SAND2009-2106C

Application of a Pragmatic Interval-Based "Real Space" Approach to Fire-Model Validation Involving Aleatory and Epistemic Uncertainty[1]

Vicente Romero[2], Anay Luketa, Martin Sherman
Sandia National Laboratories,[3] Albuquerque, NM

This paper applies a pragmatic interval-based approach to validation of a fire dynamics model involving computational fluid dynamics, combustion, participating-media radiation, and heat transfer. Significant aleatory and epistemic sources of uncertainty exist in the experiments and simulations. The validation comparison of experimental and simulation results, and corresponding criteria and procedures for model affirmation or refutation, take place in "real space" as opposed to "difference space" where subtractive differences between experiments and simulations are assessed. The versatile model validation framework handles difficulties associated with representing and aggregating aleatory and epistemic uncertainties from multiple correlated and uncorrelated source types, including:

- **experimental variability from multiple repeat experiments**
- **uncertainty of experimental inputs**
- **experimental output measurement uncertainties**
- **uncertainties that arise in data processing and inference from raw simulation and experiment outputs**
- **parameter and model-form uncertainties intrinsic to the model**
- **numerical solution uncertainty from model discretization effects.**

The framework and procedures of the model validation methodology are here applied to a difficult validation problem involving experimental and predicted calorimeter temperatures in a wind-driven hydrocarbon pool fire.

1. Introduction

One aspect of the work that Sandia National Laboratories performs for the U.S. Dept. of Energy is the design and assessment of safing systems that keep nuclear weapon firing systems inert in accidents and abnormal environments such as fires. (See e.g. reference [1] for a synopsis of weapon risk considerations and assessment methodology for fire accidents.) In a supporting activity, the model validation effort described in this paper centers around the steady-state temperature response of a fire-heated "cone" calorimeter.

---

[1] This paper is a work of the United States Government and is not subject to copyright protection in the U.S.
[2] AIAA Senior Member, corresponding author: vjromer@sandia.gov, 505-844-5890
[3] Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Reference [2] describes in detail the experimental setup, fire tests, and raw experimental data and results. Reference [3] discusses the geometry and experimental conditions important to modeling the fire tests. Also documented are the physics and discretization models in the FUEGO[4] computational fluid dynamics, combustion, and heat-transfer simulations and results. The massively parallel 3-D code involves 33 degrees of freedom at each node, representing x-y-z momentum, participating-media radiative transport directions or ordinates, chemical reactions/combustion, and turbulence, energy, pressure, and soot and chemical species conservation and transport.

The subject of this paper is validation of the FUEGO fire dynamics model. The following one-sentence contemporary definition of model validation, and close variants of it, are the accepted standard definition in [5] - [8]: *Model Validation is the process of determining the degree to which a computer model is an accurate representation of the real world from the perspective of an intended use of the model*. Despite broad agreement on this one-sentence definition, there is room within the definition for considerable debate among validation methodology developers and practitioners as to what procedures or steps lie within the scope of a completed model validation activity. For example, some consider validation to involve just the discernment of model accuracy versus reality, for those specific output quantities of the experiment and model that are relevant to *intended use of the model*, e.g. for predicting fluid drag on a vessel. Others hold that model validation further implies the discernment of model adequacy for *intended use of the model* (i.e., whether the model accuracy is deemed sufficient for making predictions to support some design, analysis, or decision-making need).

This paper takes the broader viewpoint, and to avoid confusion often refers to the adequacy component of model validation as *model affirmation*. Hence, the definition of model validation subscribed to here is (from [10]): model validation is the discernment of the accuracy and adequacy of models and model predictions as compared to reality (i.e., some subset or filter of reality that is important to predict for some purpose—in as far as we can ascertain what reality portends through appropriately designed and controlled experiments at specific validation points in the modeling space).

Not surprisingly, various modeling communities are still working out the technical procedures and criteria for assessing and deciding whether a model is considered validated (affirmed) with respect to a particular intended use. No overriding consensus yet exists. To be sure, several model validation paradigms and methodologies exist in the literature. Many of these are considered in refs. [9] - [12], where it is determined that none of those considered appear to have the full set of enabling features, yet are pragmatically simple enough, to reasonably handle the difficult model validation problem pursued in this paper. Indeed, a newly evolved model validation framework and methodology, developed in the course of model validation practice on several other challenging application problems, is applied in this paper to the very difficult fire dynamics model validation problem.

The fire experiments, simulations, and results are briefly summarized in the following. The majority of the paper is devoted to presentation of a validation-relevant subset of the

experiment and simulation results and uncertainties; processing of these into a form suitable for model validation comparisons; and description of the comparisons along with interpretation within the well-exercised and vetted model validation framework employed. For brevity here, only a representative subset of experimental and simulation results and model validation processing is presented. The full set of data, processing, and results from the validation activity can be found in [3].

2. Experiments and Experimental Conditions

As Figure 1 shows, the experiments involved a flow-through partial "enclosure" that serves the purpose of imitating foreseeable conditions where a fire heats up the walls and roof of a weapon storage or transportation room or container. The walls and roof then radiate heat to the weapon, imparting more heat and thus embodying a more severe heating condition than if no enclosure and only a shallowly engulfing fire (non-optically-thick) is present.

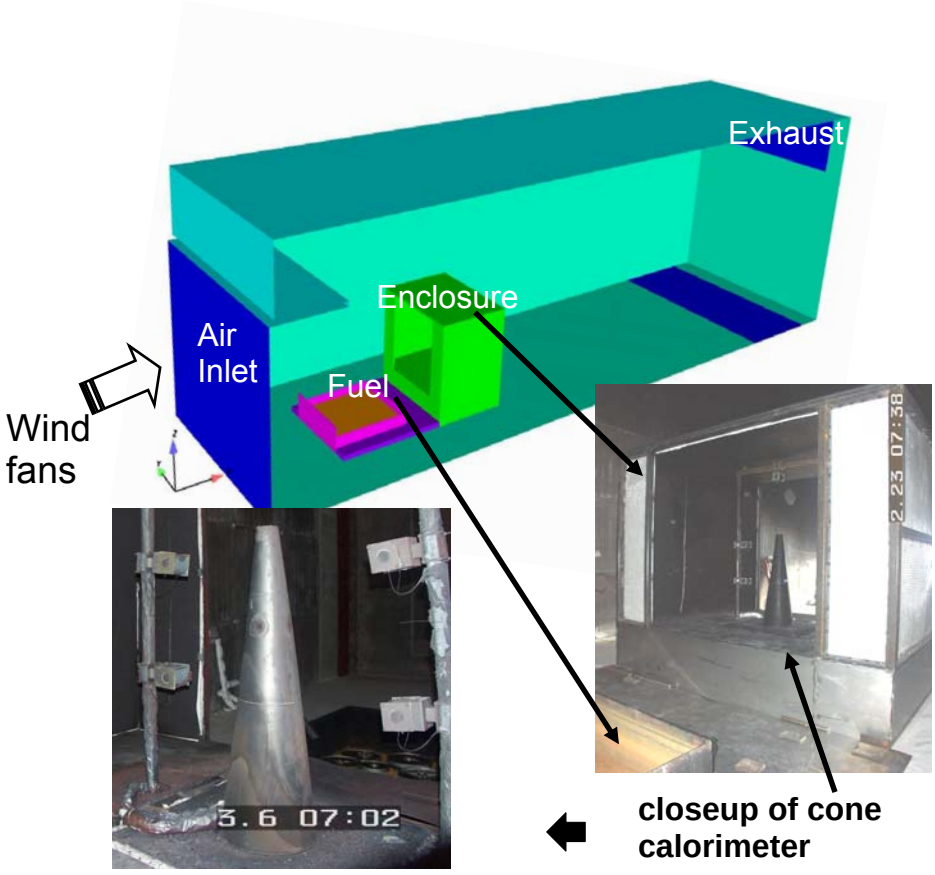

Figure 1: Cone calorimeter inside flow-through enclosure with fuel pan in front. Stands at sides of calorimeter hold thermocouples and (visible) flux gauges.

The walls and roof of the enclosure were comprised of 3/16-inch-thick steel sheet stock on the inside. These were insulated on the outside with blanket insulation to provide an easily modeled boundary condition for the validation activity. Therefore, an adiabatic boundary condition was modeled on the exterior of the enclosure walls and roof in the

FUEGO simulation. The cone calorimeter outside shell is made of 3/16-inch-thick steel sheet. For ease of modeling, the inside volume is filled with blanket insulation to prevent convective and radiative heat exchange among the inside walls of the cone. Hence, an adiabatic boundary condition is applied on the interior of the steel cone shell modeled in the FUEGO simulations. The nearly adiabatic walls (with insulated back-sides) of the calorimeter and enclosure also strongly promote a quicker arrival to steady-state wall temperatures and re-radiation conditions in the experiments.

Thermocouples (TCs) were attached to the inside of the cone shell at 12 height levels that include the two levels shown in Figure 2. TCs and flux gauges were also placed in the flow field of the fire (on the stands shown at the sides of the calorimeter in Fig. 1) and on the walls/roof of the flow-through enclosure that the calorimeter is at the center of. The steel interior walls, roof, and floor of the enclosure, and the exterior of the calorimeter, were painted with black Pyromark high-emissivity paint so that they would readily exchange radiative heat with the fire and with each other, therefore more quickly reaching steady-state temperature and re-radiation conditions. The Pyromarked surfaces have an initial emissivity ϵ of nominally 0.86 upon application to the room-temperature steel surfaces, but the emissivity after experimental "burn-in" of the paint at high temperature is a significant uncertainty in the experiments. Later it will be described how this important uncertainty is accounted for in the model validation analysis. Other uncertain inputs to the experiments will also be discussed.

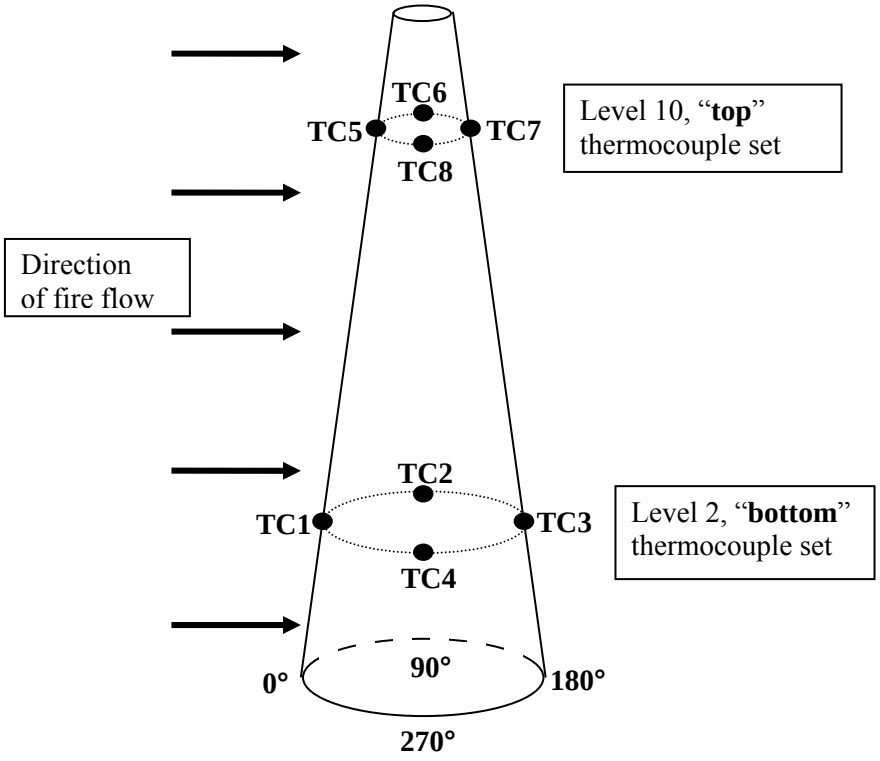

Figure 2: Thermocouple locations on Cone Calorimeter for comparison of experimental results versus FUEGO predicted results.

