

# 考虑相关性的多元输出仿真模型验证方法

林圣琳<sup>1</sup> 李伟<sup>1</sup> 杨明<sup>1</sup> 马萍<sup>1</sup>

**摘要** 复杂仿真模型一般具有多个不同类型且带有相关性的输出, 现有验证方法存在变量信息缺失、变量相关性度量不准确等问题. 为此, 提出基于变量选择和概率分布差异相结合的多变量仿真结果验证方法, 考虑不确定性的影响, 对选取到具有相关性的多个变量进行联合验证. 首先, 引入分形维数和互信息方法对多元异类输出进行相关性分析, 提取相关变量子集. 而后对相关变量子集中的各变量提取数据特征, 进而计算各相关变量子集关于数据特征的联合概率分布, 采用概率分布差异法度量仿真输出和参考输出联合概率分布的差异, 并将其转化为一致性程度; 在此基础上综合多个验证结果得到模型可信度. 通过实例应用及对比实验, 验证了方法的有效性.

**关键词** 仿真模型验证, 多变量输出, 变量相关性, 变量选择, 概率分布差异法

**引用格式** 林圣琳, 李伟, 杨明, 马萍. 考虑相关性的多元输出仿真模型验证方法. 自动化学报, 2019, 45(9): 1666–1678

**DOI** 10.16383/j.aas.c180456

## Multivariate Validation Method Under Correlation for Simulation Model

LIN Sheng-Lin<sup>1</sup> LI Wei<sup>1</sup> YANG Ming<sup>1</sup> MA Ping<sup>1</sup>

**Abstract** Complex simulation models often generate the multivariate and different types of output, some problems such as the lacking variables information and the inaccuracy correlation measurement are involved in the existing validation methods. A novel validation method combining variables selection with area metric is proposed, the multiple outputs with correlation are selected for the associated validation under uncertainty. The fractal dimension and mutual information methods are primarily applied to analyze the correlation among multivariate and divide responses and extract the correlated variable subsets. Next, the interesting data characteristics of all variables are extracted in subset and the corresponding joint cumulative distribution function (JCDF) of each subset related to any characteristic is calculated. The area metric is used to measure the difference between the simulation and reference output JCDFs of multivariate characteristics in each subset, and the differences are transformed into the consistency degrees. Then the multiple validation results are integrated to obtain the model credibility. Finally, the method is validated through the application case and comparison experiments.

**Key words** Simulation model validation, multivariate output, variable correlation, variable selection, area metric

**Citation** Lin Sheng-Lin, Li Wei, Yang Ming, Ma Ping. Multivariate validation method under correlation for simulation model. *Acta Automatica Sinica*, 2019, 45(9): 1666–1678

目前建模与仿真技术已成为人们认识和改造现实世界的重要手段. 由于仿真是一种基于模型的活动, 仿真模型是否可信成为用户十分关注的问题. 验证是仿真模型可信度评估的重要步骤<sup>[1]</sup>, 包含概念模型验证和结果验证. 仿真结果验证最直接而有效的方法是, 在相同输入条件下度量仿真输出与参考输出数据之间的一致性程度. 然而, 针对复杂系统建立的仿真模型往往具有不确定性、多元异类(动、静态)输出, 且各输出变量间可能存在相关性, 此时, 若仍采用传统仿真结果验证方法将导致验证结果不

准确. 因此, 考虑相关性及不确定性的多元输出仿真结果验证是需要重点研究的问题.

由于仿真模型和参考系统的输入及模型参数通常含有不确定性, 加之仿真模型运行和实际实验过程中引入的不确定性因素和误差, 导致仿真模型和参考系统的输出为随机变量或不确定的时间序列<sup>[2]</sup>. 考虑不确定性的影响, 静态输出结果验证方法的研究多数集中在概率框架, 形成了以参数估计<sup>[3–5]</sup>、假设检验<sup>[6–7]</sup>、贝叶斯因子<sup>[8–10]</sup>、证据距离<sup>[11]</sup>、概率分布差异法<sup>[12–13]</sup>为代表的 5 种解决方案. 其中, Oberkampff 等针对参考数据稀疏的情况, 采用插值和回归分析的方法估计参考输出的均值和标准差, 并与仿真输出的相应统计量进行对比, 得到置信区间形式的验证结果<sup>[5]</sup>; 同时, 假设检验和贝叶斯因子方法在静态仿真输出结果验证中的应用日趋完善, Jiang 等将贝叶斯区间假设检验应用于模型的分等级评估中<sup>[14]</sup>; 考虑到固有和认知不确定性的影响, 文

收稿日期 2018-07-02 录用日期 2018-11-19  
Manuscript received July 2, 2018; accepted November 19, 2018  
国家自然科学基金(61627810)资助  
Supported by National Natural Science Foundation of China (61627810)  
本文责任编辑 莫红  
Recommended by Associate Editor MO Hong  
1. 哈尔滨工业大学控制与仿真中心 哈尔滨 150080  
1. Control and Simulation Center, Harbin Institute of Technology, Harbin 150080

献 [11] 采用证据理论对动静输出进行描述, 并引入证据距离度量仿真和参考输出的一致性; Person 等提出了概率分布差异与 u-pooling 相结合的方法, 用于处理稀疏参考数据情况下的单输出仿真结果验证问题<sup>[12]</sup>, 该方法以其原理简单、可操作性强等优点得到了广泛应用.

