

## STATISTICAL VALIDATION FOR HEAT TRANSFER PROBLEMS: A CASE STUDY

S.N. SCOTT<sup>1</sup>, J.A. TEMPLETON<sup>1</sup>, P.D. HOUGH<sup>1</sup>, J.R. RUTHRUFF<sup>1</sup>,  
M.V. ROSARIO<sup>1,2</sup> & J.P. PETERSON<sup>1</sup>

<sup>1</sup>Sandia National Laboratories, USA.

<sup>2</sup>Department of Biology, Duke University, USA.

### ABSTRACT

This paper presents a proposed methodology for applying statistical techniques as the basis for validation activities of a computer model of heat transfer. To demonstrate this approach, a case study of a Ruggedized Instrumentation Package subject to heating from battery discharge and electrical resistance during normal operations is considered. First, the uncertainty in the simulation due to the discretization of the governing partial differential equations is quantified. This error is analogous to the measurement error in an experiment in that it is not representative of actual physical variation, and is necessary to completely characterize the range of simulation outcomes. Secondly, physical uncertainties, such as unknown or variable material properties, are incorporated into the model and propagated through it. To this end, a sensitivity study enables exploration of the output space of the model. Experiments are considered to be a realization of one of these possible outcomes, with the added complication of containing physical processes not included in the model. Statistical tests are proposed to quantitatively compare experimental measurements and simulation results. The problem of discrepancies between the computational model and tests is considered as well.

*Keywords:* Heat transfer, mesh resolution, numerical parameters, uncertainty quantification, validation, verification.

### 1 INTRODUCTION

Validating computational models against experimental data is challenging due to the inherent uncertainties in both realms [1]. Any uncertainty implies that a model cannot be ‘validated’ in a binary sense, but rather must be statistically assessed with regard to its prediction of reality within a given configuration. Therefore, this paper proposes that validation activities, as applied to computer simulations, focus on obtaining statistical confidence measures to provide evidence the model has predictive value. Validation can be viewed then as an application of uncertainty quantification (UQ) [2]. Illustrating this point of view is a case study involving the temperature response of a Ruggedized Instrumentation Package (RIP) in normal operating conditions. The RIP contains circuit boards, instrumentation packages and is battery powered; therefore, heat is generated through electrical resistance. As an engineering application, designers need to know the temperatures reached by the various electrical components to design the system such that these hardware pieces will function throughout the RIP’s operation. Assessing the degree to which the present model can be used for this task involves quantifying all errors in the model, which is achieved through the verification and validation (V&V) process [3–5]. Comprehensive V&V processes have been suggested for several areas of computational physics, such as computational fluid dynamics [6], as well as for some industries, such as nuclear energy [7]. While there are shared best practices across the approaches (as captured in [3–5]), to the best of our knowledge, there is no universally accepted approach to this problem.

The first step in comparing simulation output to experimental results involves quantifying the inherent uncertainties in both processes, i.e. estimating the degree to which they can

measure their underlying truth models. In the case of experiments, the underlying truths are the laws of physics, but experimental apparatuses are unable to perfectly measure any quantity in the system. In this work, established error estimates for thermocouple measurements are used. To approximate the errors associated with the discretization of the partial differential equations (PDEs) comprising the model, i.e. its truth, solution verification techniques are used. Part of the error in the solution is contributed by time integration methods, linear/nonlinear solver tolerances, etc., which are considered the numerical parameters of the model [8]. Quantification of these uncertainties for a similar configuration of the RIP was presented at the CMEM Conference [9] and is not repeated here. In addition, mesh quality and granularity also significantly impact solution fidelity. The previous paper presented an algorithm for generating a family of meshes that can be used for this task. However, a significantly improved algorithm has since been developed and is presented in Section 3 along with the solution error estimates obtained using meshes generated by it. Section 3 also provides a brief summary of the numerical parameter uncertainty estimate to provide a total measure of the verification uncertainty of the present RIP model. It is known that solution verification and validation are intertwined, with mesh resolution often being the most significant numerical parameter as well as a challenging one against which to quantify solution error [10].

Validation is focused on ensuring that the model is a reasonable representation of the physical system of interest and as such, the process seeks to assess the degree to which the model can predict quantities of interest for relevant configurations. In this work, two aspects of validation are considered. First, a sensitivity study is performed to learn which physical parameters included in the model are the primary drivers of output metrics and therefore to prioritize uncertainties around which further study will focus. The second part of validation is a comparison between model and experimental results. Any computational model necessarily is missing some physical phenomena and not all the relevant phenomena may be known *a priori*. Many metrics for validation are based on standard statistical approaches, e.g. *p*-values from correlations or confidence intervals [11]. In this work, we propose the use of the Kolmogorov–Smirnov (KS) test and assess its applicability; however, we stress much more work is required to develop more rigorous and applicable validation metrics. As the sensitivity study, experimental comparisons can suggest additional calibration activities or other model enhancements that can improve the predictive capability of the model. Both these aspects of validation are presented in Section 4.

## 2 RIP DESCRIPTION

The RIP is an assembly of electronics, batteries, and circuit boards packaged in metal housings as seen in Fig. 1. The goal was to create an FEA thermal model using the SIERRA Thermal/Fluids code [12] to monitor the temperature in specified locations to ensure that electronics remained within their operating temperature ranges. The RIP is approximately 0.15 m in diameter. Heat sources in the RIP are 48 2/3 A-cell batteries and two electronic packages (EP1 and EP2).

The analysis team received production style ProEngineer [13] (computer-aided design software) models of the RIP. To prepare this model for meshing, many of the details needed to be removed, such as screws, fillets, threaded inserts, cables, and electrical connectors and features used for tolerancing, such as small gaps. Once these changes were made, the model was exported as a STEP file and imported into CUBIT [14], a geometry creation, manipulation, and mesh generator created by Sandia National Laboratories. When the geometry was completed, the model was comprised of 119 blocks.

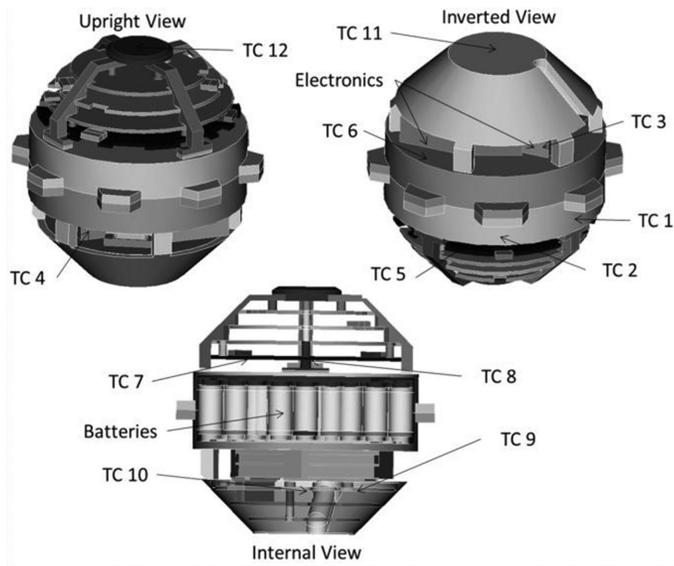


Figure 1: Locations in the RIP where temperature is monitored in this study. Additional locations were added for the validation study. Thermocouples (TCs) 13–15 are located on an external housing, not pictured.