Details of the fuel and fuel-pan, facility and structure geometries, experiment flow conditions such as the air inflow and exhaust outflow conditions, and all thermocouple and flux-gauge locations can be found in either or both of [2] and [3].

3. Initial Model Validation Considerations

Although a wealth of experimental data from the tests is available to compare against FUEGO predictions for model validation purposes, time and resource considerations dictated that comparisons in the project be confined to only calorimeter temperature response. Such response is most directly associated with the validation question driving the experimental and simulation work here: how well does FUEGO calculate surface heating conditions on objects (weapons) in wind-driven hydrocarbon fires?

To bring the scope of the validation activity into line with project resources, the validation comparisons are further limited to eight diverse and representative TC locations on the calorimeter, although many more locations were monitored in the experiments. Figure 2 shows four TC locations (at 0˚, 90˚, 180˚, and 270˚) on the upper section of the cone above a field joint (internal bolted attachment) at mid-height, and another four at the same angular positions on the lower section of the cone below the attachment joint.

On the upper section of the cone, the Level 10 set of TCs in Figure 2 was chosen in order to minimize local edge or anomaly effects on conduction behavior and therefore temperature of the calorimeter shell. Level 10 is about half way between the top edge of the cone and the height discernable in Fig. 1 where flux gauges inserted into holes of 1.5 to 2 inch diameter in the shell create local anomalies in the conduction and temperature fields. On the lower section of the cone, unknown and possibly highly varying contact resistance around the bolted attachments at the mid-height field joint and the base of the cone could produce local effects on calorimeter temperature. The Level 2 height in Figure 2 was chosen in order to minimize local effects due to these bolted joints.

Calorimeter temperature response is not the only quantity relevant to FUEGO validation. For example, enclosure temperature response is just as indicative of how well FUEGO calculates object surface heating. Certainly, it would be significantly more revealing to compare FUEGO predictions against experimental response for both the calorimeter and enclosure, instead of just one or the other. Both are time-space integrating sensors of the experimental and simulated fires, and corroboration (validation) based on two diverse such sensors is of course stronger than corroboration from one alone. However, corroboration based on two diverse integrating sensors also cannot be said to be *absolutely sufficient* to settle the model validation question with regard to the present experiments. For instance, a further check against the TCs and flux gauges on the stand in the flow field of the fire (see Figure 1) could indicate differently, and refute the fire model's predictive capability locally there. (Such further checks are essentially infinitely prescribable—see discussion in [10] on "field validation" versus the validation of only certain *resultant effects* of the field important to an engineering purpose, "effect validation".)

In fact, absolute sufficiency is probably only reachable in theoretical limits and it would appear that statistical procedures and arguments would have to be invoked to establish validation sufficiency in practical terms (e.g., odds or significance level associated with a validation conclusion). This is beyond the scope of the present treatment. In the present circumstances we can only address whether a *necessary* test of model adequacy is met by FUEGO with respect to the narrow program-relevant quantity of calorimeter steady-state temperature. Indeed, it is shown in the following that a significant validation test is passed in this regard, lending credibility to FUEGO's predictive abilities in the present experimental circumstances, although not sufficient to absolutely confirm its predictive capability—even for the present experimental conditions.

A substantial relaxation in the strength of the validation case also comes from the fact that we compare only steady-state temperature of the calorimeter, as opposed to transient temperature response. This is forced by several reasons. First, capabilities for realistic modeling of transient startup and growth of the pool fire were not yet available in FUEGO. Second, the predominant project driver, assessment of weapon safety risk in fires, is concerned with weapon internal response events that take 10's of minutes to develop and are the result of sustained fire heating and are fairly insensitive to the early-time initial flare-up and stabilization transients of the fire.

Thus, although transient response is a more stringent test of FUEGO predictive capability, project constraints dictated that we could only reasonably compare steady-state temperatures, and this was felt to be acceptable with respect to the driving purpose of predicting heating conditions for weapon-response simulations for risk analysis.

4. Plan of FUEGO Simulations for Uncertainty Quantification in the Model Validation Procedure

FUEGO simulations were used in the validation procedure to quantify modeling uncertainty. Resource constraints dictated that only 5 or 6 steady-state FUEGO simulations could be run at the medium-resolution model described in [3]. (The simulations took several weeks on 512 processors.)

Here model resolution involves spatial-discretization cell size; number of discrete-ordinate directions for solution of participating-media radiation transport; and solver numerical-error tolerance parameters employed in the steady-state FUEGO computations. The resolution level was surmised from past FUEGO modeling projects to be sufficient for the validation purpose here, i.e., to yield simulation results that would change with further resolution refinement only relatively little compared to the effects of other physical modeling uncertainties and bias errors in the validation activity.

To check that the medium-mesh discretization was indeed sufficient, a simulation was planned at a significantly higher resolution. However as explained in [3], the high-resolution simulation never completed because of difficulties with the required multi-thousand processor jobs. Thus, in assuming that the calculation results here are sufficiently numerically resolved, there is only previous experience to cite from [13] which compared results from a similar "medium" resolution to results from a

substantially higher resolution. Nonetheless, it is explained later how the validation framework used in this paper would handle any characterized uncertainty (e.g. by methods presented in [14]) due to numerical resolution in the model and simulations.

Regarding fire-physics modeling uncertainty, a consensus of Sandia fire modeling expert judgment concluded that the following physics submodel forms and coefficients were generically the largest sources of *intrinsic*[4] modeling uncertainty. The nominal values and associated uncertainties are listed in Table 1.

Table 1. Intrinsic Sources of Uncertainty in the Fire Dynamics Model

- heat of combustion (HOC): **44.66kJ/mol ± 10%**
- soot extinction coefficient (SEC): **7 ± 10%**
- convection coefficient (CC) at object surface: **calculated value -50% to +100%**
- flame volume coefficient: (FVC): **2.13 ± 30%**
- flame loading coefficient: (FLC): **0.41 ± 30%**
- turbulence model form: **TFNS (nominal) versus BVG**

To estimate the modeling uncertainty contributed by the above 6 factors, using only the budget of 5 or 6 steady-state FUEGO simulations that could be afforded, information from previous uncertainty/sensitivity studies over these modeling factors was drawn from. The plan was to get reasonable upper and lower bounds on the FUEGO predicted heating of the calorimeter by running a simulation at the combination of parameter values within the preceding parameter space (joint "uncertainty space") that gives the highest heating to the calorimeter, and then at a combination of parameter values within the uncertainty space that gives the lowest heating to the calorimeter.

The sought parameter combinations were identified to the best possible extent under all the presiding constraints and available information as follows. Reference [15] presents a blocked experimental design of 16 simulations that explore the above six-factor uncertainty space plus spatial discretization and computational solver resolution effects. Sixteen simulations were run for "Airplane Fire" and sixteen more for "Truck Fire" wind-driven fires. The sensitivity of weapon heating to the six uncertainties was

[4] This term signifies a category of modeling uncertainty that is innately associated with the "traveling" portion of the model being validated. Only the traveling portion goes on to new predictions beyond the validation exercise. That is, certain elements or aspects of the model will be carried forward to new predictions, and certain aspects are specific only to the validation setting. The intrinsic uncertainties in Table 1 all exist in the validation setting, but also travel to new predictions, as uncertainties that are propagated to simulations results in the new prediction settings as well. Other uncertainties in the experiment, such as emissivity of the calorimeter and enclosure walls in the modeled experiment, are confined to the validation experiments; in general we will simulate new environments and enclosures with different emissive properties and uncertainties than the specially prepared Pyromark painted surfaces in the validation experiments. Thus, these emissivity uncertainties will not travel to new predictions—they are not intrinsic to the FUEGO fire dynamics model. As explained in [12] and demonstrated in this paper, the validation framework handles traveling uncertainties differently from non-traveling uncertainties.

analyzed. From the sensitivity results, the parameter combinations for high and low weapon heating[5] in each accident scenario were inferred.

The investigation in [15] employed a standard k-epsilon turbulence model and alternatively a Buoyant Vorticity Generation (BVG) turbulence model to hopefully bound model-form effects of turbulence modeling error in the Reynolds-Averaged Navier-Stokes (RANS) computational fluid dynamics (CFD) model. The investigations in [3] and [13] propose CFD model-form bounds as the BVG RANS formulation and a Time-Filtered Navier-Stokes (TFNS) RANS formulation. See [3] and [13] for detailed summaries of these CFD formulations.

The formulations generally dictate that the BVG-RANS turbulence model has the potential to result in larger and hotter simulated fires than the TFNS-RANS turbulence model due to the inclusion of a baroclinic torque source-term in the k-equation which enhances mixing and therefore combustion in the BVG formulation. Hence, for model validation purposes we could minimally run just one FUEGO simulation with the BVG model and the inferred "hot fire" extreme combination of values of the other uncertain parameters (Set 4 prescribed below); and another simulation with the TFNS model and the inferred "cool fire" extreme combination of other parameter values (Set 3 prescribed below). A third simulation would be run (analogous to Simulation 6 discussed later) to account for the effects of the highly uncertain emissivity of the Pyromarked surface of the calorimeter and enclosure.

Nevertheless, to more thoroughly examine the effect of turbulence model form on predicted heating in the current physical setting, we ran both model forms with the "hot fire" and "cold fire" extreme parameter combinations listed in the first five bullets of Table 1. Thus, the following simulations were run. Table 2 summarizes the run matrix.

Set 1 (Simulation 1) – Baseline Case, TFNS with nominal parameter values in Table 1

Set 2 (Simulation 2) – High Heating parameter combination with TFNS

Set 3 (Simulation 3) – Low Heating parameter combination with TFNS

Set 4 (Simulation 4) – same as High Set #2 above, except with **BVG** turbulence model

Set 5 (Simulation 5) – same as Low Set #3 above, except with **BVG** turbulence model

[5] as a global spatial average over the heated object, but not necessarily locally at all points on its surface

Table 2. Fire Model Input Variations for the Six FUEGO Simulations Run

FUEGO run	turbul. model	HOC	SEC	CC	FVC	FLC	ε_Pyromark
Set 1	TFNS	44.66	7	100%	2.13	0.41	0.86
Set 2	TFNS	44.66+10%	7	100%	2.13	0.41 -30%	0.86
Set 3	TFNS	44.66-10%	5	200%	2.13 -30%	0.41 +30%	0.86
Set 4	BVG	[……same as row 2……]					
Set 5	BVG	[……same as row 3……]					
Set 6	[……same as row 2……]						0.96

The uncertainties of other inputs to the validation experiment(s), that are non-intrinsic to the traveling fire dynamics model, such as calorimeter/enclosure/facility geometry, fuel regression rate, air flow rate (wind velocity), and material properties, must also be considered in the model validation procedure. Of the parameters in this non-traveling category, it was surmised that the only uncertainty of first-order significance impacting calorimeter steady-state temperature was emissivity of the calorimeter exterior and of the enclosure interior.

