

基于概率分布距离的多响应模型确认度量

赵亮, 杨战平

(中国工程物理研究院 电子工程研究所, 四川 绵阳 621900)

摘要: 针对模型确认中的确认度量问题, 构造实验观测数据经验概率分布的置信包络. 通过计算其与模型响应概率分布之间距离的上/下确界, 给出基于概率分布距离确认度量的置信区间. 通过构造与实验观测数据有关的协方差矩阵, 给出基于概率分布距离的多响应模型确认度量及其置信区间的求解方式. 该度量利用了模型输出与实验观测的完整概率分布信息, 并且考虑了各模型响应间的相关性. 算例仿真结果表明其确认错误率低于现有的其他两种确认度量.

关键词: 模型确认; 概率分布距离; 置信区间; 多响应确认度量

中图分类号: O327; TH113

文献标志码: A

Multiple response model validation metric based on distance of probability distribution

ZHAO Liang, YANG Zhan-ping

(Institute of Electronic Engineering, China Academy of Engineering Physics, Mianyang 621900, China. Correspondent: ZHAO Liang, E-mail: swjtu_zhaoliang@126.com)

Abstract: For the metric for model validation, the confidence interval of the validation metric is presented by calculating the infimum and supremum of the distance between the empirical probability distribution of model response and the confidence interval of the experimental distribution. Besides, a multiple response validation metric based on the distance of probability distribution is proposed with its confidence interval by constructing experimental covariance matrix. The metric makes use of the whole data distribution and considers the correlation among multiple model responses. Simulation results show that the validation error rate of the proposed metric is lower than other two metrics.

Keywords: model validation; distance of probability distribution; confidence interval; multiple response validation metric

0 引言

建模仿真如今科学、工业领域中广泛采用的一种科研、设计方法, 由建模仿真得到的结果通常是相关领域决策的重要依据. 在一些涉及重大安全性、可靠性的问题中, 建模仿真所得结果的准确性是人们尤为关心的问题, 模型确认就是针对该问题而提出的一个重要研究方向^[1]. 如 George Box^[2]所说: “现实中的所有模型都是错的, 但有些模型是有用的”, 模型确认并非以“对”或“错”的形式评判一个模型, 而是评估模型在其预期应用环境下的准确性是否足够, 在这一过程中需要克服的困难包括: 如何量化以不确定性形式表示的模型响应与实验观测之间的差异; 如何将基于这种量化结果的判断外推至缺乏实验观测

的模型预期应用环境, 并给出相应的置信度等. 虽然, 国外已经开展了大量有关模型确认的研究^[1, 3-5], 但距离在实际问题中进行严格意义的模型确认仍有差距, 相比之下, 国内在这方面的研究仍处于起步阶段.

确认度量是模型确认中的重要环节, 是对含有不确定性的模型响应与实验观测之间差异的量化. 目前, 研究人员已经在模型确认中提出或应用了多种形式的确认度量^[6-14], 它们各有特点; 但由于模型确认本身尚缺乏统一的认识, 实际模型确认过程中在确认度量的选择上表现出很大的主观性和随意性. 本文结合模型确认的基本需求总结了确认度量应具备的性质, 以此为依据分析了目前具有代表性的几类确认度量, 提出一种新的多响应确认度量, 并通过算例仿真

收稿日期: 2014-03-16; 修回日期: 2014-07-06.

基金项目: 中国工程物理研究院科学技术发展基金项目(2012B0403058); 十二五行业预研项目(426010401).

作者简介: 赵亮(1983—), 男, 博士生, 从事模型验证与确认的研究; 杨战平(1966—), 男, 研究员, 博士生导师, 从事复杂系统建模与仿真等研究.

验证了所提出度量的有效性。

1 对现有确认度量的分析

1.1 确认度量的要求

早期评价一个模型所采用的“图形对比”的方式仅仅能够提供对其准确性的定性评估,而模型确认中的确认度量要求给出模型响应与实验观测之间差异的客观量化结果. 本文以 $d(x^m, x^e)$ 表示这种量化结果. 其中: x^m 为模型响应数据, x^e 为实验观测数据. 此处结合科学、工程领域在模型确认中的需求^[6, 13-14], 将确认度量所应具有的性质总结如下:

- 1) 给出客观的量化结果, 排除分析人员主观因素的影响;
- 2) 能反映模型响应与实验观测数据的整体之间的差异;
- 3) $d(x^m, x^e)$ 具备数学上定义度量时所要求的性质: $d(x^m, x^e) \geq 0$, 当且仅当 $x^m = x^e$ 时, $d(x^m, x^e) = 0$;
- 4) 确认度量能够提供与实验观测数量相联系的置信水平;
- 5) 判断模型准确性的标准独立于确认度量;
- 6) 当模型的不确定性增加时, 确认度量不倾向于接受模型;
- 7) 确认度量不仅能够对模型响应和实验观测进行“单点”比较, 也能将“多点”的数据合并起来进行比较.