To simulate the experimental setup, the RIP (Fig. 1) was enclosed in a housing with natural convection on all surfaces at an initial temperature of 300 K. A radiative boundary condition was placed in all enclosures. For the first minute of the test, the batteries each produce 0.07 W, while electronics on the boards produce a total of 8.66 W. For the next hour, EP1 produces 22.57 W, the batteries' heat increases to a total of 0.19 W, and the electronics on the boards increase to a total of 9.75 W. Finally, for the last 15 min of the test, EP2 turns on (22.57 W), the batteries heat increases to a total of 0.31 W, and the electronics on the boards increase to a total of 10.84 W. Fifteen thermocouples were located on the RIP, as shown in Fig. 1.

## 2.1 Computation model

A three-dimensional finite element model composed of tetrahedral elements was evaluated in the Sierra Thermal/Fluids [15] radiation–conduction code to computationally simulate the experimental configuration. The code solves the general heat equation:

$$\frac{\partial \rho c T}{\partial t} = \nabla \cdot k \nabla T + \dot{q} \quad (1)$$

where  $\rho$  is the density,  $c$  is the specific heat,  $T$  is the temperature,  $t$  is the time,  $k$  is the conductivity, and  $\dot{q}$  is the internal heat generation. Radiative and convective boundary conditions can be applied to the model. The convective boundary condition is defined as:

$$q = h(T_s - T_{inf}) \quad (2)$$

where  $q$  is the heat flux,  $T_s$  is the temperature of the surface,  $T_{inf}$  is the ambient temperature, and  $h$  is the convective heat transfer coefficient, which is defined as:

$$h = \frac{\overline{Nu_D}}{Dk} \quad (3)$$

where  $D$  is the characteristic diameter,  $k$  is the thermal conductivity of the fluid, and  $Nu_D$  is the Nusselt number correlation. In this case, the RIP was approximated as a sphere and the following imperial average Nusselt number correlation was used:

$$\overline{Nu_D} = 2 + \frac{0.589Ra_D^{1/4}}{\left[1 + (0.469 / Pr)^{9/16}\right]^{4/9}} \quad (4)$$

where  $Ra_D$  is the Rayleigh number and  $Pr$  is the Prandtl number, whose definitions can be found in [16]. For the radiative boundary condition, the Sierra Thermal/Fluids code calculates the heat transferred between elements in an enclosure by calculating the view factors of each element to every other element in the system. A full treatment of radiative heat transfer and view factor calculations can be found in [16] and details on the methodology in the code [15].

### 3 SOLUTION VERIFICATION

#### 3.1 Mesh resolution study

When performing any computer analysis of PDEs, the continuous mathematical problem must be converted into a discrete representation (a mesh). The quantification of the sensitivity of the solution to the mesh size is accomplished through a mesh resolution study involving a family of topologically similar meshes with a range of characteristic length scales. The finest mesh in the set is considered to be the ‘true’ solution, because as the mesh size decreases, the mesh-based approximations to the continuous derivatives also improve. When the coarser meshes are compared with the ‘true’ solution, the error associated with discretization can be assessed.

##### 3.1.1 Original mesh creation

CUBIT (<http://cubit.sandia.gov>) was chosen as the mesh generator for this project because the built-in Python (open source scripting language) interface allowed for the creation a set of logical rules that controlled the meshing process. By scripting these logical rules, a tetrahedral mesh of the RIP was automatically created that allowed: (1) mesh sizes that were determined by the geometry of each block and (2) continuous transitions of mesh size between blocks of differing sizes. This algorithm has been applied to both the RIP as well as other heat transfer models and has been observed to be much more robust and produce higher quality meshes than the algorithm used in the original study [9]. The meshing script worked using the following algorithm:

- Obtain geometric information about the blocks, surfaces, and curves.
- Iterate through all surfaces, divide the smallest edge length in the surface in half (to guarantee the smallest edge will be made of at least two elements), and tentatively set the mesh size of all curves in the surface to that value.

- Iterate through all vertices and record the minimum mesh size of all connected curves as a list of ‘vertex goals’.
- Set continuously changing intervals on all curves by interpolating mesh size between each set of connected vertices.
- Decompose each surface into two subsets: curves that circumscribe the surface (external curves), and curves that are internal to the surface (internal curves).
- Force internal elements of the surface to interpolate smoothly between external curves and internal curves using built-in sizing functions in CUBIT.
- Attempt to mesh all volumes.

It was found that applying this mesh scheme to all blocks of the RIP caused the program to crash. However, when used only for volumes which failed to mesh on the first iteration, we were able to generate a mesh of 4,058,066 elements in approximately 8 min. This mesh will be referred to as the nominal mesh, and serves as the starting point for all subsequent meshes in this resolution study.

### 3.1.2 Refined mesh creation

The method used to create the refined mesh is identical to that used for the nominal mesh with the exception of one step: scaling the curve and surface meshes using a user-defined ‘sizeFactor’. By manipulating the ‘sizeFactor’ value (default is 1) before running the script, meshes of differing number of elements can be generated. The number of elements in the entire model is controlled by manipulating the list of vertex goals. After the list of vertex goals is generated for the nominal mesh, the list of vertex goals is multiplied by the ‘sizeFactor’ input. By specifying a ‘sizefactor’ value  $<1$ , the script will produce a mesh with smaller size and more number of elements, while values  $>1$  produce meshes with larger size and fewer elements. The ‘sizeFactor’ also has the same effect on the internal elements of the mesh and works in concert with the previous ‘vertex goal’ scaling to globally control the amount of refinement of the mesh. Eight meshes were produced by varying the ‘sizeFactor’ input to rough scale the number of elements up or down by a factor of 2. Table 1 shows the exact number of elements used to create the differently sized meshes from the same nominal mesh (4,058,066 elements).

Table 1: Element counts for meshes generated by the remesher program.

Mesh name	No. of elements
Coarse 16×	195,103
Coarse 8×	528,434
Coarse 4×	1,177,552
Coarse 2×	2,161,911
Fine 2×	7,682,930
Fine 4×	14,073,953
Fine 8×	28,282,563
Fine 16×	58,802,650

### 3.1.3 Mesh resolution study results

The finest mesh in our study,  $16\times$  fine mesh, was taken to be the closest approximation to the analytical result, and was used as the ‘true’ solution in the convergence analyses. Since this is not an exact result, mesh convergence can only be assessed in the Cauchy sense, as discussed by Hughes [17]. Thus, the rate of convergence can be identified, but the difference between the nominal and  $16\times$  mesh only provides an estimate for the absolute difference between the numerical results and analytic solution. Two standard norms were used to assess the rate of convergence: the  $L_2$  and  $L_\infty$  norms. The  $L_\infty$  norm is the absolute maximum difference between the two solutions, also normalized by the fine mesh. This norm is mathematically defined as (in one dimension):

$$L_\infty = \lim_{n \rightarrow \infty} \frac{\left( \int_{\Omega} (T_F(x) - T_C(x))^n dx \right)^{1/n}}{\left( \int_{\Omega} T_F^n(x) dx \right)^{1/n}} \quad (5)$$

where  $\Omega$  is the domain,  $T_F(x)$  is the temperature of the fine mesh at location  $x$  and  $T_C(x)$  is the temperature of the course mesh at location  $x$ . The  $L_\infty$  norm is expected to converge as  $h^{-1}$ , where  $h$  is the mesh length scale. It provides a measure of the worst-case local error that can occur and is useful for verifying that lower dimensional parts of the mesh are accurate, such as faces and contacts. In contrast, the  $L_2$  norm measures the mean square error between two solutions, normalized by the fine mesh, and is expected to converge as  $h^{-2}$ . It is useful for assessing the overall quality of the solution.