This emissivity uncertainty is accounted for by running a simulation (#6, below) to quantify the effect of a substantial perturbation from the nominal emissivity $\epsilon_{nominal}$ = 0.86, with all other model inputs held constant. From this, a first-order finite-difference approximation to d(cone_local-steady-state-temp.)/dϵ at every point on the cone surface can be formed. Then the linearized relationship $\Delta T \cong$ [d(cone_local-steady-state-temp.)/dϵ] • ($\Delta\epsilon$) can be used to estimate the steady-state temperature change at a given location on the calorimeter surface for any value of emissivity ϵ different from $\epsilon_{nominal}$.[6] Hence, the local ΔT effect from varying ϵ over its applicable uncertainty range can be investigated analytically with the above relation, where $\Delta\epsilon = \epsilon - \epsilon_{nominal}$. This will be applied later.

The uncertainty range for emissivity of the calorimeter exterior surface and the enclosure interior surfaces (walls, roof, and floor) is taken to be 0.76 to 0.96 based on measured values ([16] - [20]) for both heavily oxidized steel surfaces and burned-in Pyromarked steel surfaces like in the present XTF experiments. Whether heavily oxidized (burned-in) steel, or burned-in Pyromarked steel, the corresponding emissivity is thought to lie within the range cited above. Set 6 employs an emissivity perturbation that goes to the high extreme of the emissivity uncertainty range, 0.96.

Set 6 (Simulation 6) – Same as Set 2, except $\epsilon = 0.96$

[6] For certain reasons not evident in what is presented in this paper, Sim. 6 is a perturbation from parameter set #2. In what follows, the assumption is made that the derivative d(cone_local-steady-state-temp.)/dϵ would be approximately the same for a similar emissivity perturbation from *any* of the simulations # 1 – 5. Although the derivative might actually change non-negligibly if computed by perturbations from the other simulations, this sensitivity could not be investigated because of the limit on the number of FUEGO simulations that could be run.

A less significant but still substantial non-traveling uncertainty in this project is the heat transfer between the bottom of the steel cone and the thermally massive steel floor it was bolted to. Unfortunately, in this project it was not practical to parameterize a contact conductance at this interface into the FUEGO simulations. Instead, a simple adiabatic boundary condition was applied at the bottom of the cone. The consequent error (and uncertainty thereof) remains unquantified in the present effort. Nevertheless, strong arguments are made in Section 6 that neglecting this factor probably does not materially change the final validation conclusions.

Geometry uncertainties may also have been possibly significant, but were not considered for lack of resources. Finally, the uncertainties associated with fuel regression rate, air flow rate, and steel and insulation thermal properties of the calorimeter were judged to have relatively little effect on calculated steady-state temperatures, based on applicable sensitivity studies from [13].

5. Experiment and Simulation Results and Uncertainty Processing for Model Validation Comparisons

Here we consider only a representative subset of validation-relevant experiment and simulation results and uncertainties. We focus on TC5, about 1/5 of the way down from the top of the calorimeter, and facing the oncoming wind-driven fire. The experiment and simulation results and validation processing for the other seven TCs in Fig. 2 can be found in [3].

Figure 3 shows experimental and predicted temperature response curves for TC5. The data for two nominal repeat experiments, #6 and #7, are shown. As described in [2], the previous five experiments were partial trial experiments needed to refine the geometry configuration and experimental conditions, and to "burn in" the Pyromark-painted radiating surfaces. Also plotted on the figure are the results of all six simulations. The simulation results will be discussed in detail later, but it is mentioned here that these were false-transient simulations to reach steady-state temperatures as quickly as possible. Therefore, the transient response for the simulations is plotted on a scale of seconds to reach steady state (whereas the experimental responses are plotted on a scale of minutes to reach steady state). It is also informative to mention here that results for Simulation 5 are falling rather than rising to a steady state because a mistake in one of the parameter values was noticed and corrected mid-way through the weeks-long calculation. Also, only the late-time portion of the Sim. 1 results were saved for plotting.

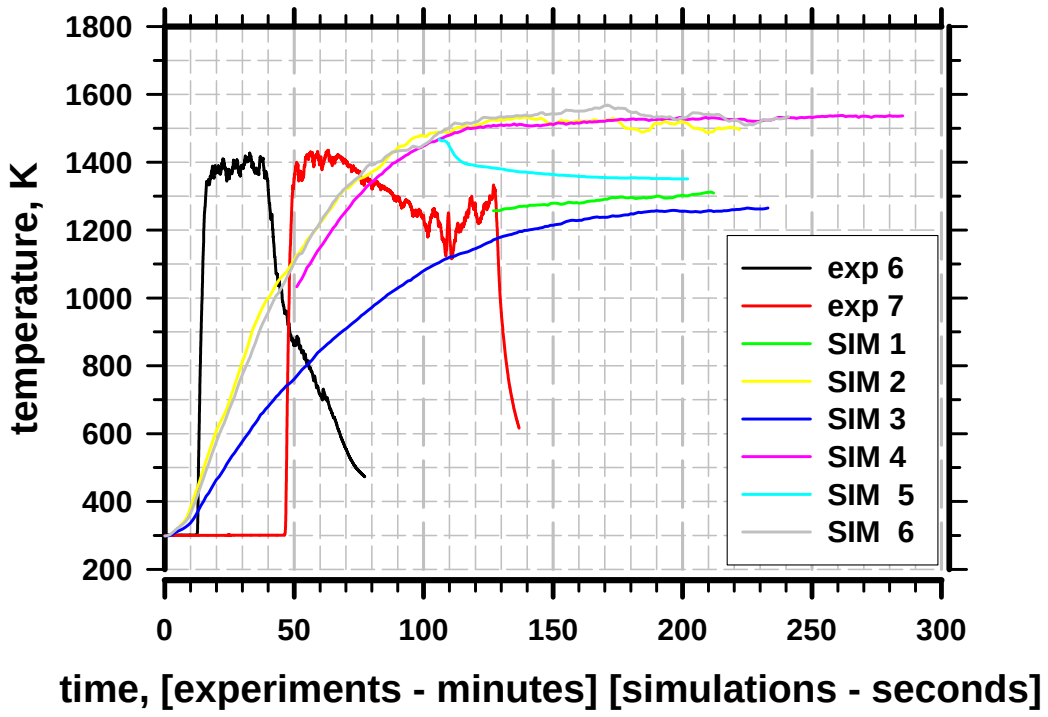

Figure 3. Calorimeter Temperature at TC #5, top - 0° facing the fire

Characterization of Raw Experimental Results: For the experimental results in Figure 3 it is obvious that it is important to use appropriate time windows over which to average TC5 temperature responses in order to get representative steady-state values for model validation comparisons. The time windows used are depicted graphically in Figure 4. These correspond to a six-minute window from 32 - 38 minutes in Exp. 6 and a 13-minute window from 54 – 67 minutes in Exp. 7.

The time-window boxes in Figure 4 have a middle horizontal line that represents the arithmetic mean of the data within the time window, where readings were recorded at 1-second intervals. The means are listed in Table 3. The upper and lower horizontal lines that define the top/bottom of the box are given respectively by the mean +/– 5K. The rationale is that other time-windows than those used here, perturbed wider/narrower and/or later/sooner in time could be just as appropriate or representative. A few such physically-reasonable perturbations were examined. These resulted in changes in time-window means of a few degrees K or less. Accordingly, a deemed-reasonable uncertainty band of +/– 5K was assigned to the mean temperatures (representative experimental steady-state point values) in Table 3. As additional information, the table also lists maximum and minimum *instantaneous* temperatures over the time windows, but these should not be mistaken to represent reasonable uncertainty bounds on the steady-state *means* over the time windows.

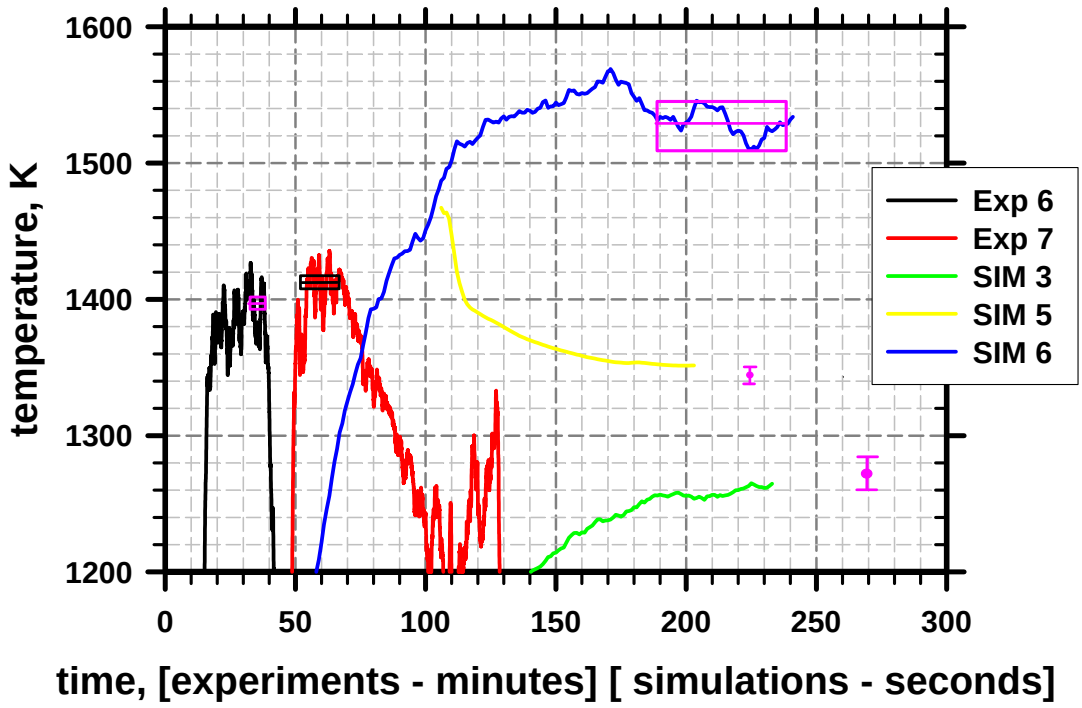

Figure 4. Uncertainty on steady-state experimental and simulation results at TC 5.

