Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 197 (2008) 2490–2495
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

Thermal challenge problem: Summary

R.G. Hills [a,*], K.J. Dowding [b], L. Swiler [b]

[a] New Mexico State University, Mechanical Engineering, P.O. Box 30001, Las Cruces, NM 88003, USA
[b] Sandia National Laboratories, Albuquerque, NM, USA

Received 14 May 2007; accepted 23 July 2007
Available online 31 December 2007

Abstract

This paper summarizes the approaches used to address the thermal validation challenge problem. The approaches differ in their characterization of the thermal properties and uncertainty, the definitions and use of validation metrics, the use of validation experimental data to characterize or improve the model predictions, and the assessment of regulatory compliance. All approaches estimated regulatory failure with the resulting estimated probabilities varying by an order of magnitude.
Published by Elsevier B.V.

Keywords: Validation; Modeling; Simulation

1. Introduction

The thermal challenge problem [10] is one of three problems addressed in this special issue on validation. The mathematical model and the solution provided by Dowding et al. [10] are based on one-dimensional, linear heat conduction in a solid slab, with heat flux boundary conditions. Experimental data from a series of material characterization, validation, and accreditation experiments related to the mathematical model were provided by Dowding et al. [10]. The authors [1–7] who address this problem were asked to evaluate the validity of the provided mathematical solution for use in a specified application with a defined regulatory criterion, and were asked to use the solution to predict regulatory compliance.

Multiple approaches were used by Brandyberry [1], Ferson et al. [2], Higdon et al. [3], Hills and Dowding [4], Liu et al. [5], McFarland and Mahadevan [6] and Rutherford [7] to address the thermal challenge problem. While many of the approaches possess some commonality, differences in the philosophy, focus, and details were observed. For example, the approaches of Higdon et al. [3] and Liu et al. [5] applied Bayesian analysis in the development of calibrated model correction terms and updated parameter distributions, using the ensemble and accreditation data, with little emphasis placed on validation metrics. This approach provides estimates of uncertainty in the model predictions and is based on a multivariate extension of the methodology developed by Kennedy and O'Hagan [8]. Paper [6] used a similar Bayesian approach, but also used statistical inference to assess model validity. The approaches of Brandyberry [1], Ferson et al. [2], Hills and Dowding [4] and Rutherford [7] are based on a non-Bayesian point of view, with several addressing model validity though the use of statistical inference, and several using statistical or engineering tools to characterize or improve the model's agreement with the data.

2. Summary of approaches

Tables 1 and 2 summarize major features of the approaches, and serve as an outline for the following discussion. Table 1 is partitioned into two basic approaches; Bayesian and non-Bayesian. The table addresses issues related to the three hierarchical validation steps discussed by Hills et al. [9]: (1) the use of material characterization, (2) ensemble and accreditation validation metrics, (3) and

* Corresponding author. Tel.: +1 505 646 5072; fax: +1 505 646 6111.
E-mail address: rhills@nmsu.edu (R.G. Hills).

0045-7825/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.cma.2007.07.032

Table 1
Thermal Challenge Problem: Summary

Bayesian Approach	Non-Bayesian Approach
The Use of Material Characterization Data	
Used to define prior distributions [3,5]	Used to define distributions [1,2,4,6,7]
T dependence handled through model bias terms and Bayesian estimation [3,5,6];	T dependence not included (pooled data) [2,4];
	T dependence handled though censored data [1,6,7];
Censored samples [5,6]	Model predictions based on effective k as a function of temperature or time [2,4]
Quantitative Ensemble and Accreditation Validation Metric	
None [3,5]	Based on means: univariate [1], multivariate [4,6]; Based on distributions [2,4,7]
Prediction of Uncertainty of Regulatory Compliance	
Includes uncertainty due to estimation of thermal properties and model bias parameters [3,5,6]	Includes uncertainty due to estimation of thermal properties [4];
	Includes uncertainty due to estimation of model bias parameters [7];
Includes uncertainty in the form of the distributions [3,5]	Analyzes impact of uncertainty in thermal properties on validation steps [2]

Brandyberry [1], Ferson et al. [2], Higdon et al. [3], Hills and Dowding [4], Liu et al. [5], McFarland and Mahadevan [6], Rutherford [7].