性质7)中的“单点”比较是指单变量输出模型在一个确认点的响应与该确认点的实验观测之间的比较, 其反映的是模型输出的局部准确性. 所谓“多点”比较则包括单变量输出模型在不同确认点的响应与对应的实验观测之间的比较, 或者多变量输出模型在一个确认点的多个变量的响应与该确认点的实验观测之间的比较. 将这种能进行“多点”比较的度量称为多响应确认度量, 它可以评估模型的整体准确性, 这一点是工程设计中尤为看重的^[10-11, 14].

1.2 目前的几类确认度量

目前比较有代表性的确认度量主要有两类, 一类为假设检验类型, 具体采用的方法包括经典假设检验法^[7-11]和贝叶斯因子法^[8, 11]. 其中经典假设检验法通常基于实验观测与模型响应同分布的假设构造检验统计量, 根据实验观测数据计算统计量, 并由此判断是否拒绝原假设. 其关注点在于已有的证据是否能拒绝模型正确的假设, 但没有足够证据拒绝一个模型并不等同于该模型是准确的. 相比之下, 基于贝叶斯假设检验的贝叶斯因子法主要估计模型正确的信任

度(后验概率), 将有关模型是否准确的两个假设的后验概率之比作为确认度量, 其关注点在于是否接受而非拒绝模型. 但是, 如前文所述, 对确认度量的第1要求是给出模型响应与实验观测间差异的客观量化结果, 而非带有主观性的“模型正确的信任度”. 总而言之, 假设检验类的确认度量在回答模型是否准确时难以提供对模型准确程度的客观量化, 这与对确认度量的要求之间存在差距. 同时, 由于假设检验类的确认度量以“拒绝模型”或“接受模型”的形式作为度量结果, 当存在“多点”数据时, 如何合理整合各“点”所得的度量结果是存在困难的.

另一类确认度量可归为距离类型, 这类方法根据模型响应和实验观测数据的均值^[4, 12-13]或概率分布之间的距离^[14]来量化两者的差异, 更符合确认度量所要求的客观量化差异的性质. 此类度量的关键在于选择合适的数据特征作为距离的计算对象. 如果只是比较模型响应和实验观测的数据均值, 或是类似的仅仅反映了数据某项趋势的特征参数, 则并不能保证所得的度量结果反映了模型响应和实验数据的整体之间的差异. 在不易确定数据的哪些特征是重要的时, 以模型响应和实验观测两者分布的距离作为确认度量是较好的选择, 因为概率分布中包含了几乎所有的数据信息. 另外, 以概率分布的差异作为确认度量也符合“ $d(x^m, x^e) \geq 0$, 当且仅当 $x^m = x^e$ 时, $d(x^m, x^e) = 0$ ”这一数学上对度量的要求, 并且 $d(x^m, x^e)$ 不会随模型的不确定性增加而减小, 即不确定性增加不会使确认度量倾向于接受模型. 这类确认度量的不足在于没有提供与实验观测数量相联系的置信水平, 且在进行多响应确认时采用概率积分变换的方式将所有实验数据无差异地转换到概率域^[14], 这样做未考虑到不同模型响应之间的相关性.

通过上述总结可以发现, 基于概率分布距离的确认度量相比于假设检验类型以及其他距离类型的确认度量更符合1.1节所述要求, 但相对于完整的确认度量应有的性质仍存在不足, 即缺乏置信区间, 且多响应确认不够合理. 本文通过构造实验观测数据经验分布的置信包络, 计算其与模型响应分布差异之间的上/下确界, 给出基于概率分布距离的确认度量与实验观测数量相联系的置信区间; 再通过构造与实验观测数据有关的协方差矩阵, 给出一种基于“单点”概率分布距离的多响应确认度量. 该确认度量考虑了各“点”数据间的相关性, 能在“多点”情况下给出模型响应与实验观测之间整体差异的量化结果. 算例仿真结果验证了本文方法的合理性和有效性.