Table 3. Mean, Max., and Min. Temperatures over Time-Windows in Figure 4

	Time-window mean, μ	Maximum over time-window	Minimum over time-window
Exper. 6	1394 K or 1121°C	μ6 + 33 K or °C	μ6 - 31 K or °C
Exper. 7	1412 K or 1139°C	μ7 + 24 K or °C	μ7 - 34 K or °C

Characterization of Raw Simulation Results: For the simulation results, representative steady-state values are actually *less* definite in many instances. In particular, Figures 4 and 5 show that the temperature results corresponding to Simulations 1, 3, and 5 have not yet stabilized completely. Simulations 1, 3, and 5 had to be terminated due to resource limitations in the project. (Each simulation took several weeks on 512 processors.) Consequently, results were extrapolated to estimated steady-state values. The procedure involved visually extrapolating a TC curve to form reasonable uncertainty bands of plausible upper and lower bounds on the asymptotic steady-state value, and then taking the midpoint of the bounds as the single most representative point-estimate for steady-state temperature. For example, for Sim. 3 it is visually judged that the steady-state asymptote of the temperature curve is within an uncertainty bar plotted in Fig. 4 that runs

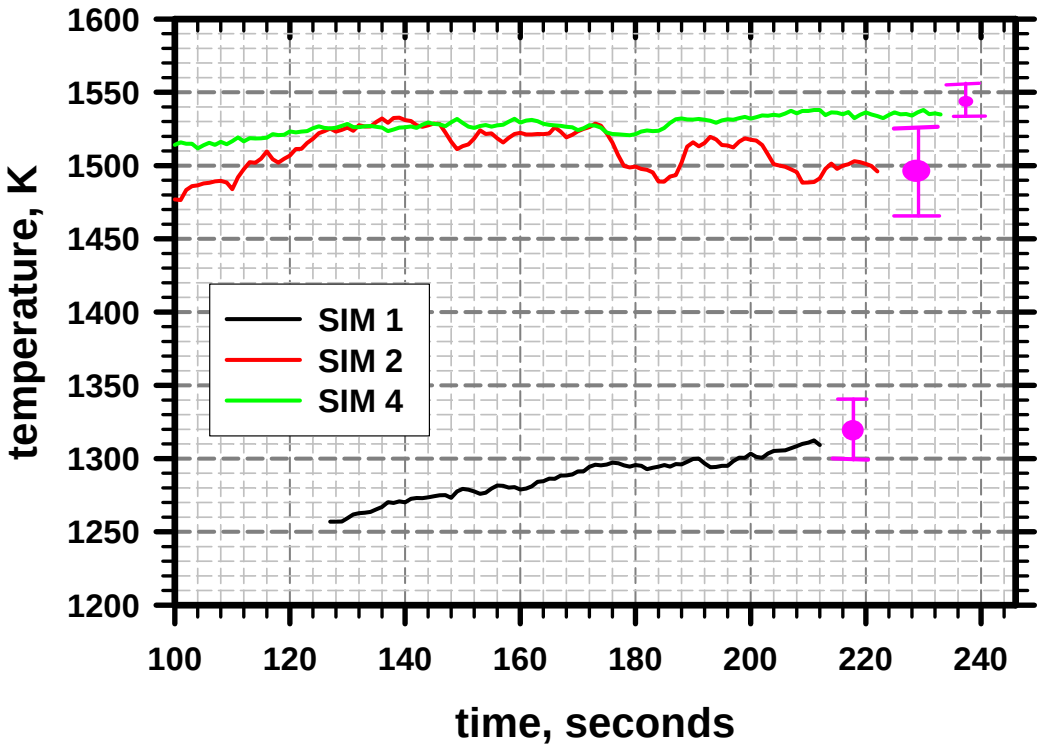

Figure 5. Uncertainty on steady-state simulation results at TC 5.

from 5K below the last plotted temperature from the simulation, to 20K above the last plotted temperature. This information is recorded in Table 4. These values of -5K and +20K are added to Sim. 3's last plotted temperature in Table 5 to get the Min. and Max. values listed for Sim. 3 in Table 5. The Min. and Max. values correspond to the lower and upper extents of Sim. 3's uncertainty bar shown in Fig. 4. The midpoint of the uncertainty bar for Sim. 3 coincides with the 'midway' temperature listed for Sim. 3 in Table 5.

Table 4. Values in brackets [] are visually projected limits to anticipated change from end-of-simulation temperatures to asymptotic steady-state values. For Simulation 6 the averaging time-window is listed from which a mean estimate for steady-state temperature is determined.

Sim. 1	Sim. 2	Sim. 3	Sim. 4	Sim. 5	Sim. 6
[-10,+30] K	[-30,+30] K	[-5,+20] K	[-5,+20] K	[-15,0] K	190 - 241 sec

Table 5. End-of-simulation temperatures at TC5 location, with mean or extrapolated-asymptotic steady-state values and uncertainty bounds

FUEGO run	ending T (K)	Max. (K)	Min. (K)	Central estimate (K)
Sim. 1	1309	1339	1299	1319 (midway)
Sim. 2	1496	1526	1466	1496 (midway)
Sim. 3	1265	1285	1260	1273 (midway)
Sim. 4	1537	1557	1532	1545 (midway)
Sim. 5	1352	1352	1337	1345 (midway)
Sim. 6	1534	1546	1510	1529 (mean)

For Simulations 2, 4, and 6, temperature results did plateau to quasi-steady values before the simulations were terminated. Processing Sim. 6 first, an estimate of steady-state temperature was obtained by averaging over the time window shown in Fig. 4 and listed in Table 4. However, unlike for the experimental data, no empirical investigation was performed to determine sensitivity of the time-window mean temperature (as the steady-state temperature estimate) to perturbations of the window wider/narrower and/or later/sooner in time. Instead, uncertainty bounds listed in Table 5 are the maximum and minimum instantaneous temperatures over the time window. These are also depicted by the upper and lower extents of the time box in Fig. 4.

This latter approach is an easy way to get bounds, but is thought to give considerably exaggerated uncertainty magnitudes for simulation results that have plateaued to a (perhaps noisy) steady state. Indeed, the time-window maximum and minimum for the experimental data are found in Table 2 to be roughly 25K to 35K higher or lower than the time-window mean, yet experimentation with different plausible time windows on the steady-state portion of the experimental data indicated that a more reasonable uncertainty to ascribe to the time-window means is ± 5K. In comparison, the upward and downward instantaneous differences from the mean for Sim. 6 in Table 5 are roughly 20K—considerably less than the 25 to 35K differences from experimental time-window means. Therefore, if the experimental and simulation results have proportionate sensitivities of their time-window means to different plausible time windows that could reasonably have been chosen, then something less than ± 5K would be more reasonable to assign for the uncertainty of steady-state temperature for Simulation 6.

Indeed, the different vertical extents of the three time-window boxes in Fig. 4 portray an exaggerated uncertainty regarding the Sim. 6 results. For Experiments 6 and 7 and for Simulation 6 the scale of temperature oscillation over their time windows is roughly the same, yet the vertical extents of the time boxes, which represent the uncertainty ascribed to the time-window means as estimates of steady-state temperature, are vastly different.

Hence, it is reasonable to conclude that an easy and conservative approach to bounding the uncertainty on time-window means, as an alternative to the labor intensive approach

of sampling different reasonable time windows (recall that we had 7 other TCs to process), is to use instantaneous maximum and minimum temperature over the time window. However, this likely exaggerates the uncertainty by a considerable amount. Because of this, it was concluded that no significant advantage results from the processes of determining a representative time window and then computing the mean within it, versus the much simpler approach of treating plateaued simulation results with the same visual procedure that is applied to the non-plateued results. Since an extrapolation does not have to be made when the results have already plateued, even if noisy, it is much easier to make a visual judgment for reasonable upper and lower bounds on steady-state temperature by considering the character of the response curve and the level of noise toward the end of the curve. This was done for Simulations 2 and 4, with the results listed in Tables 4 and 5.

Note that the error bars and time-window boxes are drawn in the figures so that readers can judge for themselves whether the subjectively determined uncertainty ranges on steady-state temperatures are reasonable. The error bars (and vertical extents of the plotted time boxes for cases of averaging over time windows) are only accurate to within "hand drawn" accuracy. These give a qualitative sense of the magnitudes of the quantities involved, but for quantitative accuracy the tables should be consulted. Regardless of how the uncertainties are determined, in the context of a time-windowing procedure or a visual extrapolation procedure, these uncertainties are referred to as "*graphical processing uncertainties*" in the following. We have seen that both the experimental and simulation results possess significant graphical processing uncertainties.

Uncertainty Processing of Experimental Data for Model Validation Comparisons

The experimental data is here processed in a specific manner for model validation comparison to simulation results according to the validation paradigm and procedures detailed in [12]. There are many other model validation approaches of various flavors that one could pursue (see [10] for a sampling of references on this subject). However, the one applied here appears to uniquely have the required features to handle all the difficult attributes of the current FUEGO validation problem.

Figure 6 presents the perceived dominant experimental uncertainties (from a model validation perspective) concerning steady-state temperature at the location of TC5. The sizes of the error bars in the figure are approximately to scale for the numerical magnitudes denoted in the figure.

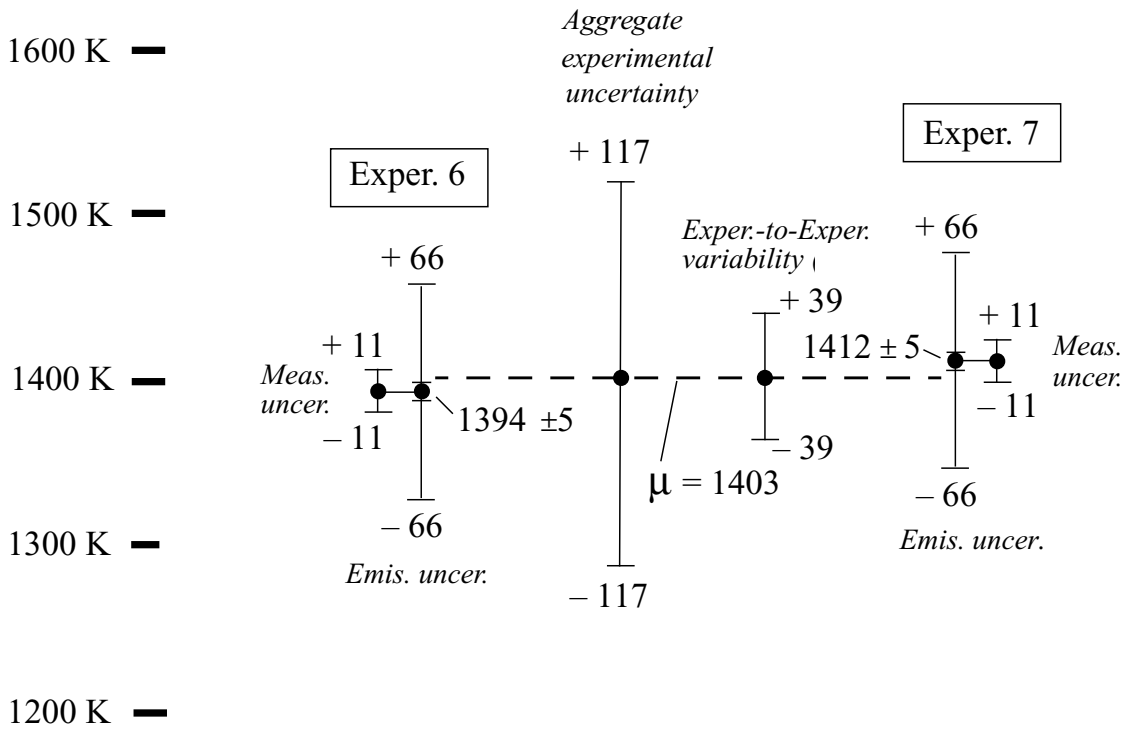

Figure 6: Experimental uncertainties and rollup to aggregate experimental uncertainty at TC5. All temperatures are in degrees K.

We first consider Experiment 6. From Table 3 it has a steady-state mean temperature of 1394K over the applicable time window. It was previously established that this mean is subject to a ± 5K graphical uncertainty, thus the designation 1394K ±5K in the figure.