Table 2
Thermal Challenge Problem: Additional Details

Approaches to Reduce Computational Complexity (Surrogate Models)
SVD basis function with GP weights [3]
GP model [5]
Cluster based approximation [1]
Sensitivity analysis [4]
Additive Function or Effective Property Approach for Bias Correction
Application specific basis functions with GP weights [3]
GP model [5]
Linear in T [1]
Linear in input parameters [6,7]
$k = f(T)$ [2,4]
Calibrated using Ensemble/Accreditation Data
Yes [3,5–7]
No [1,2,4]

Brandyberry [1], Ferson et al. [2], Higdon et al. [3], Hills and Dowding [4], Liu et al. [5], McFarland and Mahadevan [6], Rutherford [7].

the prediction of regulatory assessment. Table 2 provides additional detail on aspects of the validation approaches.

2.1. Material characterization

Several features of the approaches used for material characterization are summarized in Table 1. All authors characterized the uncertainty in the thermal conductivity and volumetric heat capacity though statistical characterization of the provided data. Statistics of the material characterization data were used to develop prior distributions for the thermal properties for Bayesian analysis [3,5]. One paper [5] based these priors on normal distributions with statistics estimated using the characterization data. Another paper [3] defined the priors using uniform distributions with ranges based on those observed in the characterization data. The prior distributions were later updated [3,5] through Bayesian analysis; using the ensemble and accreditation data with Markov Chain Monte Carlo analysis and Metropolis updates. The methodology used by Higdon et al. [3] and Liu et al. [5] required no assumption about the form of the posterior distributions. One paper [6] used Bayesian analysis to update only the means of the thermal properties and assumed uniform priors that were not conditioned on the characterization data.

In the cases of those participants who did not use Bayesian analysis, the property distributions were estimated and held fixed for the analysis of the ensemble, accreditation, and regulatory steps. Two papers [2,4] based their material characterization on independent, normally distributed, thermal properties (the challenge problem definition [10] stated that the uncertainties in the thermal parameters k, ρC_p, were independent). Paper [2] also provided analysis based on resampling values from the observed distributions of thermal properties. Paper [1] assumed normally distributed conductivity for all data levels and for heat capacity at the low data level, but developed empirical distributions based on the raw data for heat capacity at the medium and high data levels. Papers [4,6] evaluated the goodness of the fit of the resulting normal distributions to the data using standard statistical measures, whereas several authors provided no justification for the assumption of

normally distributed thermal properties. One paper [7] used kernel density estimation to characterize uncertainty in the thermal properties.

The heat conduction model provided by Dowding et al. [10] for assessment was based on constant thermal properties. Many of the authors (see Tables 1 and 2) observed a temperature-dependence in the thermal conductivity, which can lead to bias in the model predictions. All authors developed or suggested approaches to account for model bias due to this temperature-dependence, or due to this and other forms of model bias. Authors [2,4] presented methods to account for this dependence through time or temperature-dependent effective thermal conductivity to be use with the linear heat conduction model supplied by Dowding et al. [10]. Other authors [1,3,5–7] included additive correction terms to the predicted temperatures of the supplied linear heat conduction model to account for model bias. This approach required that parameters be estimated for the correction terms, which was performed using the ensemble and accreditation data. Uncertainty in the estimated parameters is an integral part of the Bayesian analysis approach and was also addressed using statistical methods by several of the other approaches. Two authors [2,4] also presented results obtained when the means and variances of the thermal properties were evaluated using characterization data obtained at all temperatures, without additional ensemble or accreditation data calibration, or without the evaluation of effective properties, to illustrate the effects on the uncertainty in the prediction of regulatory temperature. Pooling the thermal conductivity data for all temperatures resulted in larger estimated variances for the thermal conductivity than those approaches which did not pool the data in this fashion. Censoring (sampling property data over a limited range of temperature) was used by many [1,5–7] to decrease the possible temperature dependence effect of the thermal properties, and to increase the relevance of the resulting property distributions to the ensemble, accreditation, and regulatory steps.