2 改进的基于概率分布距离的确认度量

2.1 确认度量的置信区间

基于概率分布距离的确认度量如图 1 所示. 文献 [14] 通过计算模型响应与实验观测的经验概率分布之间的差异面积 (图 1 中阴影部分), 可给出对模型准确性的量化结果^[14], 此处的“面积”可视为距离的一种特例.

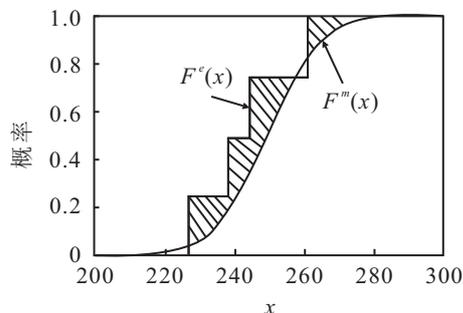


图 1 基于概率分布距离的确认度量示意图

图 1 中 $F^m(x)$ 为在某一确认点处模型响应值所构成的经验概率分布, 在不需要考虑计算代价的情况下模型输出可以足够多, 从而得到类似图 1 中的平滑 $F^m(x)$ 曲线. $F^e(x)$ 为同一确认点由实验观测得到的经验累积概率分布, 由于实验观测数据通常是有限的, 使 $F^e(x)$ 呈现明显的阶梯状. 该度量的数学表达式为

$$d(x^m, x^e) = \int_{-\infty}^{+\infty} |F^e(x) - F^m(x)| dx, \quad (1)$$

其本质是 Wasserstein 距离的一种特例.

为了给出式 (1) 所示确认度量的置信区间, 需要考虑实验观测数据的样本不确定性, 它源于有限的实验观测样本不足以反映总体变化的情况. 本文采用 Kolmogorov-Smirnov 置信包络, 对实验观测数据的这种不确定性进行量化, 其满足以下定理.

定理 1 有一组不确定性变量的观测样本数据, 其经验概率分布函数为 $F^e(x)$, 则在置信水平 $(1 - \alpha) \times 100\%$ 下, 观测样本的概率包络为

$$\overline{F^e}(x) = \min(1, \max(0, F^e(x) + \varepsilon)), \quad (2)$$

$$\underline{F^e}(x) = \min(1, \max(0, F^e(x) - \varepsilon)). \quad (3)$$

其中: $\overline{F^e}(x)$ 为概率分布 $F^e(x)$ 的上界; $\underline{F^e}(x)$ 为概率分布 $F^e(x)$ 的下界; 由 $\overline{F^e}(x)$ 和 $\underline{F^e}(x)$ 围成的概率包络决定了一族概率分布函数 $F(x)$, 且满足 $\overline{F^e}(x) \leq F(x) \leq \underline{F^e}(x)$; ε 为 Kolmogorov-Smirnov 置信包络的临界统计^[14], 且只与置信水平 $(1 - \alpha) \times 100\%$ 和样本数 p 有关.

得到实验观测数据在 $(1 - \alpha) \times 100\%$ 置信水平下的 Kolmogorov-Smirnov 置信包络后, 确认度量 $d(x^m,$

$x^e)$ 在同样置信水平下的置信区间可表示为

$$\left[\inf_{f^e(x) \in \text{KS}} \int_{-\infty}^{+\infty} |f^m(x) - f^e(x)| dx, \sup_{f^e(x) \in \text{KS}} \int_{-\infty}^{+\infty} |f^m(x) - f^e(x)| dx \right], \quad (4)$$

其中 KS 表示由式 (2) 和 (3) 所得的置信包络. 式 (4) 所示的置信区间只与置信水平 $(1 - \alpha) \times 100\%$ 和实验观测的样本数 p 有关. 式 (4) 的求解可转化为以下优化问题:

$$\begin{aligned} \text{find } \mathbf{x}, \max & \int_{-\infty}^{+\infty} |F^e(x) - F^m(x)| dx; \\ \text{s.t. } & \text{size}(\mathbf{x}) = p, F^e(x) \in \text{KS}. \end{aligned} \quad (5)$$

$$\begin{aligned} \text{find } \mathbf{x}, \min & \int_{-\infty}^{+\infty} |F^e(x) - F^m(x)| dx; \\ \text{s.t. } & \text{size}(\mathbf{x}) = p, F^e(x) \in \text{KS}. \end{aligned} \quad (6)$$

通过遗传算法之类的方法可获取其最优值. 式 (5) 和 (6) 中: \mathbf{x} 为 Kolmogorov-Smirnov 置信包络内可能的实验观测数据向量, $\text{size}(\cdot)$ 为计算向量中元素个数的算子.