考虑不确定性的同时, 复杂仿真模型可能存在多个输出变量的情况, 且各变量间可能存在函数或相关关系. 在单变量静态输出结果验证方法的基础上, 针对多元输出仿真结果验证方法的研究取得了一定的进展. 例如, Rebba 等最先提出了假设检验、贝叶斯因子与协方差相结合的多元输出结果验证方法, 并引入了非正态验证数据转化为正态数据的方法以满足假设检验的条件<sup>[15]</sup>; Jiang 等将区间贝叶斯因子方法进行推广, 将其应用至多元静态输出结果验证问题中<sup>[16]</sup>; Zhan 等提出了基于概率主成分分析 (Probabilistic principal component analysis, PPCA) 与贝叶斯因子相结合的方法, 用于解决带有不确定性和相关性的多元动态输出结果验证问题<sup>[17]</sup>. 同时, Li 等将概率积分变换 (Probability integral transformation, PIT) 与概率分布差异法相结合, 将多变量累积概率分布转化为单变量概率分布的形式, 采用概率分布差异法计算仿真和参考输出累积概率分布的差异<sup>[18]</sup>; Zhao 等分别计算仿真和参考输出与相应总体分布的马氏距离, 进而得到仿真和参考输出马氏距离的累积概率分布, 并应用概率分布差异法计算两者的差异<sup>[19]</sup>.

从单变量验证到多变量验证, 各变量间的关系是研究的重点, 现有验证方法存在验证结果不够准确和全面的问题. 利用传统单变量验证 (或结合多种预处理方法) 对各变量进行分别验证再综合, 一是对带有相关性的多个验证结果进行加权综合导致最终验证结果不够准确; 二是未考虑多变量间相关性的验证将导致验证结果不全面. 此外, 对复杂仿真模型进行结果验证, 其输出变量间的关系不够明确. 此时需首先明确输出变量间的关系 (独立/函数/相关关系) 再进行验证, 现有的多变量验证方法仅适用于变量间关系已知的情况. 同时, 多变量验证方法均利用协方差矩阵度量变量间的相关性, 这对非线性等其他相关关系将不再适用, 导致多变量间相关性度量不够准确.

为解决上述问题, 提出基于变量选择和概率分布差异相结合的多元输出仿真结果验证方法, 对具有不确定性的多元异类输出进行联合验证. 第 1 节对多元输出结果验证问题进行描述与分析, 指出现有方法存在的问题; 第 2 节给出多元静、动态输出的相关性分析及变量选择方法; 第 3 节提出基于数据特征提取和联合概率分布差异的多元输出仿真结果

验证方法; 第 4 节给出应用实例与对比实验结果; 第 5 节给出本文结论.

## 1 问题描述与分析

用  $S$  表示系统,  $S_s$  和  $S_r$  分别表示仿真模型和参考系统, 用  $\mathbf{A}_s = \{\mathbf{a}_{s1}, \mathbf{a}_{s2}, \dots, \mathbf{a}_{sp}\}$  和  $\mathbf{A}_r = \{\mathbf{a}_{r1}, \mathbf{a}_{r2}, \dots, \mathbf{a}_{rp}\}$  分别表示仿真模型和参考系统的输入变量集,  $p$  为输入变量个数,  $\mathbf{Y}_s = \{\mathbf{y}_{s1}, \mathbf{y}_{s2}, \dots, \mathbf{y}_{sm}\}$  和  $\mathbf{Y}_r = \{\mathbf{y}_{r1}, \mathbf{y}_{r2}, \dots, \mathbf{y}_{rm}\}$  分别表示仿真模型和参考系统的输出变量集,  $m$  为输出变量个数, 多元异类输出集  $\mathbf{Y}_s$ 、 $\mathbf{Y}_r$  中的静态输出表示为随机变量, 动态输出表示为多个时间序列集合的形式. 假设  $\mathbf{y}_{si}$ 、 $\mathbf{y}_{sj}$  分别为仿真模型的某一动态和静态输出, 则有

$$\mathbf{y}_{si} = \begin{bmatrix} y_{si}^1(t_1) & y_{si}^1(t_2) & \cdots & y_{si}^1(t_N) \\ y_{si}^2(t_1) & y_{si}^2(t_2) & \cdots & y_{si}^2(t_N) \\ \vdots & \vdots & \ddots & \vdots \\ y_{si}^n(t_1) & y_{si}^n(t_2) & \cdots & y_{si}^n(t_N) \end{bmatrix}$$