2.2. Validation and accreditation

Two papers [4,6] applied validation metrics with the ensemble and accreditation data that were based on standard multivariate statistics such as the χ^2 and Hotelling's T^2 statistic. One paper [1] applied univariate metrics at each measurement time for the medium and high data levels. The metrics compare measurements or means of measurements to expected values obtained from the models. Several papers [2,4,7] provided new or used existing statistical metrics to compared distributions, which characterize not only how well a model predicts mean behavior, but also how well the structure of uncertainty is predicted (i.e., the shape of the probability density or cumulative density functions). Two of the papers that used a Bayesian approach [3,5] did not present model validation metrics. These approaches focused on the development of model correction terms and updated parameter distributions, calibrated using the ensemble and accreditation data. While the Bayesian approaches did not explicitly define validation metrics, the magnitude of the resulting corrections to the predicted temperatures, and the associated estimates of uncertainty in these corrections, can be used as an indicator of model validity.

Of the authors that provided validation metrics, only Hills and Dowding [4] and Rutherford [7] found strong statistical evidence to reject the model supplied by Dowding et al. [10] based on the application of the metrics to the ensemble and accreditation data. All authors, however, developed or suggested corrected models to be used to assess regulatory performance.

2.3. Prediction of regulatory performance

All participants estimated the probability of meeting the regulatory criteria using various forms of Monte Carlo methods. For comparison, one paper [4] also used a first-order sensitivity analysis and assumed normal distributions for the regulatory predictions. With the exception of two papers [1,7], uncertainties in the characterization of the probability density functions for the thermal properties were included in the regulatory analysis, or their impact investigated. Bayesian approaches [3,5] included the impact of this uncertainty as a natural product of the methodology. One paper [6] presented a heuristic approach to account for this uncertainty through a distribution of the variances of the thermal property distributions. The uncertainty was estimated separately from their Bayesian analysis. Confidence intervals or levels of confidence on the predicted probability of failure were provided by Brandyberry [1], Ferson et al. [2], McFarland and Mahadevan [6] and Rutherford [7].

2.4. Computational cost

Several approaches were used to reduce the computational cost (number of evaluations of the solution of the mathematical model) for parameter estimation, validation, and regulatory prediction. While the analytical solution provided for the challenge problem required minimal computational resources to evaluate, many authors addressed the issue of computational cost, as real-world applications often utilize computational expensive solutions. For such applications, less computationally expensive surrogate representations of the solution are often desirable for sampling based approaches to uncertainty quantification. Gaussian process (GP) models were used as surrogate representations of the solutions of the mathematical model by two participants [3,5]. The use of GP models as surrogates for Monte Carlo simulation was introduced by Sacks et al. [11] and is well cited in the literature. The uncertainties in the GP model parameters are estimated as a product of the Bayesian analysis [8]. A cluster-based method, first used for nuclear risk assessment, was presented by paper [1]. This method uses an analytical solution of a low fidelity

approximation to the mathematical model with Monte Carlo simulation to identify clusters of temperatures, and more importantly, clusters of model parameters that span the cumulative density function for temperature for the low fidelity solution. Means or end points of these parameter clusters are then used as input for the solution of the full fidelity model, to approximate the cumulative probability distribution of the predicted temperature. A first-order sensitivity analysis approach was also presented by paper [4] to reduce model complexity by assuming that perturbations of model predictions evaluated around the mean of the parameter values, are approximately linear in the parameters. The covariance matrix for the model predictions was estimated using this approximation. The uncertainties associated with this approximation were not investigated.

2.5. Model bias

Several approaches were used to account for observations of model bias (also denoted model form bias [1], model discrepancy [3,7], or model inadequacy [6]). These approaches were based on the evaluation of effective parameters or the development of calibrated correction functions. The modified models were used by some authors for regulatory assessment, whereas others used these predictions for both regulatory assessment and validation using the ensemble and accreditation data. Authors [1,3,5–7] developed model correction functions that were calibrated using the ensemble and/or accreditation data. These correction functions were additive in the sense that they were added to the predicted temperatures of the solutions to the mathematical (or surrogate) model. The functions used to represent bias include Gaussian process models [3,5,6], and functions that are linear in temperature or in the model input parameters [1,6,7]. When used with Bayesian analysis, the probability distributions associated with the uncertainty in the bias function parameters were also provided. Papers [5,6] developed a bias term that was based on a GP model of the configuration inputs, whereas Higdon et al. [3] used a set of kernel basis functions. Two papers [2,4] reduced model bias error by using a temperature (or time) dependent thermal conductivity without ensemble or accreditation data calibration.