2.2 “多点”情况下的确认度量

式 (1) 中的确认度量利用两个累积概率分布表示所有模型响应和实验观测所包含的信息, 它只适用于“单点”确认的情况. 当数据来自 $n (n > 1)$ 个确认点, 或者数据对应模型的 n 个输出变量时, 如何利用所有数据给出有关模型整体准确性的量化结果是多响应确认度量需要解决的问题. 为此提出确认度量的形式如下:

$$M^2 = \mathbf{D}^T \mathbf{W}^{-1} \mathbf{D}. \quad (7)$$

其中: \mathbf{D} 为由式 (1) 所得的度量值 $d_i (i = 1, 2, \dots, n)$ 组成的向量, d_i 分别对应各确认点/各模型输出变量的“单点”确认值; \mathbf{W} 为一正定阵, 取

$$\mathbf{W} = \text{cov}(\mathbf{X}^e), \quad (8)$$

\mathbf{X}^e 为 $p \times n$ 维的实验观测矩阵, p 为各确认点/各模型输出变量对应的实验观测数目. 式 (5) 中的确认度量 M^2 通过实验观测数据的协方差矩阵 \mathbf{W} 反映了各“单点”度量值 d_i 之间的联系, 并且是尺度无关的, 这一点对于多变量输出模型的确认非常重要, 因为这种模型不同的输出变量通常具有不同的属性, 分别对应不同的尺度, 所以衡量此类模型整体准确性的确认度量应是尺度无关的.

注意到可能存在各确认点/各模型输出变量对应的实验观测数目不同, 或者 $p \leq n$ 的情况, 从而导致 \mathbf{W}^{-1} 为奇异矩阵. 为了避免这种情况并充分利用所有实验观测数据, 此处给出另一种矩阵 \mathbf{W} 的构造方法. 第 $i (i = 1, 2, \dots, n)$ 个确认点/模型输出变量

的“单点”度量 $d(x^m, x^e)$ 可通过下式离散化求解:

$$d_i(x^m, x^e) = \sum_{j=1}^N |F_i^e(x_{ij}) - F_i^m(x_{ij})| \cdot \Delta x_{ij} = \sum_{j=1}^N A_i(x_{ij}) \cdot \Delta x_{ij} - \sum_{j=1}^N B_i(x_{ij}) \cdot \Delta x_{ij}. \quad (9)$$

其中

$$A_i(x_{ij}) = \begin{cases} F_i^e(x_{ij}), & F_i^e(x_{ij}) \geq F_i^m(x_{ij}); \\ -F_i^e(x_{ij}), & F_i^e(x_{ij}) < F_i^m(x_{ij}); \end{cases}$$

$$B_i(x_{ij}) = \begin{cases} F_i^m(x_{ij}), & F_i^e(x_{ij}) \geq F_i^m(x_{ij}); \\ -F_i^m(x_{ij}), & F_i^e(x_{ij}) < F_i^m(x_{ij}). \end{cases}$$

式(9)将概率分布 $F_i^m(x)$ 和 $F_i^e(x)$ 的定义域划分为足够小的 N 个区间, 各区间宽度为 Δx_{ij} , x_{ij} 取各区间的中点. 注意到式(9)中的 $\sum_{j=1}^N B_i(x_{ij}) \cdot \Delta x_{ij}$ 包含了在计算 $d_i(x^m, x^e)$ 时所用到的实验观测数据的所有信息, 通过改变 N , 可以将其划分为任意数量的 $B_i(x_{ij}) \cdot \Delta x_{ij}$ 的组合. 由此可得到一种有关实验观测数据的 $N \times n$ 维矩阵为

$$Y^e = \begin{bmatrix} B_1(x_{11})\Delta x_{11} & B_2(x_{21})\Delta x_{21} & \cdots \\ B_1(x_{12})\Delta x_{12} & B_2(x_{22})\Delta x_{22} & \cdots \\ \vdots & \vdots & \ddots \\ B_1(x_{1N})\Delta x_{1N} & B_2(x_{2N})\Delta x_{2N} & \cdots \\ B_n(x_{n1})\Delta x_{n1} \\ B_n(x_{n2})\Delta x_{n2} \\ \vdots \\ B_n(x_{nN})\Delta x_{nN} \end{bmatrix}. \quad (10)$$

该矩阵的行数 N 不受实验观测数目的限制, 任何情况下均可使 $N > p$, 以此构造的协方差矩阵为

$$W = \text{cov}(Y^e), \quad (11)$$

在确保 W^{-1} 非奇异的同时也能反映各 d_i 之间联系.