The ± 11K measurement uncertainty indicated in the figure is an amalgamation of several sources. First, manufacturing variability of the 1/16-inch diameter ungrounded-junction MIMS Type-K TCs used in the experiments is specified by the manufacturer to result in a "two-sigma accuracy of 2.2°C or 0.75% of reading in °C, whichever is greater" (see [21]). This accuracy level is said to exist over a temperature measurement range from 200°C to 1000°C. The wording is generally interpreted (e.g. [21]) to imply that, although a given TC's particular transducing error ε (= $T_{indicated} - T_{true}$) is not known, the relative likelihood of what the error might be is governed by a normal distribution with standard deviation σ = ½(2.2°C or 0.75% of reading in °C, whichever is greater). (See shaded distribution associated with note ② in Figure 7.) Hence, the probability that the absolute value of the error is less than 2σ is given by integrating the shaded portion of the normal distribution that lies within ±2σ of $T_{indicated}$. This integration yields a value slightly larger than 0.95. Accordingly, >95% probability exists that the given TC's error has a magnitude less than 2σ; $|T_{indicated} - T_{true}| < 2\sigma$. Unfurling this inequality yields: $T_{indicated} - 2\sigma < T_{true} < T_{indicated} + 2\sigma$. Hence, >95% probability exists that the true bead temperature lies with ±2σ of the indicated bead temperature.

The above characterization is presumably obtained with a measurement system (particular data acquisition system, length of TC leads, length of electrical signal cables, etc.) that is properly calibrated for accuracy. Bias error is then introduced by the different particulars of the Sandia measurement system relative to the manufacturer's system. Characterization of the Sandia measurement TC channels in e.g. [2] and [21] has determined that channel accuracies usually range well within 1°C even for high-temperature calibration signals (standards) emulating 1000°C (1273 K), in the neighborhood of calorimeter temperatures in the present tests. With the error characterization from these activities, individual TC readings in the experiments could be bias-corrected, but usually are not because of the relatively small errors involved. Instead, a "blanket" uncertainty is assigned to the reported data to cover any error due to the Sandia measurement system. We here assign an interval uncertainty of ± 0.25% of °C reading. This equates to greater than ± 2.5°C in the present case, representing a reasonable upper bound on measurement system errors traditionally seen in assessments against calibration standards at these elevated temperatures.

Hence, two significant sources of uncertainty are surmised to exist with regard to face-value temperature readings in the present experiments. Figure 7 helps illustrate the uncertainty sources and their implications. The probability density functions (PDFs) in the figure represent the uncertainty in true temperature due to TC-to-TC variable error (*aleatory* uncertainty) characterized by the manufacturer, assuming channel bias in their measurement system is corrected using a reliable calibration standard. The depicted potential shifts in the PDF (and therefore where the true temperature is expected to lie) are caused by potential differences between the manufacturer and Sandia measurement systems. Uncertainty in the magnitude of the systematic differences is here taken to be an interval (*epistemic*) uncertainty of ± 0.25% of reading in °C.

Accordingly, if no Sandia measurement-system bias exists, the true TC bead temperature would be expected with >95% probability to lie within the range marked by the smaller error bar in the figure. However, when Sandia measurement-system bias does exist, and its value is characterized to lie e.g. within an uncertainty range of $-\Delta^-$ to $+\Delta^+$ as depicted in Figure 7, then the said PDF uncertainty (and therefore where the true temperature is expected to lie) might be shifted up or down by any value within the extremes shown.[7]

[7] Under such shifts, the superposition spoken of here is a simplifying approximation, but is reasonable for the present circumstances and validation assessment. True superposition requires that the PDF in Figure 7 is independent of the second uncertainty that precipitates the shifting, and that the distribution remain unchanged or invariant as it shifts. The independence condition is met here and in many other real settings. However, the invariance condition is not strictly met here, and is probably difficult to strictly meet in general. Here, the standard deviation (thus width) of the distribution is posed as a percentage of temperature reading. Since temperature magnitude changes as the distribution shifts vertically over the $[-\Delta^-, +\Delta^+]$ range in Figure 7, the width of the distribution would change as well, violating the strict superposition conditions. However, this non-invariance effect is not large enough to be material to the validation conclusions here. If the non-invariance effect *is* significant and therefore important to capture, a Monte Carlo procedure like that demonstrated in [22] can be employed.

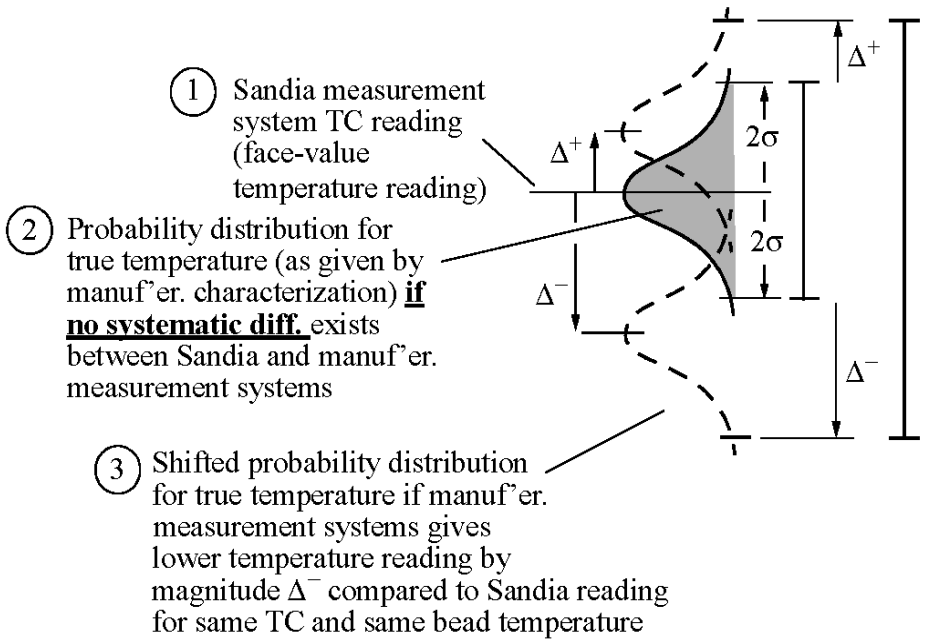

Figure 7: Uncertainty of TC reading error is governed by probabilistic uncertainty (PDF) and potential systematic shifts in the PDF. Aggregate uncertainty interval shown at right is given by shifting either PDF or its interval representation shown, through the uncertainty range $[-\Delta^-, +\Delta^+]$ of the potential error of Sandia meas. system relative to TC manuf. meas. system (that manuf. supplied PDF is referenced to).

Thus, the region within which the true temperature is expected to lie (at the 95% probability level) is given by the large error bar at right in Figure 7. The lower and upper extremes of the large error bar are formed by simply adding the respective extremes of the PDF and interval source uncertainties: $[L,U] = [(-\Delta^- + -2\sigma), (\Delta^+ + 2\sigma)]$.

For the present case we get [L,U] = [(-0.25% + -0.75%), (0.25% + 0.75%)] of °C reading = [-1%, 1%] of °C reading. This uncertainty range can also be expressed conventionally as ± 1% of °C reading. Consulting Table 3 for °C reading for Experiments 6 and 7, ± 1% yields the numerically rounded [–11K, +11K] error bars in Figure 6 (analogous to the large error bar in Figure 7).

Usually, a third—and dominant—source of temperature indication error exists, due to TC attachment effects. That is, the TC bead is not usually at the same temperature as the surface it is attached to because of contact resistance effects and heat transfer to or from the bead. See [23] for a detailed presentation of the issues, and quantification of the effects via finite-element modeling. Since it is really the surface temperature that is desired from the measurement, and not actually the bead temperature, any difference between the two is usually considered to be an error in the measurement. Such error has been indicated in Sandia investigations (e.g. [17], [21]) to be as much of 6% of °C reading under certain conditions. Such temperature differences or lags are not considered

in the present activity because steady-state temperature is the quantity of interest. For this quantity, positive and negative temperature lags are presumed to cancel out in the calculation of mean temperature as the instantaneous temperature oscillates noisily about the steady-state mean. Depending on the actual oscillation history, non-symmetric time-weighting of positive and negative lags could lead to non-complete cancellation, but any such effect is likely to be small to negligible. It seems reasonable to assume here that any such effects are small enough to be covered by the significant margin of conservatism in the assigned 0.25% uncertainty discussed previously.

Another source of validation uncertainty in the indicated experimental steady-state temperature is represented by the [–66K, +66K] error bars labeled "*Emis. uncer.*" in Figure 6. It is evident that these are dominant uncertainties in the individual experiments. These arise from an uncertain *input* to the experiments (calorimeter and enclosure emissivity) as opposed to the aforementioned uncertainties in the measurement and processing of the experimental *output* (temperature).

By far, emissivity of the calorimeter exterior and enclosure interior is thought to be the factor that yields the highest sensitivity of calorimeter steady-state temperature to experimental input-factor uncertainty. As previously explained, Simulation 6 at $\epsilon = 0.96$ is a perturbation from Sim. 2 at $\epsilon_{nominal} = 0.86$, all other simulation parameters being the same. From Sim. 6 and Sim. 2 results, a first-order finite-difference approximation to d(cone_local-steady-state-temp.)/dϵ at every point on the cone surface can be formed. Then the linear projection equation $\Delta T \cong [d(\text{cone\_local-steady-state-temp.})/d\epsilon] \bullet \Delta\epsilon$ can be used to estimate the steady-state temperature change ΔT from Sim. 2 steady-state temperature at a given point on the calorimeter surface, for any value of emissivity ϵ different from $\epsilon_{nominal}$ (where $\Delta\epsilon = \epsilon - \epsilon_{nominal}$).

The uncertainty range for emissivity of the calorimeter exterior surface and the enclosure interior surfaces (walls, roof, and floor) is indicated to be $\epsilon = 0.76$ to 0.96 as established earlier. The upward ΔT perturbation at TC5, corresponding to the upper-bound perturbation $\epsilon = 0.96$ from $\epsilon_{nominal} = 0.86$, is given directly by subtracting the central estimate for Sim. 2 steady-state temperature in Table 5 from the central estimate for Sim. 6 steady-state temperature. The result is $\Delta T = +33$K. From the fact that a downward uncertainty perturbation $\epsilon = 0.76$ from the nominal value 0.86 is the same magnitude as the upward perturbation from 0.86 to 0.96, the downward temperature perturbation is −33K by linearity of the projection equation. Therefore, the full uncertainty range of temperature at TC5 due to an emissive uncertainty [0.76, 0.96] is projected to be 66K, given by the uncertainty interval [−33K, +33K].

Indeed, in the next section this interval [−33K, +33K] is an element of the *prediction* uncertainty for steady-state temperature at TC5 (see Fig. 8). Furthermore, since the validation framework applied here assesses how closely the experimental and modeled systems transform experimental inputs to output results, an objective assessment would map any uncertainty of the experimental inputs through the modeled and physical systems, and then compare how output results differ. However, unlike the modeling case, in the physical case we do not have empirical quantification of how the physical system

outputs of calorimeter TC temperatures vary as the Pyromark emissivity varies over its uncertainty range. This experimental knowledge deficit, and the fact that the experimental output results correspond to a specific but unknown input emissivity (although it almost certainly lies within the uncertainty range discussed), have several important implications for model validation and model validation methodology (see [11], [12]).