2.6. Data level

Many papers demonstrated the impact of low, medium, and high levels of data on the validation process or the regulatory predictions. Papers [1,6] quantified the impact on validation through the use of statistical power. Statistical power quantifies the probability of failing to reject an invalid model (Type II error) and is dependent on the quantity of data used and the assumptions made concerning the structure and statistics of the uncertainty. Several authors repeated the analysis of regulatory performance using the low, medium, and high data levels to assess sensitivity on dataset size (see Table 3 and Section 4).

2.7. Regulatory compliance

The final task addressed was to evaluate the model at the regulatory criteria specified for the challenge problem, assess whether the criteria were met, and address confidence in the assessment. The estimated probabilities of the various approaches are listed in Table 3. Note that all approaches estimated that the probability was greater than the value 0.01 specified as the regulatory criterion [10]. There were considerable differences in the probabilities, with the results ranging from 0.02 to 0.28. Probabilities less than 10% were estimated by Brandyberry [1], Ferson et al. [2], Higdon et al. [3], Hills and Dowding [4] and Liu et al.

Table 3
Regulatory compliance

Authors	P ($T > 900$ °C), data level			Comments
	Low	Medium	High	
Brandyberry [1]	0.22	0.26	0.24	
	0.23	0.26	0.16	Interpolated surrogate model
	0.12	0.08	0.06	Bias correction
	0.13	0.10	0.05	Bias correction, interpolated surrogate model
Ferson et al. [2]	0.28	0.22	0.25	
		0.25		Thermal properties sampled from data
	0.10	0.05	0.05	Effective conductivity
Higdon et al. [3]	0.07	0.03	0.03	
			0.17	Sensitivity analysis
			0.25	Monte Carlo
			0.09	Effective conductivity, sensitivity analysis
			0.14	Effective conductivity, Monte Carlo
Liu et al. [5]		0.02	0.04	
McFarland and Mahadevan [6]			0.17	
Rutherford [7]			0.15	

[5]. These lower probabilities were derived from approaches that accounted for the thermal dependence of the conductivity or approaches that included additive bias terms. Probabilities of failure larger than 20% were estimated by Ferson et al. [2] and Hills and Dowding [4] for approaches that sampled the characterization data over the entire temperature range (resulting in broader probability density functions), coupled with approaches that did not attempt to correct for possible model form error. Note that regardless of approach, the added uncertainties associated with characterization of thermal properties and the parameters that appear in bias correction terms will result in additional uncertainties in the prediction of regulatory compliance. Thus, even if the model accurately represents the physics, the characterization and inclusion of these additional uncertainties will lead to uncertainties associated with prediction that were not present in the analysis performed by Dowding et al. [10] in developing the challenge problem.

3. Discussion and conclusions

The approaches used to address the thermal challenge problem included those that followed the step-by-step validation hierarchical process discussed in the introductory paper [9], or focused on improving model performance through additive corrective terms, calibrated using the ensemble and accreditation data, with less emphasis on the intermediate steps of validation. Many of these approaches were based on long-established engineering techniques and statistical inference, whereas other approaches were based on Bayesian analysis. The results summarized in Tables 1–3 and discussed above suggest the following:

1. A wide range of approaches were used to characterize the uncertainty in the thermal properties of the materials, from the assumption of normally distributed properties with or without justification, to sampling from the characterization data, to kernel density estimation using the observed characterization data, to Bayesian analysis using the observed ensemble and accreditation data to update priors for the thermal properties.Several authors observed temperature dependence in the thermal conductivity data. Approaches used to handle this dependency ranged from the simple pooling of thermal conductivity data over all temperature, to censoring the thermal data to include observations over a limited range of temperatures, to the development of effective temperature-dependent thermal conductivity approximations. The censored or effective thermal conductivity approaches resulted in smaller variances for thermal conductivity than did the pooled approach.
2. Some authors developed additive correction terms to account for observed differences (bias) between the predictions of the model and the ensemble and accreditation data. Some of the approaches also characterized and addressed the impact of the uncertainty in the estimation of the bias terms on validation and the prediction of regulatory performance.
3. Two basic approaches were used by those that defined validation metrics to assess the models based on the ensemble and accreditation data. The first compared means of the model output to multivariate values or means of the ensemble and accreditation experimental observations. The second compared distributions of the model output to distributions estimated based on the ensemble and accreditation data.
4. All approaches predicted a failure to meet the regulatory criterion. Generally, those approaches that attempted to either account for temperature-dependent thermal conductivity through a temperature or time dependent effective thermal conductivity, or those that developed model bias correction terms using the ensemble and accreditation data, predicted smaller probabilities of failure than those that did not. Most of these approaches included the effect of uncertainty due to the estimation of parameters in the correction terms. Because the true probability of failure associated with the mathematical and statistical models used to generate the data for the challenge problem was not, and will not be made available to the participants, the potential benefits of these approaches to improve prediction were not assessed.
5. The predicted probability of failure for the regulatory design ranged from 0.02 to 0.28. This order-of-magnitude range suggests that there is considerable subjectivity in the validation process, when the validation problem is defined to include many of the uncertainties associated with possible model form error, material characterization, validation, and prediction of regulatory (or design) performance. It is reasonable to expect additional subjectivity if the added complexities associated with experimental uncertainty, physics-related model conceptualization, other sources of epistemic uncertainty, and the integration of suites of validation experiments addressing different physics with the associated different sources of uncertainty, are included.