$$\mathbf{y}_{sj} = [y_{sj}^1, y_{sj}^2, \dots, y_{sj}^n]^T \quad (1)$$

式中,  $i, j \in [1, m]$ , 且  $i \neq j$ ;  $N$  为时间序列的长度;  $t_1, t_2, \dots, t_N$  表示时间序列的时刻点; 考虑不确定性的影响, 需要进行多次仿真和实际实验,  $n$  为重复实验次数. 用  $C(\mathbf{Y}_s, \mathbf{Y}_r)$  表示在  $\mathbf{A}_s = \mathbf{A}_r$  时,  $\mathbf{Y}_s$  相对于  $\mathbf{Y}_r$  的一致性程度, 且  $C(\mathbf{Y}_s, \mathbf{Y}_r) \in (0, 1]$ . 当  $\mathbf{Y}_s$  与  $\mathbf{Y}_r$  完全一致, 则有  $C(\mathbf{Y}_s, \mathbf{Y}_r) = 1$ ; 当  $\mathbf{Y}_s$  相对于  $\mathbf{Y}_r$  的一致性程度越差, 表示仿真模型越不可信, 则有  $C(\mathbf{Y}_s, \mathbf{Y}_r) \rightarrow 0$ <sup>[1]</sup>.

假设  $\mathbf{Y}_{sJ} \in \mathbf{R}^{n_s \times m}$  与  $\mathbf{Y}_{rJ} \in \mathbf{R}^{n_r \times m}$  为多元仿真模型和参考系统的静态输出变量,  $n_s$ 、 $n_r$  表示仿真和实际实验的重复运行次数. 针对带有相关性的多元静态输出结果验证方法主要有:

1) 基于假设检验和马氏距离相结合的方法. 文献 [16] 给出基于似然比检验和马氏距离相结合的验证方法, 得到最终一致性结果.

2) 基于主成分分析的方法<sup>[17]</sup>. 对  $\mathbf{Y}_{sJ}$  和  $\mathbf{Y}_{rJ}$  进行降维, 去除变量间相关性, 得到新的输出变量  $\mathbf{Y}_{sJ}^{new} = [y_{sJ1}^{new}, y_{sJ2}^{new}, \dots, y_{sJ\eta}^{new}]$  和  $\mathbf{Y}_{rJ}^{new} = [y_{rJ1}^{new}, y_{rJ2}^{new}, \dots, y_{rJ\eta}^{new}]$ ,  $\eta \leq m$  为主成分的个数, 而后采用现有静态输出验证方法对若干主成分进行逐一验证并综合得到最终验证结果.

3) 基于概率分布差异的方法<sup>[18]</sup>. 分别计算  $m$  维  $\mathbf{Y}_{sJ}$  和  $\mathbf{Y}_{rJ}$  的联合累积概率分布 (Cumulative distribution function, CDF) 函数并作差, 获得仿真和参考输出数据的差异, 得到  $[0, +\infty)$  的误差度量结果.

针对带有相关性的多元动态输出结果验证问

题,常用方法为基于数据特征和主成分分析相结合的方法<sup>[17]</sup>.首先分别提取动态输出数据的特征矩阵,而后采用基于主成分分析的多元静态输出验证方法获得最终验证结果.针对上述多元输出仿真结果验证方法进行分析,存在以下问题需要进一步研究:

1) 复杂仿真模型常存在多元输出变量间的相关或独立关系未知的情况,目前方法均是在变量关系已知的前提下进行研究,存在一定局限性;

2) 利用主成分分析获取的多元输出变量的主成分是线性变换后的结果,被提取主成分所代表的变量含义不够明确,同时对多元输出变量进行降维将导致验证信息丢失,使验证结果不够准确和全面;

3) 采用协方差矩阵度量变量相关性,需假设变量样本服从正态分布,且仅能描述多元输出变量间的线性关系,无法度量变量间非线性等其他相关关系,进而导致变量间相关性度量不准确;

4) 基于联合概率分布差异法可直接度量多元静态输出变量间的差异,需要已知变量间的独立或相关关系,同时,处理多元动态输出存在局限,得到的差异度量结果无法刻画仿真模型的可信度.

为解决上述问题,可采用基于变量选择和概率分布差异相结合的多元输出仿真结果验证方法,考虑不确定性的影响,对选取到具有相关性的多变量进行联合验证.首先,引入变量选择方法分别对  $\mathbf{Y}_s$ 、 $\mathbf{Y}_r$  进行相关性分析,提取相关变量子集(又称相关变量组,子集中各变量是相关的,各子集中变量数的和为输出变量总数),进而得到多个独立的变量子集;同时,提取相同变量子集中多变量的数据特征,对于静态输出选取数据本身作为变量特征,对于动态输出选取距离、形状以及频谱特征;而后计算变量子集中多个变量关于某特征的联合 CDF 差异,并将其转化为可信度;最后将多个变量子集关于若干数据特征的一致性与多个动态输出均值曲线的一致性进行综合得到仿真模型可信度.

## 2 多元输出相关变量选择方法

为明确复杂仿真模型中多元输出变量间的独立或相关关系,引入数据挖掘领域的相应方法对多变量进行相关性分析,进而提取相关变量子集.本文仅

考虑同种类型(静态或动态)输出变量间的相关性,利用分形维数和互信息方法分别对静、动态输出变量进行相关性分析.