式(7)中多响应确认度量的置信区间同样基于各确认点处实验观测数据的 Kolmogorov-Smirnov 置信包络进行表示, 即

$$\left[\inf_{F_i^e(x) \in \text{KS}_i} D^T W^{-1} D, \sup_{F_i^e(x) \in \text{KS}_i} D^T W^{-1} D \right]. \quad (12)$$

其中: KS_i 表示第 i 个确认点的 p 个实验观测数据在 $(1 - \alpha) \times 100\%$ 置信水平下的 Kolmogorov-Smirnov 置信包络, 该置信区间与置信水平和各确认点的实验观测数量有关. 类似于第2.1节, 通过以下优化问题对式(12)进行求解:

$$\begin{aligned} & \text{find } X^e, \max D^T W^{-1} D; \\ & \text{s.t. } F_i^e(x) \in \text{KS}_i. \end{aligned} \quad (13)$$

$$\begin{aligned} & \text{find } X^e, \min D^T W^{-1} D; \\ & \text{s.t. } F_i^e(x) \in \text{KS}_i. \end{aligned} \quad (14)$$

其中 X^e 为各确认点的 Kolmogorov-Smirnov 置信包络内可能的实验观测数据向量所构成的矩阵:

由式(7)所得的确认度量值 M^2 是模型响应和实验观测的概率分布之间整体差异的一种客观量化结果. 显然 $M^2 \geq 0$, 当且仅当所有模型响应和实验观测完全一致时才有 $M^2 = 0$. 在实际的模型确认问题中, 还需要回答 M^2 小于何阈值时才可以说明模型的准确性是足够的. 该阈值的设定取决于模型使用者对模型偏差的接受程度、数据的分布类型以及所要求的显著性水平等因素, 因为模型“足够准确”的要求本身就是相对于特定条件而言的. 正如第1.1节所述, 对确认度量的要求之一就是判断模型准确性的阈值与确认度量值应是相互独立的, 即模型确认中判断模型是否准确的阈值并非确认度量的研究内容, 故本文不讨论所提出的确认度量的阈值问题.

3 算例分析

例1 采用文献[6]中的算例验证本文方法, 所有实验观测数据都由下式产生:

$$y^e(x, \theta) = \sin(x - 0.5\pi\theta) + \cos(\theta + 0.25\pi) + 0.2x + \varepsilon_e. \quad (15)$$

其中: $x(0 \leq x \leq 8)$ 为确定性输入; θ 为实验参数, 取固定值 1.5; 观测误差 ε_e 服从正态分布 $N(0, \sigma_e^2)$. 为了检验确认度量能否分辨出正确模型, 分别构建两个模型: 模型1与式(15)完全一致, 其参数 $\theta = 1.5$, 即 $y_1^m(x) = y^e(x, \theta = 1.5)$; 模型2则将参数 θ 取为 1.2, 即 $y_2^m(x) = y^e(x, \theta = 1.2)$, 以模拟参数存在一定错误的模型.

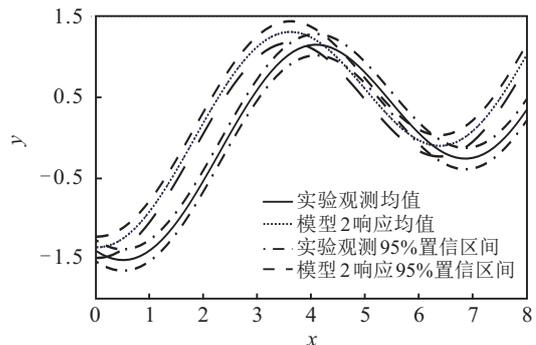


图2 实验观测与模型响应对比图

图2描述了由式(15), 模型1和模型2所得的实验观测及模型响应的均值和95%置信区间, 其中观测误差的标准差 $\sigma_e = 0.08$. 由于模型1与式(15)完全

一致,图2中模型1的响应与由式(15)产生的实验观测重合.由图2可以看出,参数 θ 错误的模型2的响应与实验观测数据存在偏差,但在 x 的部分取值区域,如 $x \in [4, 7]$,两者有一定重合.

在 $x = 2, 3, \dots, 7$ 这6个确认点各产生5个实验观测数据以及20个模型响应数据.表1为模型1和模型2在各确认点通过式(1)和(4)所得到的确认度量及对应的95%置信区间.