The results of the temperature simulation for a region of interest in the RIP, TC2, demonstrate the patterns seen throughout the model. The results of three of the meshes (the finest mesh, the nominal mesh, and the coarsest mesh) show that the maximum divergence of temperature occurs at the end of the simulation at 60 min (Fig. 2). The maximum difference between the finest mesh and the coarsest mesh was 1.222 K, whereas the maximum difference in temperature between the finest mesh and the nominal mesh was 0.078 K, indicating that the solution is converging. To compare the norms with their expected convergence rates, the norms were plotted on a log–log plot, and a linear regression was performed. The three coarsest meshes were excluded from the regression because their values of  $L_\infty$  were not indicative of convergence and skewed the regression. After this exclusion, the convergence rate of  $L_2$  was  $-2.37$  (expected value =  $-2.0$ ,  $p = 0.002$ ,  $r^2 = 0.991$ ) while the rate of convergence of  $L_\infty$  was  $-0.947$  (expected value =  $-1.0$ ,  $p = 0.0167$ ,  $r^2 = 0.849$ ), as shown in Fig. 3. The nominal mesh was chosen as the optimal mesh for two reasons: (1) the nominal mesh is included in the regression and is within the linear convergence regime and (2) it is also sufficiently ‘coarse’ to allow tractable computation times.

### 3.2 Uncertainty quantification for solution verification

The approach taken in this work is that the total numerical uncertainty in the simulation is approximated by the independent summation of the errors arising from a mesh resolution study and a numerical parameter study. Performing an integrated UQ study with both factors simultaneously would stress existing high-performance computing platforms; therefore, this definition is taken out of practicality. As the geometry and results are sufficiently similar to the numerical parameter study previously performed on the RIP [9], the key results are simply

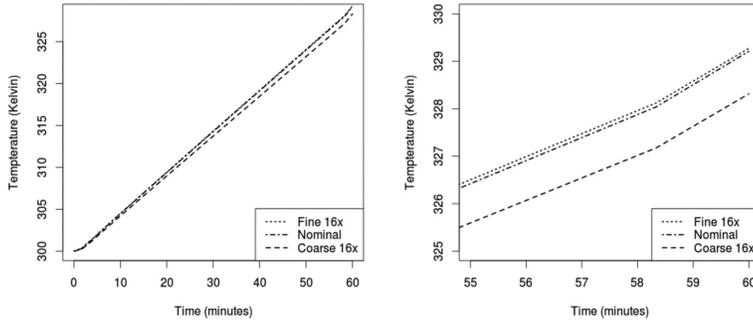


Figure 2: Temperature for TC2 for the 16× fine mesh, nominal mesh, and 16× coarse mesh: entire time history (left), zoom in at the end of the simulation (right).

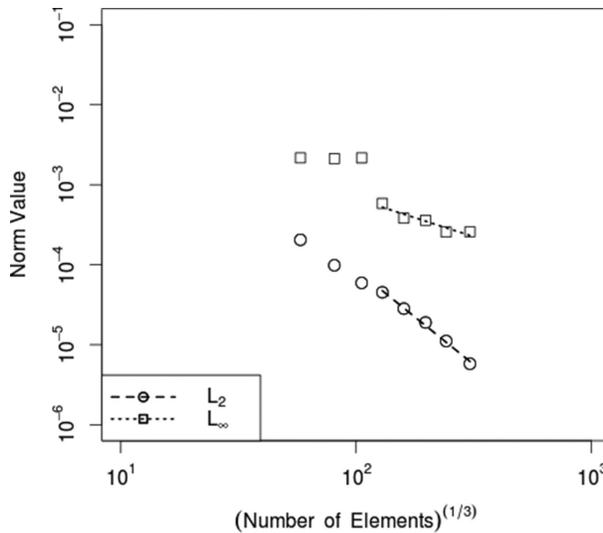


Figure 3:  $L_2$  and  $L_\infty$  norms with best fit lines ( $L_2$  best fit slope =  $-2.37$ ,  $L_\infty$  best fit slope =  $-0.947$ )

quoted here and the interested reader is referred to that work. The three most significant numerical parameters, in terms of contributing to solution error and model evaluation time, are in Table 2. The relevant results from [9] are that the maximum time step and integration order are the parameters driving the uncertainty, resulting in a 0.75 K variation in temperature.

The data from the mesh resolution and the numerical parameters studies can be used to quantify the total uncertainty in the temperature field due to mesh discretization. The maximum difference in temperature between the 16× fine and the 16× coarse meshes is a conservative estimate of the uncertainty with regard to the mesh as shown in Table 3. For each thermocouple, the maximum temperature difference ranged from 1.10 to 1.24 K. To determine the uncertainty from the mesh for all points on the RIP, not just the tracked locations, the  $L_2$  and  $L_\infty$  norms are employed. Since the mesh resolution study showed that the nominal mesh was optimal for our problem, a conservative estimate for the error introduced by the

Table 2: Significant parameters influencing temperature response and computational time.

Parameter name	Description
Maximum time step	The maximum amount of time allowed between each time step. The allowed range is 10–200, nominal is 100.
Residual norm scaling	The method for measuring the error. The choices are none or RHS, which normalizes it against the right-hand side. The nominal is none.
Time integration method	The order of accuracy is for time integration. The choices are first order and second order. The nominal is second order.

Table 3: Uncertainty associated with each mesh.

Mesh name	$L_2$ error	$L_\infty$ error
Coarse 16 $\times$	$\pm 0.21\%$	$\pm 2.05\%$
Coarse 8 $\times$	$\pm 0.18\%$	$\pm 2.02\%$
Coarse 4 $\times$	$\pm 0.15\%$	$\pm 0.86\%$
Coarse 2 $\times$	$\pm 0.13\%$	$\pm 0.38\%$
Nominal	$\pm 0.09\%$	$\pm 0.42\%$
Fine 2 $\times$	$\pm 0.05\%$	$\pm 0.24\%$
Fine 4 $\times$	$\pm 0.03\%$	$\pm 0.23\%$
Fine 8 $\times$	$\pm 0.01\%$	$\pm 0.15\%$

mesh at any point in our problem is  $\pm 0.42\%$ . In addition, the  $L_\infty$  correlates well with the temperature difference-based approach ( $335 \text{ K} * 0.42\% = 1.4 \text{ K}$ ). Thus, the total temperature error is approximately 2.15 K or  $\pm 0.62\%$ .