Accordingly, the present circumstances that the uncertain input of interest is non-traveling (see Footnote 4) and is predominantly systematic over the two experiments, dictate a somewhat non-intuitive treatment ("*data-conditioning*" of the experimental data with respect to the uncertain emissivity) to minimize the chances of committing a "Type X" model validation error.[8] Here the uncertainty added to the experimental results turns out to be [−66K, +66K] as shown in Figure 6. This is twice the uncertainty [−33K, +33K] added to the *prediction* results in Fig. 8.[9]

Finally, we address the test-to-test variability of the experimental results. Consider the uncertainty in Figure 6 associated with the nominal point results of the two experiments, 6 and 7. First note that the measurement and emissivity-related uncertainties are completely correlated. That is, since the same TC and data acquisition system and channel was used for TC5 in the two experiments, any associated bias errors in reading

---

[8] The model validation methodology employed here is skewed toward preventing a "Type X" model validation error ([11]) of an incorrect conclusion of 'no significant model bias' when in fact significant bias does exist but is hidden by systematic uncertainty in non-traveling input factors in the experiments. The drawback is that the framework likely exaggerates the uncertainty "resolution level" within which it can be established that the model emulates the real system. (The tradeoff here is analogous to the situation in statistical hypothesis testing, where the more one chooses to skew (decrease) the odds of incorrectly rejecting a true hypothesis (Type I error), the more likely it is that a Type II error will be committed of incorrectly accepting a false hypothesis. As in hypothesis testing, the presence of uncertainty forces one into a position of having to make a choice of which undesirable outcome is the least undesirable.) Thus, the methodology here favors incurring a "Type Y" error of exaggerating the range of possible model bias relative to what it is likely to actually be, and accepts this tradeoff in preference to incurring a Type X error of underestimating the model bias. This choice is argued in [11] to best support the objectives of Best Estimate + Uncertainty extrapolative predictions with the model.

[9] This factor of two arises when the model is used as an approximation for the experimental change (slope) in TC5 temperature versus change in the input factor (here emissivity). This is just a nominal estimate for what the data conditioning factor actually should be. The factor is formally two times the ratio r of experimental slope to modeled slope. Thus, the factor is two when $r = 1$, i.e., under the nominal approximation of equal experimental and modeled slopes. However, the sensitivity of validation conclusions to uncertainty in r should be investigated as part of the validation procedure. It is determined that the actual ratio of experimental to modeled slopes can be up to $r = 1.5$ (for a factor of 3) before the experimental uncertainty bar in Fig. 8 extends outside the range of prediction uncertainty and changes the validation conclusions arrived at later. The actual experimental slope is expected to be well within this allowable 50% difference from the modeled slope.

vs. true temperature are essentially the same in the two experiments. The same is true of the uncertainty due to emissivity in the experiments. Although the emissivity in the experiments is unknown to within a relatively large range of 0.76 to 0.96, it is reasonable to postulate the experiment-to-experiment differences in emissivity are small comparatively. Hence, the associated uncertainty in experiment 6 and 7 results is closely correlated (systematic over the two experiments).

Conversely, the ± 5K graphical uncertainties associated with the steady-state temperature averaging windows for the two experiments can be considered to be independent and uncorrelated among the two experiments. This dictates that this source of uncertainty, for the purposes of characterizing experiment-to-experiment variability, be treated differently than the ones above.

First, however, consider the instructional case of only one experiment (either 6 or 7). The aggregate uncertainty for the single experiment would be constructed as follows. The [–11K, +11K] measurement uncertainty in Figure 6 can be conceived as being subject to vertical shifting over the range [–66K, +66K] due to the emissive uncertainty. Such shifting is already familiar from the previous discussion pertaining to Figure 7. In the way that the uncertainty ranges are added or superposed in accordance with the assumptions in Footnote 7, the results here would yield [(-11K + -66K), (11K + 66K)] = [-77K, +77K]. This uncertainty bar about the nominal measured temperature would in turn be subject to vertical shifting of ± 5K associated with the steady-state temperature graphical processing uncertainty. An aggregate uncertainty of [-82K, +82K] would result. This would be the case for either Experiment 6 or 7 alone.

Now consider Experiments 6 and 7 together. With reference to Figure 6, the nominal results are respectively 1394 K and 1412 K. The 18 K difference between these results cannot be explained by the ± 5K uncorrelated processing uncertainties in each result. (Recall that the other uncertainties are effectively correlated or systematic between the two experiments, so cannot explain or contribute to any relative differences in the two experimental results.) Therefore, some other explanation lies behind the experimental differences. Certainly, things varied between the two experiments that we could not characterize or explicitly treat in this project due to practical limitations.[10]

---

[10] If we had quantified the input variabilities and could afford to propagate them to the simulation output, just as the experimental system propagated them to the experimental output, then we would have approximately offsetting effects in the validation comparisons to come later. Instead, neglectance of experimental input variabilities on the simulation side may show up as an under-represented uncertainty band in the validation comparison against the aggregate experimental uncertainty. Any consequent "uncertainty shortfall" might then be mapped into selected parameters of the model to add a physical variability effect to it. This so-called "*model conditioning*" ([9] - [12]) arguably supports an objective of Best Estimate + Uncertainty modeling for extrapolative prediction. It will be established later that no such model conditioning is indicated to be necessary in the present activity. Indeed, the model-intrinsic uncertainties of Table 1 will be seen to bound the experiment-to-experiment variability in the physical results.

In any case, we can reasonably posit that if many other repeat experiments were run, the results would vary according to a Normal distribution, as is often the case with complex experimental systems. We can get estimates of what the mean and variance of the Normal distribution would be by calculating these from the 1394 K and 1412 K nominal experimental results. The mean of these is 1403 K as denoted in the figure. The standard deviation S, times two, is 2S = 25 K. It must be kept in mind that this two-sigma magnitude of 25 K only *nominally* corresponds to 95% included probability in the postulated normal distribution of experiment-to-experiment steady-state temperature at TC5.[11]

The nominal treatment also ignores the ± 5K graphical processing uncertainties in the two steady-state temperatures. Because this uncertainty is uncorrelated over the two experiments, a worst-case (largest variance) treatment of this uncertainty involves decreasing the lower nominal result, 1394 K, by the maximum possibility over the applicable ± 5K uncertainty range; and increasing the upper nominal result, 1412 K, by the maximum possibility over its independent ± 5K uncertainty range. This yields adjusted results of 1394K –5K = 1389 K, and 1412K + 5K = 1417 K. The accompanying two-sigma magnitude is S_high = 39.1. This compares to the two-sigma value of 25 obtained in the previous paragraph when the uncorrelated ± 5K window processing uncertainties were not accounted for. A rounded value of 39 (shown in Figure 6) is used in the following.

The effect of the correlation treatment is very significant here. If the ± 5K uncertainties were perfectly correlated among the two experiments, then these would constitute a systematic uncertainty over the two experiments. Then, consistent with Figure 7, the aggregate uncertainty from these two factors (graphical processing and experiment-to-experiment variability) would be to shift the (two-sigma = 25K) normal distribution over a ± 5K range. The result would be [(-5K + -25K), (5K + 25K)] = [-30K, +30K] by the linear superposition approximation (Footnote 7).

Instead, if the ± 5K graphical processing uncertainties are treated as independent and uncorrelated over the two experiments, the result is [-2S_high, +2S_high] = [-39K, +39K] as already determined. Therefore, the added effect of the uncorrelated ±5K

---

[11] The actual standard deviation of a large number of experimental repeats could be very different from the S = 25K/2 calculated from just the two experiments. The small-sample uncertainty in the standard deviation value, and also in the mean value, is not accounted for in this paper. Hence, we cannot state with reasonable statistical confidence that the mean ±2S = 1403K ±25K defines an interval that encompasses 95% of the postulated normal distribution for experiment-to-experiment steady-state temperature variability. We can only state that this interval gives a nominal quantification of the physical variability. In fact, if the small-sample uncertainty on the apparent experimental variability is taken into account, this "uncertainty on the variability" or "epistemic uncertainty on the aleatory uncertainty" is large enough that it can potentially overturn our validation affirmations at some of the eight TC locations. Nonetheless, within the sampling uncertainty, it is also possible that an even stronger affirmation of model validity could occur, depending on the direction of the actual errors in the calculated values $\tilde{\mu}$ =1403K and S = 25K/2.

processing uncertainties is a ±14K increment to the standing ±2S = ±25.0 nominal uncertainty from experimental variability. The ±14K is nearly three times the added effect if the ± 5K processing uncertainties are treated as perfectly correlated (systematic) over the two experiments.

To close out this section, we combine the experimental measurement and emissivity-related interval uncertainties with the normal PDF for experiment-to-experiment variability (mean 1403K and standard deviation S_high = 39K/2). Following the paradigm of Figure 7, the [–11K, +11K] measurement uncertainty (interval) is superposed/added with the PDF uncertainty to get: [(-11K + -39K), (11K + 39K)] = [-50K, +50K]. The interval uncertainty [–66K, +66K] due to uncertain emissivity is combined in by further superposition to get: [(-66K + -50K), (66K +50K)] = [-116K, +116K]. By using the numbers from each term with more decimal-place precision, we get the more accurate result [-117K, +117K] depicted by the uncertainty bar labeled 'Aggregate experimental uncertainty' in Figure 6.

Analogous quantities and results to those presented in Figure 6 for TC5 are presented in [3] for the other TCs.

Uncertainty Processing of Simulation Results for Model Validation Comparisons

Figure 8 shows FUEGO simulation results at TC5 location for Simulations 1 – 5. These were all run with the nominal value of emissivity, $\epsilon_{nominal} = 0.86$. The range of aggregate experimental uncertainty from Fig. 6 is plotted in Figure 8 to lend a sense of scale and location relative to the simulation uncertainty. The central-estimate values of steady-state temperature from Table 5 are printed in the figure beside the plotted filled dots. The maximum and minimum reasonable steady-state temperature bounds from Table 5 are also plotted in Figure 8, as unfilled diamond symbols.

The *nominal* steady-state temperature values indicate that the BVG turbulence model yields greater object heating than the TFNS model—all other simulation parameters being equal. This agrees with expectations.

When the emissivity uncertainty is accounted for, the expectations still hold up. For instance, consider the TFNS-High simulation (#2) and the BVG-High simulation (#4). The emissivity-related [−33K, +33K] uncertainty bars of these two simulations overlap some in Fig. 8. However, recall that these uncertainty bars represent *correlated* or systematic uncertainty over the set of simulations. That is, whatever the true value of emissivity is, it is the same for all simulations. If the true value (or any value) were input to the simulations, it would not yield e.g. a value on the upper portion of the uncertainty bar of Sim. 2 and a value on the lower portion of the uncertainty bar of Sim. 4. Rather, the results would be correlated such they would both lie at closely the same vertical position within each error bar. Therefore, although the error bars of Sim. 2 and Sim. 4 overlap some, giving the *appearance* that it is possible to get a temperature realization from BVG-High Sim. 4 that is lower than a corresponding realization from TFNS-High Sim. 2, this is not really the situation here.