由表1可见,模型1的确认度量值在所有确认点均小于模型2,说明基于概率分布距离的确认度量能够分辨这两个模型的好坏.需要说明的是,虽然模型1与产生实验观测数据的式(15)完全一致,但由于观测误差 ε_e 的存在,模型1的确认度量没有出现0值.同时也注意到,这两个模型的确认度量的95%置信区间在部分确认点处存在重叠(与图2中模型响应与实验观测数据重叠的趋势一致),说明有限实验观测数据所存在的样本不确定性有可能使模型1的确认度量大于模型2的确认度量,从而导致在模型确认上出现错误.

下面通过统计蒙特卡罗实验中的确认错误率来比较几种度量的“多点”确认能力.假设感兴趣的模型应用涵盖多个确认点处的模型响应,由式(15)在

多个确认点各产生5个实验观测数据,模型1和模型2相应地各产生20个模型响应数据,计算模型1和模型2的确认度量值,将此过程重复10000次.表2为在不同确认点数据组合的条件下,本文确认度量与文献[6, 14]中基于均值距离的度量和u-pooling度量所得到的确认错误率的比较结果,其中本文度量1和度量2分别采用式(8)和(11)的协方差矩阵.

表2中选择的几种“多点”数据组合中都包含了确认点 $x = 4, 6$.图2显示模型1和度量2的响应在这两个确认点处存在较严重的重叠.由表2的结果可以看出,当“多点”数据包含的确认点数量小于各确认点的实验观测数量时,本文度量1和度量2的确认错误率小于基于均值距离的度量和u-pooling度量.这是因为本文度量1和度量2利用了各确认点数据的完整概率分布,并考虑了其相关性.当“多点”数据包含的确认点数量大于等于确认点的实验观测数量时,本文度量1失效.这是因为确认点数量大于单个确认点对应的实验观测数量,导致式(8)求逆时出现了奇异矩阵.本文度量2,均值距离度量和u-pooling度量不受上述实验观测数量的限制,它们的确认错误率随“多点”数据的增加而逼近于0.本文度量2在各种“多点”实验观测数据组合下的确认错误率均不高于其他度量,体现出本文确认度量的有效性.

表1 不同确认点的“单点”确认度量及置信区间

确认点	模型1		模型2	
	确认度量	确认度量的95%置信区间	确认度量	确认度量的95%置信区间
$x = 2$	0.046 4	[0.021 0, 0.092 6]	0.716 8	[0.637 7, 0.792 9]
$x = 3$	0.046 2	[0.020 7, 0.089 8]	0.550 4	[0.474 8, 0.631 7]
$x = 4$	0.045 7	[0.020 6, 0.092 1]	0.111 4	[0.042 7, 0.182 4]
$x = 5$	0.046 3	[0.021 1, 0.092 5]	0.199 1	[0.127 1, 0.278 6]
$x = 6$	0.045 8	[0.020 1, 0.089 0]	0.098 0	[0.034 3, 0.174 5]
$x = 7$	0.047 3	[0.021 5, 0.094 9]	0.330 3	[0.249 2, 0.409 4]

表2 不同度量进行“多点”确认的错误率

模型预期应用所包含的确认点	本文度量1	本文度量2	均值距离度量	u-pooling度量
$x = 4, 6$	0.047 5	0.050 2	0.054 0	0.325 0
$x = 4, 5, 6$	0.015 3	0.007 3	0.025 0	0.057 8
$x = 3, 4, 5, 6$	0.011 0	0.002 9	0.012 8	0.030 5
$x = 2, 3, 4, 5, 6$	0.507 8	0.002 0	0.005 1	0.002 8
$x = 2, 3, 4, 5, 6, 7$	0.487 0	0	0.000 7	0

例2 采用美国桑迪亚国家实验室提出的热传导模型确认挑战问题^[16]验证本文确认度量.该问题以特定环境下材料的温度响应为具体对象,探讨模型确认过程中需要解决的关键问题.问题中描述材料温度响应的模型为

$$T(k, \rho C, T_0, L, q; x, t) = T_0 + \frac{qL}{k} \left[\frac{kt/\rho C}{L^2} + \frac{1}{3} - \frac{x}{L} + \frac{x^2}{2L^2} - \sum_{n=1}^6 \frac{2}{\pi^2 n^2} \exp\left(-\frac{n^2 \pi^2 kt}{L^2 \rho C} \cos\left(n\pi \frac{x}{L}\right)\right) \right]. \quad (16)$$

其中: k 为材料的导热系数, ρC 为体热容, T_0 为初始温度 (25°C), L 为材料的厚度, q 为热流密度, t 为时间. 这里前两者为材料特性参数, 具有不确定性. 问题同时给出了多个确认点的实验观测数据, 分别称为整体确认 (EN) 数据和认证确认 (AC) 数据, 如图3所示. 研究该挑战问题需要回答的内容之一是使用所给的材料特性数据描述参数 k 和 ρC 的不确定性, 并根据所提供的实验观测数据评估模型的准确性.