#### 4 VALIDATION

With the model verified such that errors originating from the numerical discretization of the model are quantified and controlled, the model's predictive capabilities can be assessed through a validation process. In this work, simulation responses will be compared with experimental temperature measurements of the RIP designed to approximate its real-world operating conditions. Thermocouples were placed at locations important to the functionality of the device. Therefore, the primary validation metrics are (1) the model's assessment of the physical parameters which control the RIP's thermal performance and (2) the model's ability to reproduce these data. The first part of this section will correlate the change in the model's predicted temperatures with changing physical input parameters. The second subsection presents the comparisons of the time–temperature histories from the model and experiment, using statistical tests to provide a quantitative metric.

##### 4.1 Identification of the most significant computational model parameters

For this study, the LHS capability in DAKOTA optimization and uncertainty analysis package [15] (<http://dakota.sandia.gov>) was used to sample the uncertain material model

parameters and to generate and record thermocouple responses. The study used a total of 570 samples, corresponding to a sample rate of 10 samples per parameter for each of the 57 physical parameters in the model. A complete list of these parameters and their uncertainty bounds is presented in Appendix B. To broadly sample the parameter space, a uniform distribution is used for the parameter uncertainties as well as relatively large uncertainty bounds. Additionally, some parameters, such as convection coefficients, effective thermal properties for circuit boards, and battery heat output, are highly uncertain. Each simulation is post-processed to compute the temperature histories at the locations corresponding to thermocouples in the experiment. The time-averaged mean-square error between each temperature series and the experimental value is computed, as well as  $L_2$  and  $L_\infty$  norms between the run and the nominal model (the ‘nominal model’ uses the nominal value for the parameters).

The data generated by the LHS study were analysed using the R open-source statistical software [18]. We first considered the trends of the thermocouple responses relative to the parameters varied. We quantified those trends using Pearson correlations [19]. The Pearson correlation coefficient is given by

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \quad (6)$$

where  $x_i$  and  $y_i$  are given samples from their respective populations, and  $\bar{x}$  and  $\bar{y}$  are the corresponding sample means. Pearson correlations with magnitudes close to 1 indicate a strong linear relationship between the two populations. To aid in understanding the physics that drive the RIP’s thermal performance, scatter plots are also used to visually inspect the data. Of the statistically significant parameters, the batteries’ thermal output is the most important driver of the temperature profile over time and space, which in turn is the most important parameter for all the thermocouples other than TC9 (where the convective coefficient on the inside of the RIP was the most important). The next most important parameters were the internal convection coefficient and the heat generation from EP1. The thermocouples on the circuit boards (TCs 7, 8, 9, and 10) were in general less affected by the heat sources (the batteries, electronics packages, and the electronics on the boards) than the rest of the monitored locations. The only exception was the resistor that was directly adjacent to TC8. The material properties of the batteries, the silicone compression pads that are adjacent to many of the heat generating electronics, the paraffin wax that encapsulates the batteries, the polycarbonate housing for the batteries, and aluminium that comprises most the structure of the RIP are also significant parameters to most of the monitored locations. This indicates that the material properties of the parts surrounding the heat sources have a strong effect on the temperature of the monitored locations. TC7 and TC8 are also partially governed by the circuit boards’ in plane conductivity while almost all of the thermocouples are sensitive to the heat capacity of the circuit boards.

R (<http://www.r-project.org>) was also used to complete an ANOVA study [20–22]. ANOVA consists of constructing the best average linear regression through the sampled data and then determining the ratio of the sum of squares difference between the samples and regression line to the sum of squares difference between the samples and the sample mean. The latter is used to assess the extent to which the variability in any given model parameter contributes to the overall variability in the model output. Furthermore, significance tests then allow for the

identification of the parameters that have the most influence on the uncertainty in model responses. Of particular interest is the  $p$ -value, which establishes the probability that the variability attributed to a given parameter is instead due to chance. Thus, low  $p$ -values correspond to influential parameters. Appendix A contains the ANOVA results. The  $p$ -value serves as an indicator to filter out unimportant parameters in that no parameter with a high  $p$ -value was assessed to be significant either by the Pearson correlation coefficients or by visual inspection. However, some parameters which have low  $p$ -values appear to have limited influence on the results because the  $p$ -value only measure the probability a relationship exists and not the strength of that relationship. As a result of the statistical analysis and visual inspection of the data, of the 57 parameters, only 18 appear to have any bearing on the predicted temperature field, while just one, the heat generation from the batteries, is the most controlling parameter for the temperature at 14 of the 15 thermocouple locations. Other parameters that proved important to most of the thermocouple locations were the interior convection coefficient, the specific heat of the batteries, and the heat generation of EP1. The global sensitivities to these parameters are illustrated in Figs. 4 and 5.

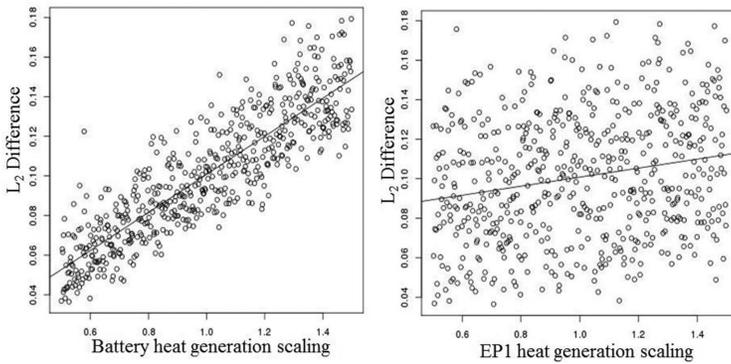


Figure 4: Global  $L_2$  differences for battery heat generation (left) and EP1 heat generation (right).

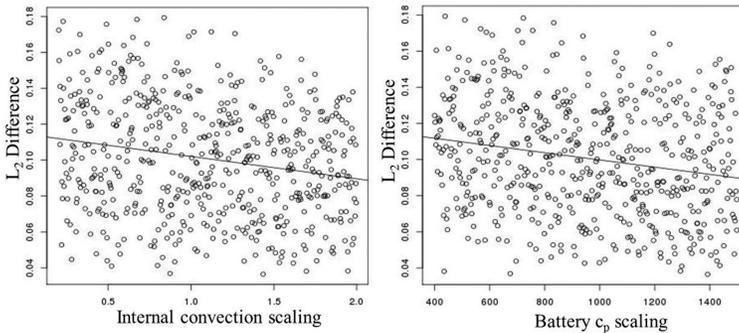


Figure 5: Global  $L_2$  differences for internal convection coefficient (left) and battery specific heat (right).

## 4.2 Comparison with experiment

To provide validation data for the RIP model, an experiment was performed to evaluate self-heating effects that occur when the instrumentation assembly is operated under representative electrical and thermal conditions simulating a 75-min period of use in the field. Fifteen thermocouples were attached to the RIP in addition to air temperature gauges inside and outside. The location of the thermocouples is shown in Fig. 1. The test was conducted with the following three specific goals: (1) demonstrate electrical performance of all components in the RIP, i.e. the battery (TCs 1, 2, 5, and 6), two EPs (TCs 3 and 4), and circuit boards (TCs 7, 8, 9, and 10), (2) determine the peak temperatures reached at the external aluminium housings, the battery housings, EP housings, and circuit boards under conditions representative of thermal loading due to self-heating and boundary conditions, and (3) collect temperature data at various points throughout the RIP to support the development and validation of the transient thermal FEA model. The latter included the internal housing (TC11), bracket cap (TC12), and external housings (TCs 13, 14, and 15). From a development perspective, the most important TCs are 13–15, followed by TCs 1–6, with TCs 7–12 being the least important.