There were major issues for which there was a consensus in approaches and one issue for which there was a difference of opinion:

1. All participants developed and characterized statistical models for the uncertainties in the model input parameters (thermal properties) using the provided data.
2. All participants used uncertainty quantification methods to propagate the uncertainties in model input parameters through the mathematical model (or surrogates), to estimate probability distributions for predictions of temperatures associated with the ensemble and accreditation experiments, and for the prediction of regulatory temperature. Several participants also developed estimates of uncertainty in the resulting probability distributions.

3. The participants who developed validation metrics for the ensemble and accreditation experiments included the effect of the uncertainties in the model input parameters in the evaluation of these metrics. The metrics were all probability-based.
4. There was disagreement as to the proper role of calibration to improve agreement between the model predictions and the ensemble or accreditation data, and on the assessment of model validity and the prediction of regulatory performance when such prediction requires extrapolation from the conditions of the validation experiments.

One of the desired goals of the workshop and this special issue was to increase awareness of methodologies associated with various approaches to model validation, including the quantification of the impact of uncertainties in the model validation process on the prediction of regulatory performance. The diversity of approaches and results, for such a simple and carefully defined validation problem, suggests that there is work do be done to reduce the subjectivity of the validation process. The presentation of these multiple approaches at a single workshop and in this special issue, represents an important step toward this goal, and toward the communication of the benefits of different approaches to various aspects of model validation under a common framework.

Acknowledgements

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. This work was supported by Sandia National Laboratories Advanced Simulation and Computing (ASC) Verification and Validation Program. We wish to thank the authors of the thermal challenge problem papers appearing in this issue for their suggestions and review of this paper.

References

[1] M.D. Brandyberry, Thermal problem solution using a surrogate model clustering technique, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2390–2407.
[2] S. Ferson, W.L. Oberkampf, L. Ginzburg, Model validation and predictive capability for the thermal challenge problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2408–2430.
[3] D. Higdon, C. Nakhleh, J. Gattiker, B. Williams, A Bayesian calibration approach to the thermal problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2431–2441.
[4] R.G. Hills, K.J. Dowding, Multivariate approach to the thermal challenge problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2442–2456.
[5] F. Liu, M.J. Bayarri, J. Berger, R. Paulo, J. Sacks, A Bayesian analysis of the thermal challenge problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2457–2466.
[6] J. McFarland, S. Mahadevan, Multivariate significance testing and model calibration under uncertainty, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2467–2479.
[7] B.M. Rutherford, Computational modeling issues and methods for the "regulatory problem" in engineering – solution to the thermal problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2480–2489.
[8] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, J. R. Stat. Soc. 63 (2001) 425–464.
[9] R.G. Hills, M. Pilch, K.J. Dowding, J. Red-Horse, T.L. Paez, I. Babuška, R. Tempone, Validation challenge workshop, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2375–2380.
[10] K.J. Dowding, M. Pilch, R.G. Hills, Formulation of the thermal problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2385–2389.
[11] J. Sacks, W.J. Welch, T.J. Mitchell, H.P. Wynn, Design and analysis of computer experiments, Stat. Sci. 4 (1989) 409–423.