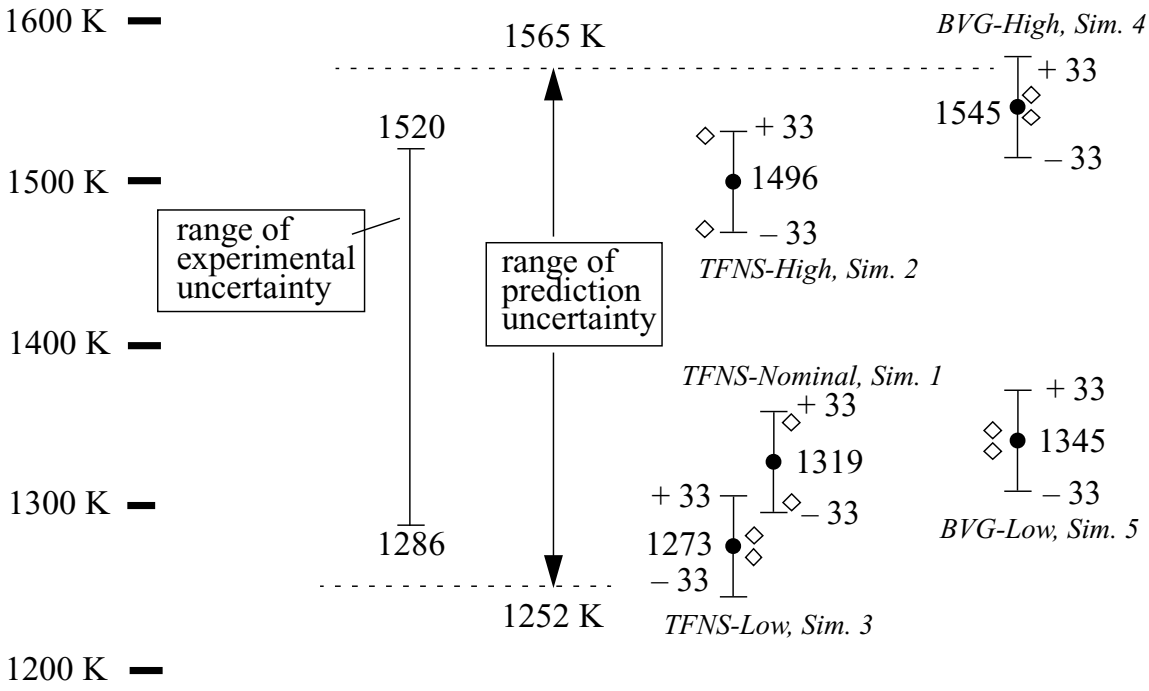

Figure 8: Simulation results and uncertainties at TC5, with range of prediction uncertainty compared against range of aggregate experimental uncertainty. All temperatures in degrees K.

The expected ordering holds even up when the graphical processing uncertainty is accounted for. In contrast to the emissivity related uncertainty, the graphical processing uncertainties (min and max bounds in Table 5) are *uncorrelated* across the various simulation results, and *can* significantly shift simulation results relative to each other. For example, in the extreme worst combination the graphical uncertainty could allow: *a*) a shift of Sim. 4's uncertainty bar downward until its midpoint (filled dot) aligns with the local lower diamond; and *b*) a shift of Sim. 2's uncertainty bar upward until its midpoint aligns with the local upper diamond. Even in this worst case the BVG-High Sim. 4 uncertainty bar remains higher in vertical position than the TFNS-High Sim. 2 bar. This implies a higher simulated temperature for BVG-High than for TFNS-High. Recall also that this is under a worst-case realization of the uncorrelated graphical uncertainties, and that the graphical uncertainty magnitudes are likely highly exaggerated for the simulation results, as discussed earlier.

It is otherwise observed that, for a given turbulence model form (TFNS or BVG), the prescribed parameter sets for Low, Nominal, and High heating give consecutively hotter fires, or at least consecutively greater heating of the calorimeter as indicated at TC5 and the other seven TC locations (see [3]). This ordering is retained at all TC locations under all possible realizations of the graphical and emissivity uncertainties.

These results support the proposition that, for the model validation purposes here (and for extrapolative predictions in general) it is not necessary to expend simulations at the BVG-Low and TFNS-High and Medium parameter sets. These parameter sets (#5, 2, and 1 respectively) routinely yield results that lie between the heating extremes of parameter sets #3 (TFNS-Low) and #4 (BVG-High).

Thus, the effects of the six dominant sources of intrinsic modeling uncertainty in the fire dynamics model (Table 1) can be effectively bounded by running FUEGO simulations at just the two "extreme" parameter sets #3 and #4. Of course, other simulations have to be run to assess the effects of other sources of uncertainty (like uncertain emissivity, numerical discretization, etc.) in a given prediction.

We now turn to the model validation objective of our analysis. Ultimately the aggregate experimental uncertainty in Figure 8 will be compared against an aggregate simulation uncertainty that is appropriate to a validation assessment. The criterion for the model to be affirmed is that the model predictions, with uncertainties properly accounted for, yield an uncertainty band that encompasses the aggregate experimental uncertainty.

Otherwise, the processed experimental uncertainty, as presumably the best empirical evidence of where "reality" lies and where the next experimental result(s) would occur, lies outside the range predicted by the modeling. Thus, the predictions would fall short of spanning and capturing the empirical evidence—whether through improper model form or via misrepresented parameter uncertainties, or both. In any case, it could not be asserted that the model was fully 'valid'. Again, this is the criterion of the present validation framework, but other validation criteria and frameworks are actively being discussed in the literature. No over-riding consensus has emerged yet, but the current framework arguably directly supports a goal of Best Estimate + Uncertainty extrapolative prediction. See [10] for further discussion.

In the present framework, the preliminary aggregate prediction uncertainty in Fig. 8—before accounting for graphical uncertainties—ranges from the low end of the lowest uncertainty bar (given by TFNS-Low Sim. 3), to the high end of the highest uncertainty bar (from BVG-High Sim. 4). That is, the emissivity uncertainty and the six uncertainties in Table 1 have possible combinations or realizations over their joint uncertainty space (where these seven factors are justifiably assumed to be independent of each other) that can yield model predictions which vary from the low end to the high end of the said uncertainty range. This preliminary range of prediction uncertainty is impacted by the graphical uncertainties as explained next.

The graphical processing uncertainty for the BVG-High Sim. 4 results is given by the min and max bounds in Table 5. The graphical uncertainty allows that Sim. 4's uncertainty bar in Fig. 8 can really lie anywhere within an upward or downward shift where its midpoint (filled dot) remains between the upper and lower unfilled diamonds to the side of the uncertainty bar. Analogous freedoms are allowed for the Sim. 3 uncertainty bar to be shifted between the upper and lower diamonds at its side. Recall that

the graphical uncertainties are not correlated with each other, so the allowable shifting of Sim. 4's uncertainty bar is independent of Sim. 3's.

At the upper end of the simulated temperature range, treatment of the graphical uncertainty for Sim. 4 impacts the validation determination relative to the high end (1520 K) of the experimental temperature uncertainty range shown in Fig. 8. Analogous considerations hold for the graphical uncertainty for Sim. 3 and the validation determination at the low end (1286 K) of the experimental uncertainty range. Within the graphical uncertainty, the Sim. 4 uncertainty bar could be translated upward, and the Sim. 3 results could be translated downward, such that validation margins are greater at both the upper and lower ends. The opposite extreme possible combination is a downward shift in Sim. 4 results and an upward shift in Sim. 3 results, such that validation margins decrease at both the upper and lower ends.

A treatment which increases validation margins here, or which tends to create a closer comparison when positive margins like those in Fig. 8 do not exist, is said to be *non-conservative*.[12] This type of treatment could enable the validation criterion to be (falsely) met or approached closer, while the actual value of the factor (for no graphical processing error) might correspond to a more biased model than the validation assessment leads one to believe. Hence, this is one possible way to commit Type X model validation error. Type X error can arise from many other sources, such as model discretization uncertainties, systematic uncertainties of non-traveling experimental inputs (here, emissivity as already discussed), and other sources catalogued in [12].

To guard against the potential for Type X error that the graphical uncertainties pose, the framework takes a *conservative* approach of attempting to eliminate the risk entirely. To do this, the most extreme possible combination is invoked of shifting the Sim. 4 uncertainty bar downward the full allowable amount until its midpoint (filled dot) is beside the lower diamond at its side, and shifting the Sim. 3 uncertainty bar upward the full amount until its midpoint is beside the upper diamond at its side. This maximally decreases the validation margins at both the upper and lower ends of the data range. Hence, this likely causes a Type Y model validation error of the framework exaggerating the perceived extent of potential model bias. Unfortunately, eliminating the risk of Type X validation error comes with a tradeoff of committing a Type Y error. An alternative is

[12] Accounting for simulation uncertainty of the intrinsic modeling factors of Table 1 and for the emissivity uncertainty have the effect of expanding the prediction uncertainty as well. Yet, this expansion is not considered to be non-conservative in the validation formulation. As [12] explains, from a model validation perspective (in the context of extrapolative predictions and hierarchical modeling) these factors are different in nature from the graphical uncertainty, so are handled differently. The intrinsic modeling factors of Table 1 proceed to any new predictions with the (validated) model, so their uncertainty is transported inherently to new prediction results. This is not the case with the graphical uncertainties being discussed here. New graphical uncertainties/magnitudes will be present in new simulation results. For example, if the new simulations are terminated after arriving at a smooth, flat asymptotic steady-state, no graphical uncertainty will be present at all. In terms of the emissivity uncertainty, like the uncertainties in Table 1 this is an uncertain *input* to the model, not a (graphical) uncertainty from processing of the *outputs* of the model. Furthermore, the emissivity uncertainty is propagated into both the simulation and experimental results. This is not the case with the graphical uncertainties discussed here. They are relevant to only the simulation results.

to just simply ignore the graphical processing uncertainties. However, this incurs a substantial risk of Type X error in the validation conclusions.

Hence, the "validation conservative" lower temperature limit of prediction uncertainty is obtained as follows. The Sim. 3 uncertainty bar in Fig. 8 is centered on the upper diamond to its right, which has a temperature of 1285 K (= 'max' value for Sim. 3 in Table 5). The temperature at the bottom of this uncertainty bar is therefore 1285 K – 33K = 1252K. This temperature is marked by the lower horizontal dashed line in the figure.

The validation-conservative upper temperature limit is obtained by a mirrored procedure. The Sim. 4 nominal 1545K result is shifted downward to the lower diamond at its right, at 1532K (= 'min' value for Sim. 4 in Table 5). The temperature at the top of this uncertainty bar is therefore 1532K + 33K = 1565K. This temperature is marked by the upper horizontal dashed line in the figure. Accordingly, the final aggregate range of prediction uncertainty is 1252K to 1565K.

6. Analysis and Discussion of Validation Comparisons

Analogous quantities and results to those plotted in Figures 6 - 8 for TC5 can be found in [3] for all eight TCs. Speaking to the TC5 results here, there is considerable margin for error in the uncertainty estimates and processing of experimental and simulation results before the model affirmation would be overturned. The upper and lower simulation bounds in Figure 8 are seen to fairly spaciously encompass the range of the aggregate experimental uncertainty within which steady-state experimental temperatures are provisionally expected to lie. Errors would have to "conspire" (i.e., coordinate in enough antagonistic directions and magnitudes) to overcome the existing margin for error of 45K (= 1565K – 1520K) at the top. At the bottom, a margin of 34K (= 1286K – 1252K) exists against errors that might conspire. The existence of errors is a given. However, it is much less likely that they would be sufficiently large and sufficiently conspiring that the net result would exceed the indicated margins for error.