The test was performed in a convective thermal chamber. The chamber was initially set to 307 K (34°C) with the test unit mounted at three points on a thermally isolating cork block. The test unit was allowed to soak to ensure that all internal components reached the chamber temperature. Once electrical testing commenced, the chamber was turned off for the remainder of the 75-min test, to prevent the blower from turning on and artificially increasing the convective heat transfer from the external housings. Since the thermal chamber was well insulated, the air temperature only dropped approximately 3°C over the 75-min test. The RIP was connected to various external devices for verifying electrical functionality throughout test, such as laptop computers, data loggers, digital multi-meters, and oscilloscopes.

Despite the well-controlled experiment, there are still many unknowns in the as-tested RIP. These include performance measures such as heat generation (especially from the batteries), thermo-physical properties of the constituent materials, and air-mediated thermal transport models with convection coefficients. While impossible to capture with a single experiment, some of the uncertainty is also due to differences in operating conditions (e.g. power outputs), deviation in materials used to build the RIP (e.g. conductivities), and poorly characterized physical processes (e.g. convection coefficients). In contrast, these uncertainties can be accounted for in the model by creating an ensemble of runs over instantiations of the model parameters.

To compare the model with the experiment, the R statistical package was used to compute the mean temperature response, standard deviation, and extreme values associated with the sample. Figures 6 and 7 provide representative results for TCs 1 (worst statistical agreement) and 7 (best statistical agreement).

The first observation of the results is one of general overprediction of the experimental data by the model, which was observed in all thermocouples (Fig. 8). One hypothesis is that this is most likely the result of the nominal model value for the battery heat generation being too high. Evidence supporting this hypothesis is found in the  $L_\infty$  norm plot in [18]. This plot measures the maximum difference in temperature from the nominal model for each time step. It can be seen that the norm indicates jumps happen at the initial time, when EP1 turns on, and again at 3660 s when EP2 turns on, which act as additional heat sources. These data suggest that these events are triggering significant changes in the model response. Aspects of

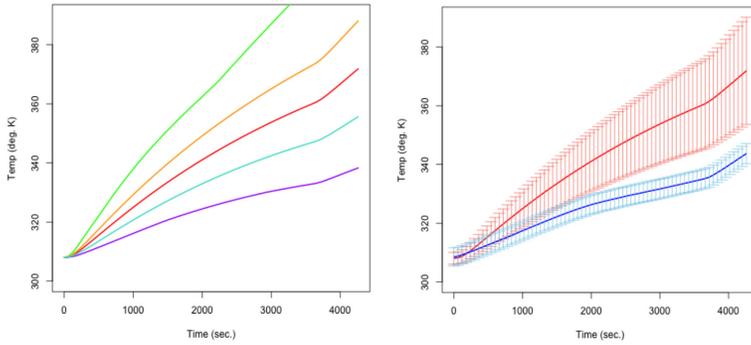


Figure 6: Left: Uncertainty in TC1 model predictions – maximum (green), one positive standard deviation (orange), mean (red), one negative standard deviation (turquoise), and minimum (purple). Right: Comparison of model (red) and experimental (blue) results with error bars denoting the associated standard deviations.

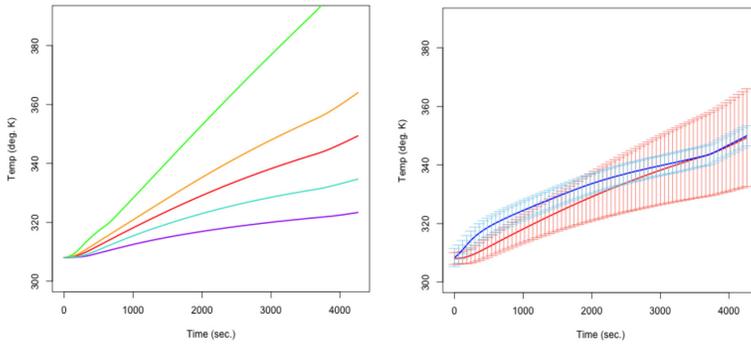


Figure 7: Left: Uncertainty in TC7 model predictions – maximum (green), one positive standard deviation (orange), mean (red), one negative standard deviation (turquoise), and minimum (purple). Right: Comparison of model (red) and experimental (blue) results with error bars denoting the associated standard deviations.

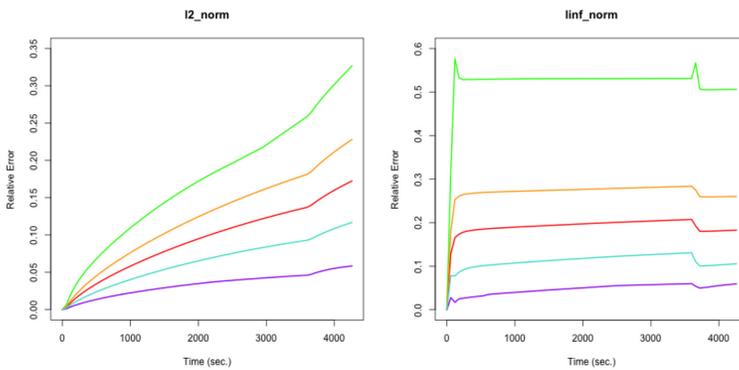


Figure 8:  $L_2$  (left) and  $L_\infty$  (right) norms of the sample compared with the nominal model – maximum (green), one positive standard deviation (orange), mean (red), one negative standard deviation (turquoise), and minimum (purple).

this hypothesis will be tested in the next subsection when the most significant parameters are identified. In this subsection, it can also be partially assessed by statistically comparing various offsets from the mean as a fraction of the standard deviation. To quantitatively assess the offset, the KS test is used. This statistical test is designed to compare distributions from two sets of data with the null hypotheses being that the data are drawn from the same distribution. The test statistic is given by

$$D = \max_{1 \leq i \leq N} \left( F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i) \right) \quad (7)$$

where  $F$  is the theoretical cumulative distribution,  $i$  is the given sample, and  $N$  is the total number of samples. More information on this test can be found in [23–25]. A theory governing how to apply this test to the simulation is currently lacking; rather, we see the use of this test as a first step in bringing established statistical methods to computational validation by using it to compare the experimental and computational temperature–time histories at each thermocouple location. It is advantageous in this work in that it provides a quantitative metric that can be used as an aggregate measure of agreement between model and test data.

The last column shows the amount of sample variance explained by this regression. For conciseness, short-hand notation is used for the variable names: battery thermal output (BAT), electronics package 1 thermal output (EP1), polycarbonate heat capacity (POL), internal convection coefficient (CON), resistor 2 heat capacity (RES), circuit board heat capacity (CIR),

Table 4: Linear regression coefficients for the three most important variables contributing to each thermocouple response.