### 2.1 基于分形维数的静态相关变量选择方法

对随机变量的相关性分析集中于 Pearson 相关系数,它仅能度量变量的线性关系,并对变量间强相关性较敏感,其结果受奇异值的影响较大,无法适应具有非线性、不确定性以及非正态分布的数据集.其他一些相关系数,如 Kendall 系数、Spearman 系数等虽可以描述非线性相关关系,但却不能完全地刻画变量间的相关性结构.此外,数据挖掘领域常用的变量选择方法,如奇异值分解法(Singular value decomposition, SVD)、主成分分析法(Principal component analysis, PCA)、基于神经网络的方法(Neural networks, NN)、基于 k-邻近方法(K-nearest neighbor, KNN)、基于决策树的方法(Decision tree, DT)、基于贝叶斯网络的方法(Bayesian network, BN)以及基于分形维数的方法(Fractal dimension, FD)等,具有不同的特点,对其进行对比分析如表 1 所示.

由表 1 可知, SVD 和 PCA 方法得到的变量子集失去了其原有的含义,且只能对具有线性相关性的变量集进行分析;而基于机器学习的方法需要训练样本集作为支撑,其运行速度受到变量个数的影响较大,导致变量个数较多时运行速度较慢;而基于分形维数的方法不仅能够度量线性相关性,还能度量非线性等其他相关关系,具有不需要训练样本集和运行速度快等优点.因此,本文引入基于分形维数<sup>[20]</sup>的方法对多元输出变量进行分析,提取相关变量子集.假设  $\mathbf{Y}_{s,J}$ 、 $\mathbf{Y}_{r,J}$  为仿真和参考多元静态输出变量集,以  $y_{s,J}^i, i \in [1, m]$  为例给出  $\mathbf{Y}_{s,J}$  的相关变量子集提取步骤如下.

**步骤 1.** 根据自相似性原理计算  $y_{s,J}^i$  的局部固有维度  $pD(\cdot)$ :

$$pD(y_{s,J}^i) \equiv \frac{\partial \log \left( \sum_i C_{a,i}^2 \right)}{\partial \log(a)}, \quad a \in [a_1, a_2] \quad (2)$$

表 1 常用变量选择方法对比

Table 1 Comparison of general variable selection methods

| 变量选择方法 | 是否为原变量集的子集 | 是否支持非线性相关关系 | 个体决策所占比例 | 是否需要训练样本集 | 运行速度与变量个数的关系 |
|--------|------------|-------------|----------|-----------|--------------|
| SVD    | 否          | 否           | 是        | 否         | 线性增长         |
| PCA    | 否          | 否           | 是        | 否         | 线性增长         |
| KNN    | 是          | 是           | 是        | 是         | 指数增长         |
| DT     | 是          | 是           | 是        | 是         | 指数增长         |
| BN     | 是          | 是           | 是        | 是         | 指数增长         |
| FD     | 是          | 是           | 是        | 否         | 线性增长         |

式中,  $r$  表示将  $y_{s,J}^i$  划分成  $2^v$  个相等大小区间的长度,  $v$  为划分深度,  $C_{a,i}$  表示  $y_{s,J}^i$  中落入第  $i$  个区间的样本个数;

**步骤 2.** 设  $c = 1$ , 移除  $\mathbf{Y}_{s,J}$  中  $pD(y_{s,J}^i) < \xi$  的变量  $y_{s,J}^i$ ,  $\xi$  为预定义的固有维度阈值, 排除的变量为独立变量, 并按照  $pD(\cdot)$  大小将  $y_{s,J}^i$  进行降序排列, 形成新变量集  $\mathbf{Y}'_{s,J}$ , 其变量个数为  $m'$ ;

**步骤 3.** 计算  $pD(\{y_{s,J}^1\}), pD(\{y_{s,J}^1, y_{s,J}^2\}), \dots$ , 直到  $|pD(\{y_{s,J}^1 \cdot y_{s,J}^k\}) - pD(\{y_{s,J}^1 \cdot y_{s,J}^{k-1}\})| < \xi \cdot pD(y_{s,J}^k), k = 1, 2, \dots, m'$ ;

**步骤 4.** 若  $k = m'$  且  $|pD(\{y_{s,J}^1 \cdot y_{s,J}^k\}) - pD(\{y_{s,J}^1 \cdot y_{s,J}^{k-1}\})| \geq \xi \cdot pD(y_{s,J}^k)$ , 则算法结束;

**步骤 5.** 设相关性变量超集  $\xi SG_c = \{y_{s,J}^1, \dots, y_{s,J}^k\}$ , 并提取  $\xi SG_c$  中的相关变量子集  $\xi G_c$  和相关变量基  $\xi B_c$ , 具体算法见文献 [20], 并设循环变量  $j = k + 1$ ;

**步骤 6.** 若  $|pD(\xi B_c \cup \{y_{s,J}^j\}) - pD(\xi B_c)| < \xi \cdot pD(y_{s,J}^j)$ , 则执行下一步, 否则转至步骤 8;

**步骤 7.** 对于  $\xi B_c$  中的每个变量  $y_{s,J}^b$ , 若  $|pD(\xi B_c \cup \{y_{s,J}^j\}) - pD((\xi B_c - \{y_{s,J}^b\}) \cup \{y_{s,J}^j\})| < \xi \cdot pD(y_{s,J}^b)$  和  $|pD(\xi B_c \cup \{y_{s,J}^j\}) - pD((\xi B_c - \{y_{s,J}^b\}) \cup \{y_{s,J}^j\})| \geq \xi \cdot pD(y_{s,J}^j)$  同时成立, 则将  $y_{s,J}^j$  加入  $\xi G_c$ ;

