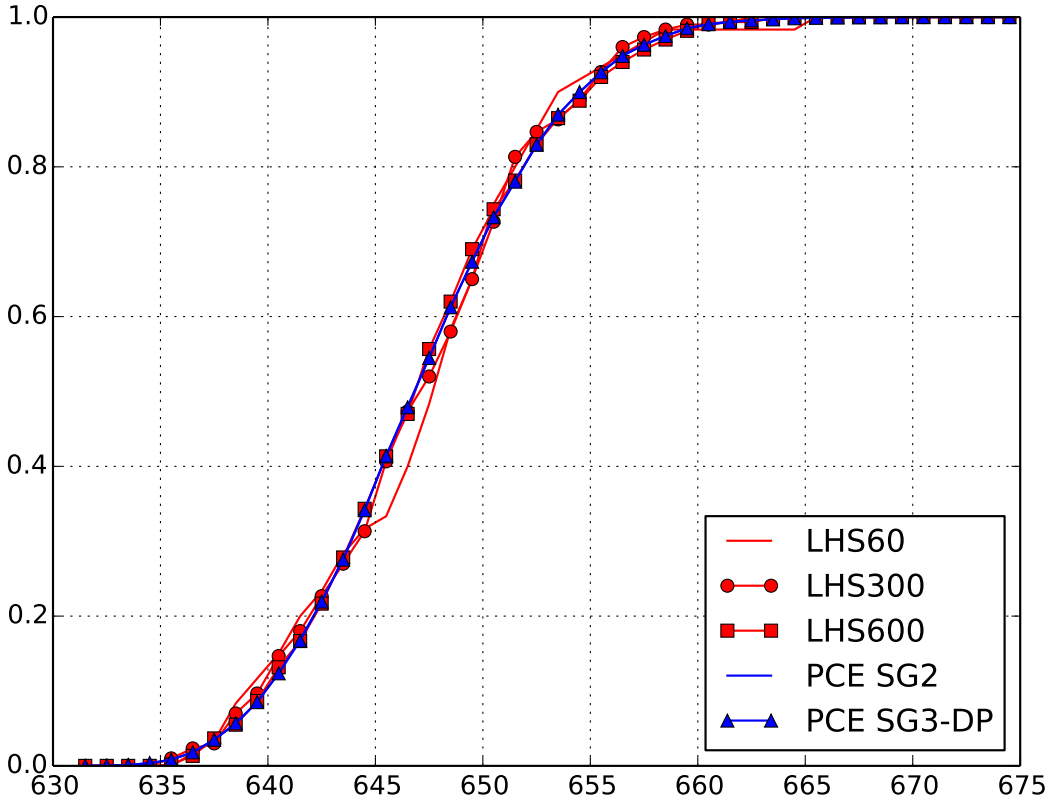

Figure 3.1. Approximate CDFs for each method.

response value off of the CDF. However, approximate CDFs may not be sufficiently accurate for small sample sizes and may be sensitive to the choice of bins. Finally, the most often desired POF is in the tail of the distribution, e.g., 1%, 0.1%, or 99%, 99.9%, etc., which is the most difficult part to obtain accurately for sampling methods. Instead, different approaches to get more accurate probability of failure estimates are used.

For sampling methods, a surrogate model can be constructed from the sample points. The surrogate model can be evaluated very inexpensively, so the surrogate can be sampled with a large sample size, and focused on the most interesting part of the distribution, to reduce the sampling error. Nevertheless, many function evaluations are required to build an accurate surrogate. Stochastic expansion methods can be viewed as a class of surrogate models, where the interpolating polynomial is the surrogate. Finally, as mentioned in the introduction, reliability methods can be very efficient at computing the POF because the either make strong assumptions about the response (local reliability methods) or adaptively choose sample points to accurately resolve a point or small region of the respond distribution (EGRA).

Dakota can numerically estimate the probabilities corresponding to particular response values, or response values corresponding to particular probabilities, as the user requests. Table 3.2 presents the probability that $T_{NA,max}$ will remain below $660.0K$, as computed by the different methods. For the LHS studies, no surrogate models were used; the POF is estimated from an approximate CDF computed by Dakota. For the PCE methods, a 10,000 point sample of the orthogonal polynomial was used to construct the approximate CDF.

Method	Function Evaluations	$P(T_{Na,max} < 660.0K)$
LHS 60	60	0.983333
LHS 300	300	0.990000
LHS 600	600	0.981667
PCE SG2	117	0.985000
PCE SG3-DP	102	0.985000
Reliability MV	13	0.989538
Reliability XTM	14	$0.984160^{\dagger}$
Reliability EGRA	78	0.982953

Table 3.2. Probability that $T_{Na,max}$ is less than $660.0K$ as computed by different methods. † The response level for the Reliability XTM method is $660.8024K$

For the SAHEX model, all of the methods gave similar results. The Mean Value and X Taylor Mean reliability methods perform well for a small number of function evaluations because to response is a simple function of the input variables. As shown by the sensitivity analysis earlier, $T_{Na,max}$ is strongly dependent on q_{vol} and v_z, there are no significant interactions between the inputs, and with respect to all the inputs except the least important, the response is linear.