Thermocouple	Variable #1 coefficient		Variable #2 coefficient		Variable #3 coefficient		% Variance
	Name	Value	Name	Value	Name	Value	
TC1	BAT	0.301	EP1	0.104	POL	5.6E-5	78
TC2	BAT	0.314	EP1	0.0971	CON	−0.0534	77
TC3	BAT	0.290	EP1	0.184	RES	−5.8E-5	79
TC4	BAT	0.300	EP1	0.133	RES	−6.5E-5	79
TC5	BAT	0.313	CON	−0.0499	POL	−5.7E-5	78
TC6	BAT	0.312	EP1	0.130	POL	−5.7E-5	78
TC7	BAT	−0.097	CON	0.467	CIR	4.2E-5	13
TC8	BAT	0.173	CON	−0.0953	N/A		57
TC9	BAT	0.0289	CON	−0.0168	N/A		4
TC10	BAT	0.0654	CON	−0.0220	CIR	−2.3E-5	13
TC11	BAT	0.292	EP1	0.141	CON	−0.0670	76
TC12	BAT	0.271	CON	−0.100	AL	0.0926	73
TC13	BAT	0.283	EP1	0.105	CON	−0.550	73
TC14	BAT	0.293	EP1	0.108	CON	−0.0555	73
TC15	BAT	0.271	CON	−0.0765	EXT	−0.0761	65

6061 T6 aluminium conductivity (AL), external convection coefficient (EXT). Only parameters with a regression coefficient  $>1.E-5$  are shown.

To determine the degree of deviation between the experiment and the model, the sum of the  $p$ -values for the KS test of each thermocouple was computed. The  $p$ -value is the KS test statistic, which indicates that the null hypothesis can be rejected with probability  $1 - p$ . Hence, a  $p$ -value of 1 indicates excellent agreement between the model and data, whereas a  $p$ -value of 0 indicates poor agreement (note the  $p$ -values from KS and ANOVA tests are interpreted differently). At present, there is no theory to determine what the  $p$ -value means in terms of the predictive capability of the model, so rather they are used only to compare two hypotheses. In this case, the hypotheses are different heat outputs from the battery, which was known ahead of time to be poorly characterized.  $p$ -values can be used to quantitatively assess whether different values of this parameter result in improved or degraded agreement with the test data. To efficiently test this hypothesis, the observation that the batteries' heat generation has a general uniform and linear trend in the model's predicted thermocouple responses is leveraged. The linear fit coefficient for each thermocouple's response to each input parameter was computed, with the three most significant parameters for each thermocouple shown in Table 4. For alternate battery thermal sources, each thermocouple time history was then perturbed by a function proportional to this coefficient to obtain an estimate of the time history under the alternate hypothesis. To appropriately scale the response in time, the normalized time-dependent variance was used in lieu of computing the linear fit at each time a measurement was taken. The summation of all the  $p$ -values was used as a metric for the degree of accuracy of each possible battery heat generation value, which was then maximized to estimate the heat output, which best-fits the data. Table 5 provides the nominal and best-fit offset  $p$ -values for each thermocouple, as well as the summation used as the objective function. The best agreement corresponds to 52% of the original battery heat generation rate, which was not fully characterized prior to the experiment. Given the uniform improvement, it is therefore likely that uncertainty in this quantity explains most of the discrepancy.

Two other observations can be drawn from the results. First, the uncertainty in the model is greater than that in the experimental measurements. This is to be expected as the error quantified in the solution verification process is similar in value to the error in the thermocouple measurements. Discretization uncertainty was intentionally driven to this range as there is little to be gained from going further. As a result, the largest contribution to model uncertainty, as captured in the validation study, is derived from uncertain physics. To account for the lack of understanding in some of these properties, their ranges are taken to be larger than the expected actual variation of their properties. Correctly assessing these variations through testing would likely have been prohibitively expensive. Therefore, the model uncertainty arises from the ensemble nature of the study, which mimics the physical variability of the model parameters as well as the degree to which we know them. An ensemble of experiments would be expected to have a smaller uncertainty if the model accounted for the dominant physics processes. This uncertainty enters in the KS statistic in that thermocouples for which the distributions had significant overlap at the level of 1 standard deviation had  $p$ -values orders of magnitude greater than those with larger separations, even if the experimental trajectory was within the bounds of the model ensemble (Tables 6–14).

Using the KS metrics and the geometry of the RIP suggests that there are three 'families' of thermocouple locations that behave similarly. The thermocouples on the batteries and EP (TCs 1–6) show a rapid divergence in the model and experimental data with the model over-predicting the results. On both, the top (TCs 7 and 8) and bottom circuit boards (TCs 9 and 10), responses are more difficult to interpret because the temperatures vary between slight

Table 5: Comparison of KS  $p$ -values for original data and best-fit data offset, including the  $p$ -value summation used as the objective function.

Thermocouple	Nominal $p$ -value	Optimal $p$ -value
TC1	4.4E-08	0.50
TC2	3.4E-07	1.00
TC3	2.3E-06	0.89
TC4	3.2E-06	1.00
TC5	4.4E-08	1.00
TC6	2.3E-06	1.00
TC7	0.49	0.27
TC8	0.0023	1.00
TC9	0.27	0.27
TC10	0.13	0.37
TC11	3.0E-05	0.89
TC12	0.0012	0.0075
TC13	1.3E-07	0.63
TC14	1.3E-07	1.00
TC15	5.6E-06	0.77
Summation	0.8987	10.57

over- and under-predictions. Finally, thermocouples on the housing (TCs 11–15) show a similar trend to TCs 1–6 of over-predicting the experimental data. These groupings are in-line with respect to which parameters are most significant from the sensitivity study and suggest that the discrepancies between model and experiment for many of the thermocouples have a common cause (Fig. 9).

To identify other types of model errors or missing physics, first consider the best-fit model/experiment comparisons shown in [20], which also include the numerical uncertainty in the model error bars. The largest visual discrepancies (based on engineering judgment) occur for TCs 7, 9, and 10, with moderate discrepancies in TCs 3 and 12. Reviewing the  $p$ -values in Table 4, there is good correspondence between the  $p$ -values and visual inspection, although the best-fit  $p$ -value for TC12 is significantly lower with this test, supplying some limited empirical evidence of the utility of this measure. Of the important regions identified during the model building process and test procedure, only TC3 on EP1 was not in statistical good agreement after attempting to correct for the unknown heat generation. This is most likely because EP1 is the main source of heat, other than the batteries, in the system and was not adjusted in the fitting procedure. TCs 7–10 monitor circuit board temperatures and have a stronger sensitivity to the internal convection coefficient as well as the thermal transport properties of the circuit boards. Thermal properties of these boards were homogenized during the model building process as their exact geometries are very complex with layers of plastic and metal. Most importantly, the ratio of metal to plastic was estimated rather than known, suggesting this uncertainty is also important to the model response. TC12 is also particularly sensitive to the internal convection coefficient, so improving this aspect of the model with either experimental measurements or convective heat transfer simulations would add fidelity. Finally, TC3 was the only TC sensitive to