**步骤 8.** 执行  $j \leftarrow j + 1$ , 若  $j > m'$ , 则转至下一步, 否则转至步骤 6;

**步骤 9.** 移除  $\mathbf{Y}'_{s,J}$  中  $\xi G_c - \xi B_c$  的变量, 并输出相关变量子集  $\xi G_c$  和相关变量基  $\xi B_c$ ;

**步骤 10.** 执行  $c \leftarrow c + 1$ , 并转至步骤 3.

通过上述步骤提取  $\mathbf{Y}_{s,J}$  和  $\mathbf{Y}_{r,J}$  的相关变量子集  $\mathbf{G}_{s,J}^i, \mathbf{G}_{r,J}^j$  如下.

$$\begin{cases} \mathbf{G}_{s,J}^i = [y_{s,J}^{i1}, y_{s,J}^{i2}, \dots, y_{s,J}^{im_{si}}], i = 1, 2, \dots, \beta_s \\ \mathbf{G}_{r,J}^j = [y_{r,J}^{j1}, y_{r,J}^{j2}, \dots, y_{r,J}^{jm_{rj}}], j = 1, 2, \dots, \beta_r \end{cases} \quad (3)$$

式中,  $\beta_s, \beta_r$  分别为提取  $\mathbf{Y}_{s,J}$  和  $\mathbf{Y}_{r,J}$  相关变量子集的个数,  $m_{si}, m_{rj}$  分别为  $\mathbf{G}_{s,J}^i, \mathbf{G}_{r,J}^j$  中变量的个数, 且有  $m_{s1} + m_{s2} + \dots + m_{si} = m_{r1} + m_{r2} + \dots + m_{rj} = m$ .

### 2.2 基于互信息的动态相关变量选择方法

与随机变量不同, 多元动态输出变量与时间有关, 其相关性分析与变量选择需从时间序列整体的角度进行分析. 一些传统的随机变量相关性分析方法对于多元动态变量同样适用, 例如 Pearson 系数、Kendall 系数、Spearman 系数等, 但无法用于动态输出变量具有多个样本 (时间序列) 的情况. 此外, 一些统计学分析方法, 如 Granger 因果关系分

析<sup>[21]</sup>、典型相关分析<sup>[22]</sup>、Copula 分析<sup>[23]</sup>、灰色关联分析<sup>[24]</sup> 以及互信息分析<sup>[25]</sup> 等同样能够用于多变量的相关性分析. Granger 因果关系分析只能定性地分析变量间的因果关系, 而无法得到定量的结果; 典型相关分析对观测值的顺序不会做出响应, 因此无法解决时间序列问题; Copula 分析需要建立在对边缘分布的合理假设之上, 使其应用受到限制; 灰色关联分析仅从形状相关性的角度对时间序列进行分析, 其相关性度量不够全面.

基于互信息的相关性分析方法能够度量动态输出变量间任意类型的关系, 互信息以信息熵为理论基础, 它能够度量变量取值的不确定性程度, 进而描述变量的信息含量大小<sup>[26]</sup>, 通常用于多种类型时间序列的特征提取和结构化预测<sup>[27]</sup>. 然而, 互信息同样存在不能完整刻画变量集相关性结构的缺点, 因此本文引入类可分性和变量可分性提取多元动态输出的相关变量子集<sup>[28]</sup>. 假设  $\mathbf{Y}_{s,D}, \mathbf{Y}_{r,D}$  为仿真和参考多元动态输出变量集, 同样以  $\mathbf{Y}_{s,D}$  为例, 给出变量选择步骤如下.

**步骤 1.** 计算  $\mathbf{Y}_{s,D}$  的  $m \times m$  维互信息矩阵, 具体算法见文献 [26];

**步骤 2.** 分别计算每一维变量的类间离散度  $\Omega_{bi}$  和类内离散度  $\Omega_{wi}$ :

$$\begin{cases} \Omega_{bi} = \sum_{i=1}^{C_{sam}} q_i (\mu_i - \mu) (\mu_i - \mu)^T \\ \Omega_{wi} = \sum_{i=1}^{C_{sam}} \sum_{j=1}^{q_i} (\mu_i - y_{s,D}^j) (\mu_i - y_{s,D}^j)^T \end{cases} \quad (4)$$