At the other seven TC locations the margins for error are even greater at 13 of the 14 upper and lower ends. In no case does the range of experimental uncertainty extend outside the prediction uncertainty. Therefore, a provisional but foreseeably robust conclusion of model affirmation is arrived at based on validation analysis and comparisons at the eight diversely spaced TCs on the calorimeter.

The word 'provisional' here signifies that large error may exist in our quantification of experiment-to-experiment variability. Nominal mean and standard-deviation results based on only two experimental repeats were used. The uncertainty on these small-sample statistics, and also in the assumption itself of normally distributed variability, could result in errors in an antagonistic direction and large enough to overturn the nominal affirmations at most TC locations. However, note that antagonistic error from this source could be partially offset or even completely overwhelmed by the substantial conservatisms introduced at several points in our experimental and simulation results processing. Moreover, the small-sample errors could go the other way, to more strongly affirm the validation conclusions at any and all TC locations. In any case, uncertainty due

to experiment-to-experiment variability is at least nominally accounted for in our validation analysis, results, and conclusions.

Finally, a mitigating factor is pointed out concerning contact resistance between the bottom of the steel cone and the thermally massive steel floor it was bolted to. It was not practical in this project to model the contact resistance at this interface, so an adiabatic boundary condition was applied at the bottom of the cone. This modeling error could be substantial, depending on the particular location considered on the calorimeter. The error causes higher temperatures to be predicted in the calorimeter than would occur if heat transfer from the cone to the steel floor was modeled. It turns out that the experimental uncertainty bars at all TC locations are skewed toward the lower ranges of the prediction bars. That is, modeling the contact conductance at the interface would shift the predicted temperatures downward, especially for the bottom four TCs 1 – 4. This would move things in the "right" direction toward better central agreement between the experimental and predicted temperatures. Since the corrections would be expected to be only 10's of degrees (not hundreds), there is sufficient margin at the top of the simulation uncertainty bars to absorb such corrections and still encompass the experimental data at the upper end.

In fact, the lowest upper margin for such a correction is 45K at TC5. All other TCs have a significantly larger upper margin. Also note that TC5 is from the upper set of TCs (level 10 in Fig. 2), where the effect of the correction would be expected to be fairly small. Indeed, the margin for correction at TC5 and the other TCs in the top row is considered to be far more than adequate. Among the lower set of TCs, the smallest available margin at the upper end for correction is 168K at TC1. This seems to be well in excess of what a correction would reasonably be expected to yield.

All in all, correcting for what is thought to be by far the largest unaccounted-for source of physics modeling error in the validation comparisons would be expected to move things in a direction that strengthens the validation conclusions, but not by a magnitude so large that it goes too far in this direction and overturns the affirmation at the upper end.

Finally, the issue of calculation verification is considered. As previously explained, project constraints required the use of "medium" spatial discretization cell sizes and a "medium" number of discrete-ordinate directions for resolving the participating-media radiation transport. The solver error-tolerance parameters that control numerical resolution in the steady-state FUEGO computations were also set at "medium" levels. (See [3] for details of discretization and solver resolution levels.) The medium levels used for spatial discretization, radiation transport, and solver computations are representative of what are routinely used for production calculation work by Sandia fire analysts. This is a pragmatic choice to achieve reasonable calculation run times (a few weeks on a thousand processors or less), but is also heuristically supported by comparisons against considerably finer resolution levels also tried on other projects (see e.g. [13]).

In any case, if any discretization-related errors happen to be in the "wrong" (antagonistic) direction that would work against validation margins, the allowable margins for error are

10's of degrees K as already established. This is significant room for error, but experience indicates that discretization effects *could* be this larger or larger. If a quantification of discretization-error uncertainty was available from resolution-refinement studies (e.g. by the methods presented in [14]), then in the validation analysis these would be handled like the simulation graphical processing uncertainties were. Finally, it is noted that substantial code verification efforts for FUEGO have been undertaken ([24]), which counts toward the veracity of the validation conclusions here.

7. Conclusions

In view of the arguments just made and the many conservatisms in the validation processing, it is reasonable to state that *based on the validation results at the eight diversely representative TC locations on the calorimeter, FUEGO modeling for wind-driven fire conditions is nominally affirmed to capture the experimental results here (steady-state temperature) according to the pragmatic validation criterion and methodology applied.*

Furthermore, there is considerable room for errors in the uncertainty estimates and processing of experimental and simulation results before this validation conclusion would be overturned. Moreover, the errors would have to be sufficiently large in magnitude and sufficiently conspiring (combining in antagonistic directions and/or sufficiently avoiding cancelation) to exceed the indicated substantial margins for error. The probability of these joint events occurring may be very low, and cannot reasonably be expected to be high—*but could occur*. In particular, the greatest concerns are: 1) the error associated with using the small-sample (two sample) standard deviation as representative of the standard deviation from a large number of repeat fire tests; and 2) error in the calculated steady-state results due to under-resolution in the discretized model and computation.

Nevertheless, on balance the analysis supports the validation conclusion provisionally arrived at in the present experimental/physical setting. This is a significant result for FUEGO and for Sandia's fire modeling program. The current work also represents a significant advancement in model validation methodology. The versatile model validation analysis framework demonstrated here handles difficulties associated with representing and aggregating aleatory and epistemic uncertainty from multiple correlated and uncorrelated source types.

References

[1] Romero, V.J., "Some Issues and Needs in Quantification of Margins and Uncertainy (QMU) for Phenomenologically Complex Coupled Systems," paper AIAA-2006-1989, 8th AIAA Non-Deterministic Methods Conference, May 1- 4, 2006, Newport, RI.

[2] Nakos, J. T., "The W76-1, JT4A-14B1 Abnormal Thermal Environment Qualification Activity: Test Data", Sandia National Laboratories report SAND2007-4286, printed July 2007.

[3] Luketa, A., Romero, V., Domino, S., Glaze, D., Sherman, M., "Validation of FUEGO Predictions of Calorimeter Heating in XTF Facility Crosswind Fire Tests JT4A-14B1," Sandia National Laboratories report in preparation (2009).

[4] Domino, S.P., Moen, C.D., Burns, S.P., and Evans, G.H., "SIERRA/Fuego: A Multi-Mechanics Fire Environment Simulation Tool," AIAA Paper 2003-0149, 41st AIAA Aerospace Sciences Meeting, Reno, NV, January 2003.

[5] AIAA, 1998, *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, AIAA G-077-1998, American Institute of Aeronautics and Astonautics, Reston, Va.

[6] Department of Energy – Defense Programs (2000), "Accelerated Strategic Computing Initiative (ASCI) Program Plan," DOE/DP-99-000010592.

[7] Department of Defense, "Key Concepts of Verification, Verification, & Accreditation," VV&A RPG document dated Sept. 15, 2006.

[8] ASME, V&V 10 – 2006, *Guide for Verification and Validation in Computational Solid Mechanics*, available at http://catalog.asme.org/Codes/PrintBook/VV_10_2006_Guide_Verification.cfm

[9] Romero, V.J., "A Paradigm of Model Validation and Validated Models for Best-Estimate-Plus-Uncertainty Predictions in Systems Engineering," Paper 200-01-1746 for Soc. Automotive Engineers 2007 World Congress April 16–20, Detroit, Michigan.

[10] Romero, V.J., "Validated Model? Not So Fast. The Need for Model 'Conditioning' as an Essential Addendum to the Model Validation Process," 9th AIAA Non-Deterministic Methods Conference, April 23 - 26, 2007, Honolulu, Hawaii.

[11] Romero, V.J., "Type X and Y Errors and Data & Model Conditioning for Systematic Uncertainty in Model Calibration, Validation, and Extrapolation," SAE paper 2008-01-1368 for Society of Automotive Engineers 2008 World Congress, April 14-17, 2008, Detroit, MI.

[12] Romero, V.J., "A Pragmatic Framework for Characterizing and Propagating Uncertainty in Hierarchical Modeling: Data and Model Conditioning, Model Verification and Validation, Extrapolative Prediction with Uncertainty," Sandia National Laboratories report in review (2009).

[13] Black, A. R., Hobbs, M. L., Dowding, K. J., Blanchat, T. K., 2007, "Uncertainty Quantification and Model Validation of Fire/Thermal Response Predictions", 18th AIAA Computational Fluid Dynamics Conference, Miami, FL, June 25-28, paper AIAA-2007-4204.

[14] Roache P.J., *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers, Albuquerque, NM, 1998.

[15] Romero, V.J., V.G. Figueroa, S.S. Yoon, J.M. Nelson, S.R. Tieszen, A.R. Black, "Uncertainty/Sensitivity Analysis of CFD Fire Modeling of Object Heating in Various Accident Scenarios," draft Sandia National Laboratories report (2003). See also "Verification, Validation, and Uncertainty/Sensitivity Studies of Fire Modeling at Sandia National Laboratories," V. Romero, A. Black, A. Brown, S. Tieszen, S. Domino, T. Blanchat, J. Nakos, A. Hanlin, K. Dowding, V. Nicolette, M. Hobbs, in proceedings CD of National Institute of Standards (NIST) Building and Fire Research Laboratory Annual Conference, Gaithersburg, MD, March 31 – April 2, 2008.

[16] Figueroa, V.G., "Emittance of Inconel 600, SS304, 17-4PH Stainless Steel, and Aluminum 7075" Sandia internal memo to Distribution, dated Feb. 20, 2005.

[17] Nakos, J.T., J.M. Suo-Anttila, W. Gill, "Shroud Boundary Condition Characterization Experiments at the Radiant Heat Facility," Sandia National Laboratories report SAND2004-5080 printed October 2004.

[18] Mehling, H., J. Kuhn, M. Valentin, J. Fricke, "Change of infrared emissivity of metal surfaces during oxidation," *High Temperatures-High Pressures*, 1998, Vol. 30, p. 333-341.

[19] Worner, B., and G. Neuer, "Emittance measurements on steel during oxidation, *High Temperatures-High Pressures*, 1998, Vol. 15, p. 455-462.

[20] Siegel, R., and Howell, J., *Thermal Radiation Heat Transfer*, Taylor & Francis, 2002.

[21] Nakos, J.T., "Uncertainty Analysis of Thermocouple Measurements Used in Normal and Abnormal Thermal Environment Experiments at Sandia's Radiant Heat Facility and Lurance Canyon Burn Site," Sandia National Laboratories report SAND2004-1023, printed April 2004.

[22] Romero, V.J., M.P. Sherman, J.F. Dempsey, J.D. Johnson, L.R. Edwards, K.C. Chen, R.V. Baron, C.F. King, "Development and Validation of a Component Failure Model," paper AIAA-2005-2141, 45th AIAA/ASME/ ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, April 18-21, 2005, Austin, TX.

[23] Figueroa, V., "Effects of Parameter Variations Associated with 1/16" Mineral Insulated Metal Sheathed Thermocouples Installation on the Surface Temperature Measurement of a Uniformly Heated Plate," University of New Mexico School of Mechanical Engineering M.S. Thesis report, April 27, 2006.

[24] Domino, S.P., G. Wagner, A. Luketa, A. Black, J. Sutherland, "Verification for Multi-Mechanics Applications," 9th AIAA Non-Deterministic Methods Conference, April 23 - 26, 2007, Honolulu, Hawaii.