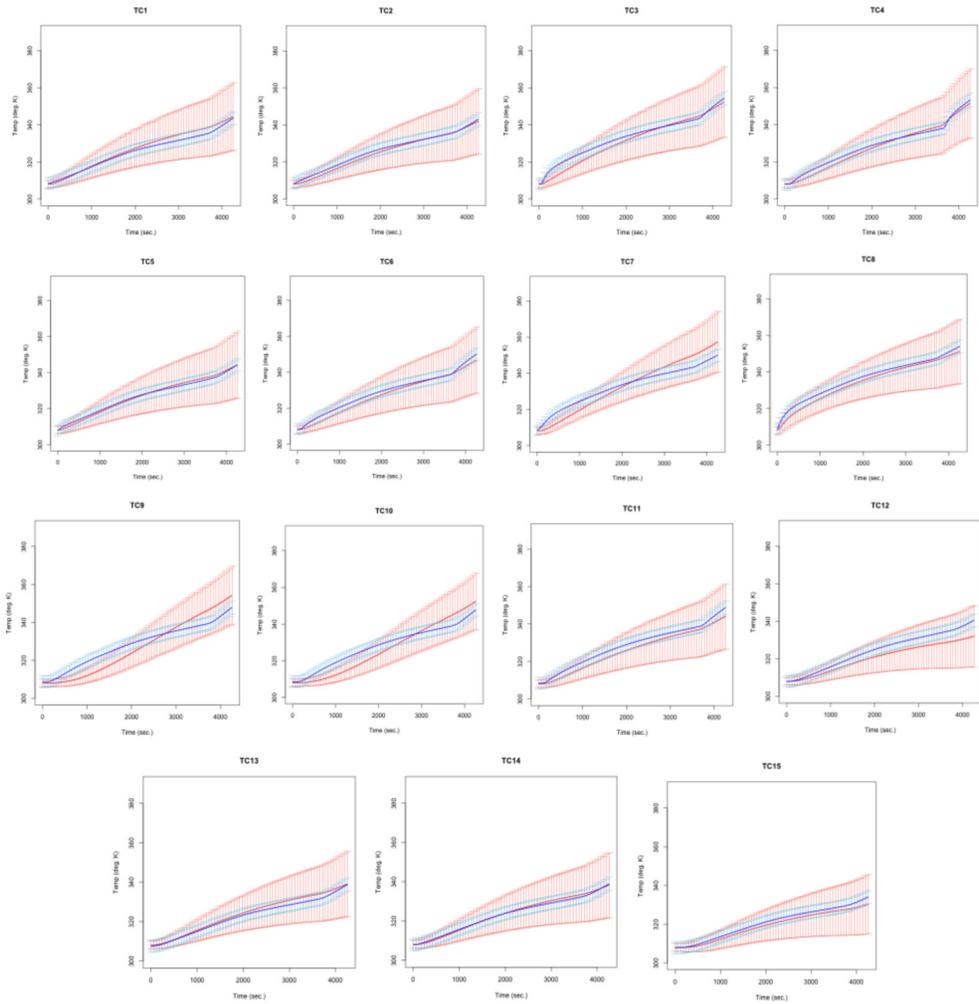


Figure 9: Comparison of regression model with the optimal offset (red) and experimental (blue) thermocouple temperatures.

the material properties of the Al 7075-T6 aluminium, suggesting that greater accuracy of this material model would also be of benefit.

## 5 SUMMARY

A verification and validation analysis was conducted on a finite element heat transfer model of the RIP. Verification consisted of two studies: a mesh resolution study and a numerical parameter sensitivity study. The former is a standard part of solution verification; however, for many models of realistic geometries, it is difficult to obtain the number of meshes necessary to demonstrate convergence rates. In particular, it is often quite challenging to coarsen the mesh. This work presented algorithms for highly automated test mesh generation and refinement. These algorithms enabled nine meshes to be constructed such that convergence rates for  $L_2$  and  $L_\infty$  norms could be identified and the error in important metrics quantified as a function of mesh resolution.

The second verification effort, described in [9], involved performing a LHS study over all the numerical parameters in the model. Custom scripts were written to allow general text replacement needed to modify text-based input files and drive the ensemble of simulations. The study quantified the differences in solutions based on parameter settings and identified the significant numerical parameters, which control both uncertainty and execution time. As a result, the numerical parameters could be chosen to simultaneously control error while minimizing execution time. A conservative estimate of the error introduced into the solution by the numerical parameters and mesh is  $\pm 2$  K. This study, once the infrastructure was prepared, took less than two working weeks to conduct. This time scale was acceptable to the development engineers who requested the thermal model. This paper shows that rigorous model verification can be obtained within the constraints of an aggressive design, development, and production schedule.

Completing an iteration of the V&V process, validation activities were performed on the model to understand its sensitivity to physics and material properties included in the model, following which the model was compared with an experiment of the hardware during physically simulated nominal performance. Test comparisons serve to estimate the uncertainty in the model relative to the experiment, suggest improved parameter values for uncertain physical parameters in the model, and assess the impact of unmodelled physics (both known unknowns and unknown unknowns) on the quantities of interest. While there are 57 physical parameters in the model, the sensitivity study indicated that the power generated by EP1 and the batteries are the most important uncertainties to minimize to improve the predictive fidelity of the model for engineering needs.

### 5.1 Future work for the RIP

It should be emphasized that V&V is an iterative process for which it is difficult to determine what constitutes 'completion'. In practice, the decision of whether or not to continue the process is based on a cost/benefit analysis comparing the work needed to improve the model versus the improved predictive power it will have for quantities of interest. Examples can include enhancing the model's geometric fidelity, choosing mesh spacing or numerical parameters to reduce error, performing characterization tests to improve constitutive models, or adding additional physics to the model. This paper has presented a single iteration through this process. Based on this iteration, the model could be most improved by developing a more accurate model of heat output in the electronics packages based on their temperature and electrical load. The next most beneficial activity would be to better characterize the circuit boards to improve their parameterized material model. Assessment by the system design team will determine if the current model, with an empirically tuned power output, will be sufficient to use to support their design activities. If so, further model improvements will not be made unless dictated by development needs. Otherwise, the cost associated with the testing program needed to bring the model up to the required fidelity will be used to drive decisions regarding future model building efforts.

### 5.2 Conclusions

This paper has presented a case study demonstrating a proposed approach to validation of computational heat transfer models that seeks to establish confidence in the ability of the model to predict quantities of interest using UQ and statistical analysis. The motivating goal was to communicate a more rigorous methodology than those previously employed in a

manner that will aid practitioners. By quantifying uncertainty at each step in the process, the proposed approach can provide reasonable approximations for uncertainties arising from equation discretization and variation in included physical parameters. Both are useful to the analyst to help assess the quality of the model as well as the types of behaviour one might expect from the real device. Importantly, physical and numerical parameters leading to uncertainty in the outputs can be at least partially identified, enabling feedback to improve the model through either enhancements to the computational discretization or further testing to better quantify material properties, etc.