式中,  $C_{sam}$  为样本类别总数,  $q_i$  为属于第  $i$  类的样本个数,  $\mu = (1/n_s) \sum_{i=1}^{n_s} y_{s,D}^i, \mu_i = (1/q_i) \sum_{i=1}^{q_i} y_{s,D}^i$ . 按照每个变量的类可分离性大小, 进行变量排序:

$$J_i = \frac{\Omega_{bi}}{\Omega_{wi}}, \quad i = 1, 2, \dots, m \quad (5)$$

**步骤 3.** 取  $J_i$  值最大的变量为变量子集  $\mathbf{G}_{s,D}^i$  的第一个变量;

**步骤 4.** 选择使下式最大的变量为  $\mathbf{G}_{s,D}^i$  的下一个变量:

$$\begin{cases} J_i = \frac{\Omega_{bi} + \Omega_{fi}}{\Omega_{wi}} \\ \Omega_{fi} = \frac{1}{|\mathbf{G}_{s,D}^i|} \sum_{k=1}^{|\mathbf{G}_{s,D}^i|} \sum_{o=1}^{C_{sam}} q_{ko} (\mu_o - \mu_{ko}) \cdot (\mu_o - \mu_{ko})^T \end{cases} \quad (6)$$

式中,  $|\mathbf{G}_{s,D}^i|$  为子集  $\mathbf{G}_{s,D}^i$  的变量个数,  $\mu_{ko}$  为子集  $\mathbf{G}_{s,D}^i$  中属于第  $o$  类的第  $k$  个变量的均值;

**步骤 5.** 当  $|\mathbf{G}_{s,D}^i| = \varepsilon$ , 则算法终止, 其中,  $\varepsilon$  为预设值, 否则转至步骤 4. 通过上述步骤得到相关变

量子集  $\mathbf{G}_{sD}^i$ 、 $\mathbf{G}_{rD}^j$  如下:

$$\begin{cases} \mathbf{G}_{sD}^i = [y_{sD}^{i1}(t), y_{sD}^{i2}(t), \dots, y_{sD}^{im_{si}}(t)] \\ \mathbf{G}_{rD}^j = [y_{rD}^{j1}(t), y_{rD}^{j2}(t), \dots, y_{rD}^{jm_{rj}}(t)] \end{cases} \quad (7)$$

式中,  $i = 1, 2, \dots, \beta_s, j = 1, 2, \dots, \beta_r$ .  $\beta_s$ 、 $\beta_r$  的含义与式 (3) 相同.

### 3 多元输出仿真结果验证方法

考虑不确定的影响, 若对每一时刻的多元动态输出变量进行分析势必导致维数爆炸. 为此提出基于特征的验证方法, 首先提取用户关注的输出数据特征, 而后计算每个特征下多变量联合概率分布的差异, 并将其转化为可信度结果, 最后综合多个验证结果得到模型可信度.

#### 3.1 多元输出数据特征提取方法

针对静态输出, 选取数据本身作为变量特征. 假设  $\mathbf{Y}_s \in \mathbf{R}^{n_s \times m}$  与  $\mathbf{Y}_r \in \mathbf{R}^{n_r \times m}$  为多元仿真和参考静态输出变量, 其数据特征描述为

$$\begin{cases} \mathbf{e}_{sij} = [y_{sij}]_{n_s \times m}, i = 1, 2, \dots, n_s \\ \mathbf{e}_{rij} = [y_{rij}]_{n_r \times m}, i = 1, 2, \dots, n_r \end{cases}, j = 1, 2, \dots, m \quad (8)$$

对于动态输出  $y_{sij}(t)$ 、 $y_{rik}(t)$ ,  $i = 1, 2, \dots, m, j = 1, 2, \dots, n_s, k = 1, 2, \dots, n_r$  则选取  $n_s$ 、 $n_r$  次系统运行得到的输出均值曲线  $\bar{y}_{si}(t)$ 、 $\bar{y}_{ri}(t)$  作为基准, 与每次实验得到的输出曲线进行对比, 求取相应的均值曲线的第  $l$  个特征  $e_{sij}^l$ 、 $e_{rik}^l$ :

$$\begin{cases} e_{sij}^l = \Phi_l(\bar{y}_{si}(t), y_{sij}(t)) \\ e_{rik}^l = \Phi_l(\bar{y}_{ri}(t), y_{rik}(t)) \end{cases}, l = 1, 2, \dots, L_i \quad (9)$$

式中,  $L_i$  为第  $i$  个输出的特征数,  $\Phi_l(\cdot)$  为第  $l$  个特征度量模型.

提取动态输出特征前, 需要先对动态输出进行归类<sup>[1]</sup>. 以第  $j$  个动态输出的第  $i$  次实现  $y_{ij}(t)$  为例, 其对应的时间变化序列为  $[t_1, t_2, \dots, t_N]$ . 则定义  $y_{ij}(t)$  随时间变化的频率为:

$$F_{ij} = \frac{\sum_{k=1}^{N-1} \left| \frac{\Delta y_{ij}(t_k)}{\Delta t_k} \right|}{|\bar{y}_{ij}|} \quad (10)$$

式中,  $F_{ij} \geq 0$  为  $y_{ij}(t)$  的变化频率;  $\Delta y_{ij}(t_k) = y_{ij}(t_{k+1}) - y_{ij}(t_k)$ ;  $\Delta t_k = t_{k+1} - t_k$ ;  $|\bar{y}_{ij}| = \sum_{k=1}^N |y_{ij}(t_k)|/N \neq 0$ . 给定  $F_0$  为判断时间序列变化快慢的临界值, 若  $F_{ij} \geq F_0$ , 则认为  $y_{ij}(t)$  为速变数据, 否则为缓变数据.