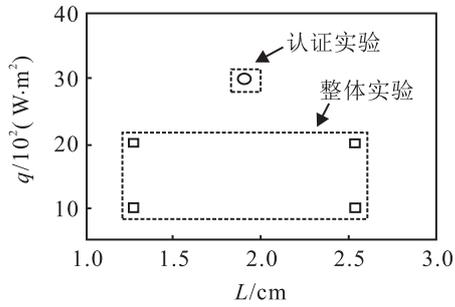


图3 实验观测对应的不同确认点

在部分有关该挑战问题的研究^[17]中, 直接以正态分布描述模型(16)的不确定性参数 k 和 ρC , 但由图4可以看出, k 与温度 T 之间具有相关性. 此处采用回归分析的方式考察 k 与 T 之间的联系, 可得

$$k(T) = 0.05035 + 2.479 \times 10^{-5}T + N(0, 0.0047).$$

(17)

对式(17)所示的回归方程进行显著性检验, 结果显示了方程中自变量 T 的回归系数“显著的”非零, 由此证实了 k 与 T 之间的相关性. 虽然严格的模型确认是区别于模型修正的^[1], 但忽略 k 与 T 的相关性势必导致模型预测中存在较大的不确定性, 故在该开放性的挑战问题中, 本文针对不确定性参数 k 的温度

依赖性进行了多种方式的模型修正^[14, 18]. 本文采用式(17)描述 k 的温度依赖性, 通过其与模型(16)的迭代得到有关材料温度的模型响应分布. 将这种考虑了参数温度依赖性的模型称为修正模型, 将式(16)称为原始模型.

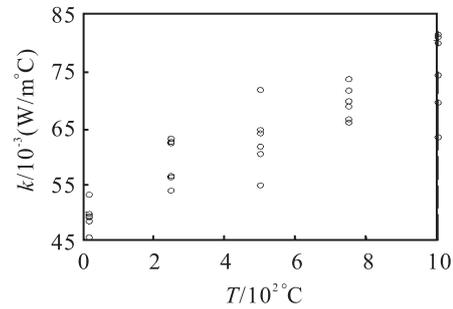


图4 模型参数 k 的材料特性数据

下面采用本文提出的确认度量来比较原始/修正模型与实验观测数据的差异. 该挑战问题给出了 EN 数据对应模型参数 L 和 q 的4种配置, 在每种配置下, 从 $0 \sim 1000$ s 每间隔 100 s 各测量了4次材料温度, 将 0 s 时的材料温度视为初始温度 T_0 , 则 EN 数据包含了 4×10 个确认点, 每个确认点有4次实验观测. 通过原始/修正模型在这 4×10 个确认点处各产生50次的模型响应, 再利用式(7)对这两个模型进行“多点”确认, 确认度量的95%置信区间可由式(12)得到, 具体结果如表3所示.

前文已经介绍过本文所提出的确认度量是与尺度无关的, 由表3可见, 修正模型的确认度量小于原始模型, 这说明修正模型与 EN 实验数据的差异更小. 类似地, 基于 AC 数据对原始/修正模型进行“多点”确认, 结果如表4所示. 由表4可见, 修正模型的响应更接近实验观测数据, 该结论与文献[14, 18]一致.

表3 原始/修正模型基于 EN 数据的“多点”确认度量

参数	原始模型	修正模型
确认度量	1.52e+04	5.46e+03
确认度量的95%置信区间	[1.35e+04, 1.67e+04]	[2.76e+03, 7.21e+03]

表4 原始/修正模型基于 AC 数据的“多点”确认度量

参数	原始模型	修正模型
确认度量	2.30e+04	7.83e+03
确认度量的95%置信区间	[1.96e+04, 2.61e+04]	[5.75e+03, 8.92e+03]

注意本文讨论该算例的目的并非想说明式(17)是针对该问题所给模型的理想修正方式, 而是通过这样一个精心设计的模型确认问题, 验证本文所提出的确认度量的有效性. 由表3和表4中的数据可知, 基于该挑战问题所给的“多点”确认数据, 本文确认度量

更倾向于接受修正模型. 该结论与其他采用同样模型修正方式的研究者所得的结论是一致的, 这表明本文提出的确认度量实现了将“多点”模型响应和实验观测数据整合在一起以量化模型与实验之间整体匹配程度的目标.