The other benefit of this proposed framework is that a body of evidence for assessing the model quality is produced as an outcome of validation activities. Therefore, quantitative evidence can be provided to support or reject the ability of models to predict quantities of interest in relevant operating conditions. Precise metrics enable improved characterization of small numbers of important physical parameters in the common situation that not all model inputs are completely characterized. However, it is the opinion of the authors that much work remains to be done in this area. For example, the statistical metrics here do not have their usual interpretation in terms of rejecting a null hypothesis with respect to the physical model because it has too many parameters and governing equations that are required to produce its output probability distributions. Future efforts will be directed to obtaining more fine-grained statistical measures such that the statistical confidence in each model component can be assessed, followed by agglomeration to compute metrics for the entire model. Another important statistical method that is needed is the ability to separate distinct versus coincident sources of uncertainty. For example, the variation of material properties between the model and experiment would be expected to be similar for a high quality model, whereas thermocouple errors are unique to experiments and not present in the model. The ability to account for these parts of the error in statistical tests would both lead to better validation metrics and have the potential to provide evidence regarding whether or not unmodelled physics are influencing the system's response in a statistically significant manner.

#### ACKNOWLEDGEMENTS

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the United States Department of Energy under Contract DE-AC04-94AL85000. Funding for this work was partially provided by the Sandia Advanced Simulation and Computing (ASC) Program and its support is gratefully acknowledged. This document is SAND2014-0011J.

#### APPENDICES

The Appendices can be found online at the following link <http://www.witpress.com/journals/cmcm>

Appendix A: SENSITIVITY STUDY RESULTS

Appendix B: MATERIAL PROPERTIES AND UNCERTAINTIES

#### REFERENCES

- [1] Sargent, R.G., Verification and validation of simulation models, *Proceedings of the 2005 Winter Simulation Conference*, pp. 130–143, 2005. doi: <http://dx.doi.org/10.1109/wsc.2005.1574246>
- [2] Babuska, I. & Oden, J.T., Verification and validation in computational engineering and science: basic concepts, *Computer Methods in Applied Mechanics and Engineering*, **193**, pp. 4057–4066, 2004. doi: <http://dx.doi.org/10.1016/j.cma.2004.03.002>

- [3] Roache, P.J., *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers: Socorro, NM, 1998.
- [4] Oberkampf, W.L. & Roy, C.J., *Verification and Validation in Scientific Computing*, Cambridge University Press: Cambridge, England, 2010. doi: <http://dx.doi.org/10.1017/cbo9780511760396>
- [5] Trucano, T.G., Pilch, M. & Oberkampf, W.L., General concepts for experimental validation of ASCI code applications, *Technical Report SAND2002-0341*, Sandia National Laboratories: Albuquerque, NM, 2002. doi: <http://dx.doi.org/10.2172/800777>
- [6] Oberkampf, W.L. & Trucano, T.G., Verification and validation in computational fluid dynamics, *Progress in Aerospace Sciences*, **38**, pp. 209–272, 2002. doi: [http://dx.doi.org/10.1016/s0376-0421\(02\)00005-2](http://dx.doi.org/10.1016/s0376-0421(02)00005-2)
- [7] American Nuclear Society, *Guidelines for the Verification and Validation of Scientific and Engineering Computer Programs for the Nuclear Industry*, ANSI/ANS-10.4-1987, American Nuclear Society: La Grange Park, IL, 1987.
- [8] Roy, C.J., Review of code and solution verification procedures in computational simulation, *Journal of Computational Physics*, **205**, pp. 131–156, 2005. doi: <http://dx.doi.org/10.1016/j.jcp.2004.10.036>
- [9] Scott, S.N., Templeton, J.A., Ruthruff, J.R., Hough, P.D. & Peterson, J.P., Computational solution verification applied to a thermal model of a Ruggedized Instrumentation Package, *WIT Transactions on Modeling and Simulation*, **55**. doi: <http://dx.doi.org/10.2495/cmем130021>
- [10] Logan, R.W. & Nitta, C.K., Comparing 10 methods for solution verification, and linking to model validation, *Journal of Aerospace Computing, Information, and Communication*, **3**, pp. 354–373, 2006. doi: <http://dx.doi.org/10.2514/1.20800>
- [11] Oberkampf, W.L. & Barone, M.F., Measures of agreement between computation and experiment: validation metrics, *Journal of Computational Physics*, **217**, pp. 5–36, 2006. doi: <http://dx.doi.org/10.1016/j.jcp.2006.03.037>
- [12] Sierra Core Team, *Sierra Thermal/Fluids Code*, Sandia National Laboratories: Albuquerque, NM, 2012.
- [13] Pro Engineer Core Team, *Pro Engineer WildFire 5*, PCT: Needham, MA, 2009.
- [14] Cubit Core Team, *Cubit: Geometry and Mesh Generation Toolkit*, Sandia National Laboratories: Albuquerque, NM, 2011.
- [15] Adams, B.M., Bohnhoff, W.J., Dalbey, K.R., Eddy, J.P., Eldred, M.S., Gay, D.M., Haskell, K., Hough, P.D. & Swiler, L.P., DAKOTA: a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: version 5.0 user's manual, *Sandia Technical Report SAND2010-2183*, Sandia National Laboratories: Albuquerque, NM, 2011.
- [16] Incropera, F.P. & Dewitt, D.P., *Introduction to Heat Transfer*, 4th edn., John Wiley and Sons: New York, NY, 2002.
- [17] Hughes, T.J.R., *The Finite Element Method*, Dover Publications: Mineola, NY, pp. 52–56, 2000.
- [18] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing: Vienna, Austria, 2012.
- [19] Pearson, K., On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philosophical Magazine Series 5*, **50**, No. 302, pp. 157–175, 1900. doi: <http://dx.doi.org/10.1080/14786440009463897>

- [20] Saltelli, A., Chan, K., Scott, E.M. eds., *Sensitivity Analysis*, Wiley Series in Probability and Statistics, Wiley: Chichester, West Sussex, England, 2000.
- [21] Saltelli, A., Tarantola, S., Campolongo, F. & Ratto, M., *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, John Wiley & Sons: Chichester, West Sussex, England, 2004.
- [22] Speed, T.P., What is an analysis of variance? (with discussion), *Annals of Statistics*, **15**, pp. 885–941, 1987. doi: <http://dx.doi.org/10.1214/aos/1176350472>
- [23] Lehmann, E.L. & Romano, J.P., *Testing Statistical Hypotheses*, Springer Science and Business Media, New York, NY, 2005. doi: <http://dx.doi.org/10.1007/0-387-27605-x>
- [24] Kolmogorov, A.N., On the empirical determination of a distribution function, *Giornale dell'Istituto Italiano degli Attuari*, **4**, pp. 83–91, 1933.
- [25] Smirnov, N.V., On the estimation of the discrepancy between empirical curves of distribution for two independent samples, *Bulletin of Moscow*, Vol. 2, pp. 3–16, 1939.
- [26] Helton, J.C. & Davis, F.J., Sampling-based methods for uncertainty and sensitivity analysis, *Technical Report SAND99-2240*, Sandia National Laboratories: Albuquerque, NM, 2000. doi: <http://dx.doi.org/10.2172/760743>
- [27] Iman, R.L. & Shortencarier, M.J., A Fortran 77 program and user's guide for the generation of Latin hypercube samples for use with computer models, *Technical Report NUREG/CR-3624, SAND83-2365*, Sandia National Laboratories: Albuquerque, NM, 1984. doi: <http://dx.doi.org/10.2172/7091452>