为刻画不确定性对系统输出的影响, 从距离和形状两方面提取缓变数据的特征. 在前期工作<sup>[29]</sup> 的基础上, 给出第  $j$  个仿真输出的第  $i$  次实现  $y_{sij}(t)$  与其均值曲线  $\bar{y}_{sj}(t)$  的距离和形状差异  $e_{sd}^{ij}$ 、 $e_{sc}^{ij}$  的度量公式如下.

$$\begin{cases} e_{sd}^{ij} = \frac{1}{N} \sqrt{\sum_{t=t_1}^{t_N} z^2(t)} \\ e_{sc}^{ij} = \frac{1}{N} \sqrt{\sum_{t=t_1}^{t_N} (z(t) - \bar{z})^2} \end{cases} \quad (11)$$

式中,  $z(t) = y_{sij}(t) - \bar{y}_{sj}(t), t = t_1, t_2, \dots, t_N, \bar{z} = \sum_{t=t_1}^{t_N} z(t)/N$ . 另外, 选取谱密度特征度量速变数据  $y_{sij}(t)$  与相应均值曲线  $\bar{y}_{sj}(t)$  的差异  $e_{sh}^{ij}$ , 定义如下.

$$e_{sh}^{ij} = 1 - \frac{\gamma}{M} \quad (12)$$

式中,  $e_{sh}^{ij}$  表示速变数据  $y_{sij}$  与  $\bar{y}_{sj}$  的谱密度差异;  $M$  表示  $y_{sij}(t)$  和  $\bar{y}_{sj}(t)$  转换至频域中的点数;  $\gamma$  表示通过相容性检验的点数. 根据式 (2) ~ (7) 得到  $\mathbf{Y}_s$  的第  $i$  个相关量子集  $\mathbf{G}_{si}$ 、 $\mathbf{Y}_r$  的第  $j$  个相关量子集  $\mathbf{G}_{rj}$  关于第  $l$  个特征的特征度量矩阵分别为

$$\begin{aligned} \mathbf{E}_{si}^l &= \begin{bmatrix} e_{si1}^{l1} & e_{si2}^{l1} & \dots & e_{sim_{si}}^{l1} \\ e_{si1}^{l2} & e_{si2}^{l2} & \dots & e_{sim_{si}}^{l2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{si1}^{ln_s} & e_{si2}^{ln_s} & \dots & e_{sim_{si}}^{ln_s} \end{bmatrix} \\ \mathbf{E}_{rj}^l &= \begin{bmatrix} e_{rj1}^{l1} & e_{rj2}^{l1} & \dots & e_{rjm_{rj}}^{l1} \\ e_{rj1}^{l2} & e_{rj2}^{l2} & \dots & e_{rjm_{rj}}^{l2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{rj1}^{ln_r} & e_{rj2}^{ln_r} & \dots & e_{rjm_{rj}}^{ln_r} \end{bmatrix} \end{aligned} \quad (13)$$

式中,  $i = 1, 2, \dots, \alpha, j = 1, 2, \dots, \beta$ . 若  $\mathbf{G}_{si}$  和  $\mathbf{G}_{rj}$  均为静态输出变量子集, 则  $L_i = 1$ ; 若  $\mathbf{G}_{si}$  和  $\mathbf{G}_{rj}$  均为缓变输出变量子集, 则  $L_i = 2$ ; 若  $\mathbf{G}_{si}$  和  $\mathbf{G}_{rj}$  均为速变输出变量子集, 则  $L_i = 1$ . 需要说明的是, 在某些特殊仿真应用中, 除了上述特征外, 通常还需关注数据本身的一些特征, 例如控制系统阶跃响应中的上升时间、超调量以及稳态误差, 位置数据中的变化趋势, 测量数据中的噪声等. 在进行实际验证中, 特征矩阵包含两部分内容, 一部分为上文给出的数据特征, 另一部分为根据具体领域知识确定的数据特征.

#### 3.2 基于概率分布差异的特征一致性分析方法

以  $\mathbf{E}_{si}^l$  为例进行分析, 用  $v$  维随机变量

$x_1, x_2, \dots, x_v$  代替其列向量  $[e_{s1}^l, e_{s2}^l, \dots, e_{sm_{si}}^l]$ ,  $v = m_{si}$ , 采用多维随机变量概率分布定义  $x_1, x_2, \dots, x_v$  的联合 CDF:

$$F(x_1, x_2, \dots, x_v) = P\left\{ (X_1 \leq x_1) \cup (X_2 \leq x_2) \cup \dots \cup (X_v \leq x_v) \right\} \quad (14)$$

将  $v$  维空间划分为等尺寸的  $\rho^v$  个区域, 遍历  $v$  维变量的  $\rho$  个取值区间, 若  $x_1 < X_1^0, x_2 < X_2^0, \dots, x_v < X_v^0$ , 则  $F(x_1, x_2, \dots, x_v) = 0$ ; 若  $x_1 < X_1^k, x_2 < X_2^k, \dots, x_v < X_v^k$ , 则  $F(x_1, x_2, \dots, x_v) = k/\rho^v$  等. 如果变量集  $x_1, x_2, \dots, x_v$  在第  $k$  个区间内的样本量为 1, 则  $F$  在  $x_1, x_2, \dots, x_v$  点的跳跃度为  $1/\rho^v$ , 如果变量集  $x_1, x_2, \dots, x_v$  在第  $k$  个区间内有  $\varepsilon$  个样本, 则  $F$  在  $x_1, x_2, \dots, x_v$  点的跳跃度是  $\varepsilon/\rho^v$ . 给出  $E_{si}^l$  和  $E_{rj}^l$  联合 CDF 间的差异如下.