4 结 论

本文总结了模型确认中的确认度量所需的性质, 以此为依据分析了具有代表性的几类确认度量, 指出距离类确认度量相比于假设检验类确认度量更符合实际需求. 本文通过构造实验观测数据经验分布的置信包络, 计算出其与模型响应分布之间差异的上/下确界, 并给出了基于概率分布距离的确认度量与实验观测数量相联系的置信区间. 除此以外, 通过构造与实验观测数据有关的协方差矩阵, 提出了一种基于“单点”概率分布距离的多响应确认度量. 该确认度量考虑了各“点”数据间的相关性, 能在“多点”情况下给出模型响应与实验观测之间整体差异的量化结果, 符合文中所总结的“确认度量所需的性质”. 最后, 通过算例仿真中与其他确认度量的比较验证了本文度量的有效性.

参考文献(References)

- [1] Trucano T G, Swiler L P, Igusa T, et al. Calibration, validation, and sensitivity analysis: What's what[J]. *Reliability Engineering and System Safety*, 2006, 91(10/11): 1331-1357.
- [2] Box G E, Draper N R. *Empirical model-building and response surfaces*[M]. Minnesota: John Wiley Ltd, 1987: 424.
- [3] Trucano T G, Easterling R G, Dowding K L, et al. Description of the Sandia validation metrics project[R]. Albuquerque: Sandia National Laboratories, 2001.
- [4] Oberkampf W L, Trucano T G. Verification and validation benchmarks[R]. Albuquerque: Sandia National Laboratories, 2007.
- [5] Roy C J, Oberkampf W L. A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing[J]. *Computer Methods in Applied Mechanics and Engineering*, 2011, 200(25-28): 2131-2144.
- [6] Liu Y, Chen W, Arendt P, et al. Toward a better understanding of model validation metrics[J]. *J of Mechanical Design*, 2011, 133(7): 48-60.
- [7] Hills R G, Trucano T G. Statistical validation of engineering and scientific models: Background[R]. Albuquerque: Sandia National Laboratories, 1999.
- [8] Hills R G, Trucano T G. Statistical validation of engineering and scientific models: A maximum likelihood based metric[R]. Albuquerque: Sandia National Laboratories, 2002.
- [9] Buranathiti T, Cao J, Chen W, et al. Approaches for model validation: Methodology and illustration on a sheet metal flanging process[J]. *J of Manufacturing Science and Engineering*, 2006, 128(2): 588-597.
- [10] Rebba R, Mahadevan S. Validation of models with multivariate output[J]. *Reliability Engineering and System Safety*, 2006, 91(8): 861-871.
- [11] Rebba R, Mahadevan S. Computational methods for model reliability assessment[J]. *Reliability Engineering and System Safety*, 2008, 93(8): 1197-1207.
- [12] Oberkampf W L, Trucano T G, Hirsch C. Verification, validation, and predictive capability in computational engineering and physics[J]. *Applied Mechanics Reviews*, 2004, 57(3): 345-384.
- [13] Oberkampf W L, Barone M F. Measures of agreement between computation and experiment: Validation metrics[J]. *J of Computational Physics*, 2006, 217(1): 5-36.
- [14] Ferson S, Oberkampf W L, Ginzburg L. Model validation and predictive capability for the thermal challenge problem[J]. *Computer Methods in Applied Mechanics and Engineering*, 2008, 197(29-32): 2408-2430.
- [15] Miller L H. Table of percentage points of Kolmogorov statistics[J]. *J of the American Statistical Association*, 1956, 51(273): 111-121.
- [16] Dowding K J, Pilch M, Hills R G. Formulation of the thermal problem[J]. *Computer Methods in Applied Mechanics and Engineering*, 2008, 197(29-32): 2385-2389.
- [17] 张保强, 陈国平, 郭勤涛. 模型确认热传导挑战问题求解的贝叶斯方法[J]. *航空学报*, 2011, 32(7): 1202-1209. (Zhang B Q, Chen G P, Guo Q T. Solution of model validation thermal challenge problem using a Bayesian method[J]. *Acta Aeronautica et Astronautica Sinica*, 2011, 32(7): 1202-1209.)
- [18] McFarland J, Mahadevan S. Multivariate significance testing and model calibration under uncertainty[J]. *Computer Methods in Applied Mechanics and Engineering*, 2008, 197(29-32): 2467-2479.

(责任编辑: 滕 蓉)