Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 217 (2006) 5–36
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Measures of agreement between computation and experiment: Validation metrics

William L. Oberkampf [a,*], Matthew F. Barone [b,1]

[a] *Validation and Uncertainty Quantification Department, Mailstop 0828, P.O. Box 5800, Sandia National Laboratories, Albuquerque, NM 87185-0828, United States*
[b] *Aerosciences and Compressible Fluid Mechanics Department, Mailstop 0825, P.O. Box 5800, Sandia National Laboratories, Albuquerque, NM 87185, United States*

Received 26 August 2005; received in revised form 17 March 2006; accepted 30 March 2006
Available online 12 June 2006

Abstract

With the increasing role of computational modeling in engineering design, performance estimation, and safety assessment, improved methods are needed for comparing computational results and experimental measurements. Traditional methods of graphically comparing computational and experimental results, though valuable, are essentially qualitative. Computable measures are needed that can quantitatively compare computational and experimental results over a range of input, or control, variables to sharpen assessment of computational accuracy. This type of measure has been recently referred to as a validation metric. We discuss various features that we believe should be incorporated in a validation metric, as well as features that we believe should be excluded. We develop a new validation metric that is based on the statistical concept of confidence intervals. Using this fundamental concept, we construct two specific metrics: one that requires interpolation of experimental data and one that requires regression (curve fitting) of experimental data. We apply the metrics to three example problems: thermal decomposition of a polyurethane foam, a turbulent buoyant plume of helium, and compressibility effects on the growth rate of a turbulent free-shear layer. We discuss how the present metrics are easily interpretable for assessing computational model accuracy, as well as the impact of experimental measurement uncertainty on the accuracy assessment.
Published by Elsevier Inc.

Keywords: Validation of computational models; Assessment of model accuracy; Model credibility; Uncertainty quantification; Experimental uncertainty; Fluid dynamics; Solid dynamics

1. Introduction

It is common practice in all fields of engineering and science for comparisons between computational results and experimental data to be made graphically. The graphical comparisons are usually made by plotting some

* Corresponding author. Tel.: +1 505 844 3799; fax: +1 505 844 4523.
E-mail addresses: wloberk@sandia.gov (W.L. Oberkampf), mbarone@sandia.gov (M.F. Barone).
[1] Tel.: +1 505 284 8686; fax: +1 505 844 4523.

0021-9991/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.jcp.2006.03.037

Nomenclature

C	confidence level chosen, $C = 100(1-\alpha)\%$
$\left\lvert\frac{\mathrm{CI}}{\bar{y}_e}\right\rvert_{\mathrm{avg}}$	average confidence indicator associated with the average of the absolute value of the relative estimated error over the range of the experimental data, see either Eq. (19) or (26)
$\left\lvert\frac{\mathrm{CI}}{\bar{y}_e}\right\rvert_{\mathrm{max}}$	confidence interval associated with the maximum absolute value of the relative estimated error over the range of the experimental data, see either Eq. (21) or (27)
E	true error of the computational model as compared to the true mean of the experimental measurements, $y_m - \mu$
$\widetilde{E}$	estimated error of the computational model as compared to the estimated mean of the experimental measurements, $y_m - \bar{y}_e$
$\left\lvert\frac{\widetilde{E}}{\bar{y}_e}\right\rvert_{\mathrm{avg}}$	average of the absolute value of the relative estimated error over the range of the experimental data, see Eq. (18)
$\left\lvert\frac{\widetilde{E}}{\bar{y}_e}\right\rvert_{\mathrm{max}}$	maximum of the absolute value of the relative estimated error over the range of the experimental data, see Eq. (20)
$F(v_1, v_2, 1-\alpha)$	F probability distribution, where v_1 is the first parameter specifying the number of degrees of freedom, v_2 is the second parameter specifying the number of degrees of freedom, and $1-\alpha$ is the quantile for the confidence interval chosen
n	number of sample (experimental) measurements
s	sample (estimated) standard deviation based on n experimental measurements
SRQ	system response quantity
t_v	t distribution with v degrees of freedom, $v = n-1$
$t_{\alpha/2v}$	$1-\alpha/2$ quantile of the t distribution with n degrees of freedom, $v = n-1$
$\bar{y}_e$	sample (estimated) mean based on n experimental measurements
y_{m}	mean of the SRQ from the computational model
α	arbitrarily chosen total area from both tails of the specified distribution
μ	population (true) mean from experimental measurements
$\vec{\theta}$	vector of coefficients of the chosen regression function, Eq. (22)
$\vec{\hat{\theta}}$	vector of regression coefficients that minimize the error sum of squares, Eq. (24)

computational system response quantity (SRQ) with the experimentally measured response over a range of some input parameter. If the computational results generally agree with the experimental data, the computational model is commonly declared, "validated". Comparing computational results and experimental data on a graph, however, is only incrementally better than making a qualitative comparison. With a graphical comparison, one rarely sees quantification of numerical solution error or quantification of computational uncertainties, e.g., due to variability in modeling parameters, missing initial conditions, or poorly known boundary conditions. In addition, an estimate of experimental uncertainty is not typically quoted, nor its statistical character quantified. A graphical comparison also gives little quantitative indication of how the agreement between computational results and experimental data varies over the range of the independent variable, e.g., a spatial coordinate, time, or Mach number. Further, a simple graphical comparison is ill suited for the purpose of quantitative validation because statistical methods are needed to quantify experimental uncertainty. It should be noted that some journals, such as those published by the American Institute of Aeronautics and Astronautics (AIAA) and the American Society of Mechanical Engineers (ASME), now require improved statements of numerical accuracy and experimental uncertainty.

The increasing impact of computational modeling on engineering system design has recently resulted in an expanding research effort directed toward developing quantitative methods for comparing computational and experimental results. In engineering and physics, the form of the computational models is predominantly given by partial differential equations (PDEs) with the associated initial conditions and boundary conditions. Although statisticians have developed methods for comparing models (or "treatments") of many sorts, their emphasis has been distinctly different from the modeling accuracy assessment perspective in engineering. Much