Chapter 4

Concluding Remarks

In this report, several methods for uncertainty quantification and sensitivity analysis were applied to the SAHEX model. The purpose was to demonstrate the combined use of Dakota and NEK-5000 for a simplified but practical scenario. Latin Hypercube Sampling and Polynomial Chaos expansion methods analyze the whole response space, and are used for both sensitivity analysis and uncertainty quantification. Correlation coefficients and Sobol' sensitivity indices showed that for the SAHEX model, the maximum temperature of the liquid sodium, $T_{Na,max}$, was a simple function of the uncertain inputs. There were no significant interactions among the inputs and $T_{Na,max}$ was a linear function of most of them; the heat generation rate, q_{vol}, and the flow velocity, v_z, were the most important inputs. In addition to the LHS and PCE studies, several reliability methods were used to answer a probability of failure question, namely, to compute the probability that $T_{Na,max} < 660.0K$. All the methods performed comparably for this simple model.

Dakota provides great flexibility to the user in the development of an interface to their simulation code. The uncertain inputs for this report were all specified in the NEK-5000 .rea input file, and the response was extracted from the NEK-5000 text output file using common UNIX commands. However, more significant changes to the SAHEX model, including manipulation of the mesh, extensive analysis of the flowfield with another software package, or even requiring recompilation of NEK-5000 for each function evaluation, could be accommodated by an analyst's changes to the Dakota-NEK5000 interface. Each function evaluation was a 128-core NEK-5000 simulation, executed on an institutional-scale HPC through a fair-share queue system. Dakota dynamically managed multiple, concurrent NEK-5000 simulations through this queue system with minimal user intervention.

References

[1] B. M. Adams, W. J. Bohnhoff, K. R. Dalbey, J. P. Eddy, M. S. Eldred, P.D. Hough, S. Lefantzi, L. P. Swiler,and D. M. Vigil "DAKOTA: A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis. Version 5.2 Users Manual," *Sandia Technical Report SAND2010-2183*, updated February 2013 (Version 5.3).

[2] B. M. Adams, W. J. Bohnhoff, K. R. Dalbey, J. P. Eddy, M. S. Eldred, P.D. Hough, S. Lefantzi, L. P. Swiler,and D. M. Vigil. "DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 5.2 Theory Manual." *Sandia Technical Report SAND2011-9106*. Updated February 2013 (Version 5.3).

[3] B. J. Bichon, M. S. Eldred, L. P. Swiler, S. Mahadevan, and J. M. McFarland. "Efficient Global Reliability Analysis for Nonlinear Implicit Performance Functions," *AIAA Journal*, **46**, no. 10, pp. 2459–2468 (2008).

[4] M. S. Eldred, C. G. Webster and P. Constantine, "Evaluation of Non-Intrusive Approaches for Wiener-Askey Generalized Polynomial Chaos," *Proceedings of the 10th AIAA Non-Deterministic Approaches Conference*, 2008.

[5] P. F. Fischer, J. W. Lottes and S. G. Kerkemeier, *nek5000 Web page*, http://nek5000.mcs.anl.gov, 2008.

[6] A. Haldar and S. Mahadevan. *Probability, Reliability, and Statistical Methods in Engineering Design.* Wiley, New York: 2000.

[7] J. C. Helton and F. J. Davis. "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems," *Reliability Engineering and System Safety*, **81**, no. 1, pp. 23–69 (2003).

[8] R. L. Iman and M. J. Shortencarier, "A Fortran 77 Program and User's Guide for the Generation of Latin Hypercube Samples for Use with Computer Models," Sandia National Laboratories, NUREG/CR-3624, SAND83-2365, 1984.

[9] V. Mahadevan, E. Merzari, R. Jain, A. Obabko, M. Smith, T. Tautges, P. Fischer, W. D. Pointer, R. Ferencz, "SHARP Fuel Assembly Coupled Simulation Demonstration," *2013 ANS Winter Meeting and Nuclear Technology Expo*, 2013.

[10] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis: The Primer.* Wiley, 2008.

[11] D. Xiu, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton University Press, Princeton, NJ, 2010.

DISTRIBUTION:

1 Keith Bradley
National Technical Directory NEAMS
Argonne National Laboratory
9700 S. Cass Avenue
Argonne, IL 60439

1 David Pointer
Oak Ridge National Laboratory
Reactor and Nuclear Systems Division
P. O. Box 2008, Bldg. 5700, MS-6167
Oak Ridge, TN 37831

1 Aleksandr Obabko
Argonne National Laboratory
9700 S. Cass Avenue
Argonne, IL 60439

1	MS 1318	J. R. Stewart, 01441
1	MS 1318	B. M. Adams, 01441
1	MS 1318	L. P. Swiler, 01441
1	MS 1320	V. G. Weirs, 01446
1	MS 1322	V. A. Mousseau, 01441
1	MS 1321	R. M. Summers, 01446
1	MS 0899	Technical Library, 9536 (electronic copy)

Sandia National Laboratories