$$D(F_s, F_r) = \int \int \dots \int |F_s(x_1, x_2, \dots, x_v) - F_r(x_1, x_2, \dots, x_v)| dx_1 dx_2 \dots dx_v \quad (15)$$

为计算联合 CDF 的差异  $D(F_r, F_s)$ , 可将上式改写为下面积分之差的形式:

$$D = \int F_s(\mathbf{x}) d\mathbf{x} - \int F_r(\mathbf{x}) d\mathbf{x} = I_s - I_r \quad (16)$$

式中  $\mathbf{x} = [x_1, x_2, \dots, x_v]$ . 假设分别用  $\hat{I}_s$  和  $\hat{I}_r$  估计  $I_s$  和  $I_r$ , 并用  $\hat{D} = \hat{I}_s - \hat{I}_r$  估计  $\hat{D}$ , 则  $\hat{D}$  的方差为:

$$Var(\hat{D}) = Var(\hat{I}_s) + Var(\hat{I}_r) - 2Cov(\hat{I}_s, \hat{I}_r) \quad (17)$$

显然, 在  $Var(\hat{I}_s)$  和  $Var(\hat{I}_r)$  确定后,  $\hat{I}_s$  和  $\hat{I}_r$  的正相关度越高, 则  $\hat{D}$  的方差越小. 本文采用重要性抽样法估计  $I_s$  和  $I_r$ , 即改写  $D$  为

$$D = \int H_s(\mathbf{x}) g_s(\mathbf{x}) d\mathbf{x} - \int H_r(\mathbf{x}) g_r(\mathbf{x}) d\mathbf{x} \quad (18)$$

其中,  $\mathbf{x} = x_1, x_2, \dots, x_v$ ,  $g_s(\mathbf{x})$ 、 $g_r(\mathbf{x})$  是两个密度函数,  $H_s(\mathbf{x}) = F_s(\mathbf{x})/g_s(\mathbf{x})$ ,  $H_r(\mathbf{x}) = F_r(\mathbf{x})/g_r(\mathbf{x})$ . 首先, 由  $g_s(\mathbf{x})$ 、 $g_r(\mathbf{x})$  各产生  $P$  个相互独立的  $v$  维随机数  $\mathbf{T}_{s1}, \dots, \mathbf{T}_{sP}$  和  $\mathbf{T}_{r1}, \dots, \mathbf{T}_{rP}$ , 并计算

$$\hat{D} = \frac{1}{P} \sum_{k=1}^P (H_s(\mathbf{T}_{sk}) - H_r(\mathbf{T}_{rk})) \quad (19)$$

采用逆变换方法由同一个  $v$  维联合均匀分布  $U(0, 1)$  产生  $\mathbf{T}_{s1}, \dots, \mathbf{T}_{sP}$  和  $\mathbf{T}_{r1}, \dots, \mathbf{T}_{rP}$ , 能够

保证两组随机数具有较高的正相关程度, 进而使  $Var(\hat{D})$  较小, 对  $\hat{D}$  的估计值趋于稳定.

需要说明的是,  $D(F_s, F_r) \in [0, \infty)$  仅是仿真和参考输出特征的联合 CDF 的差异 (如图 1 所示), 其取值范围为  $[0, \infty)$ , 此时无法给出仿真和参考输出的一致性程度 (即取值为  $[0, 1]$  的相对值). 因此, 提出将  $D(F_s, F_r)$  向可信度  $C(F_s, F_r)$  转化的公式如下.

$$C(F_s, F_r) = \frac{\prod_{i=1}^v (X_i^{\max} - X_i^{\min}) - D(F_s, F_r)}{\prod_{i=1}^v (X_i^{\max} - X_i^{\min})} \quad (20)$$

式中,  $\prod_{i=1}^v (X_i^{\max} - X_i^{\min})$  表示  $v$  维样本空间所占区域的大小;  $X_i^{\min} = \min(X_{si}^{\min}, X_{ri}^{\min})$ ,  $X_i^{\max} = \max(X_{si}^{\max}, X_{ri}^{\max})$  表示第  $i$  维变量的样本极值. 显然,  $C(F_s, F_r)$  满足如下性质<sup>[13]</sup>, 进而能够用于度量仿真模型可信度.

- 性质 1. 非负性:  $C(F_s, F_r) \geq 0$ ;
- 性质 2. 交换性:  $C(F_s, F_r) = C(F_r, F_s)$ ;
- 性质 3. 有界性:  $C(F_s, F_r) \in [0, 1]$ ;
- 性质 4. 同一性:  $C(F_s, F_r) = 1$ , 当且仅当  $F_s = F_r$ .

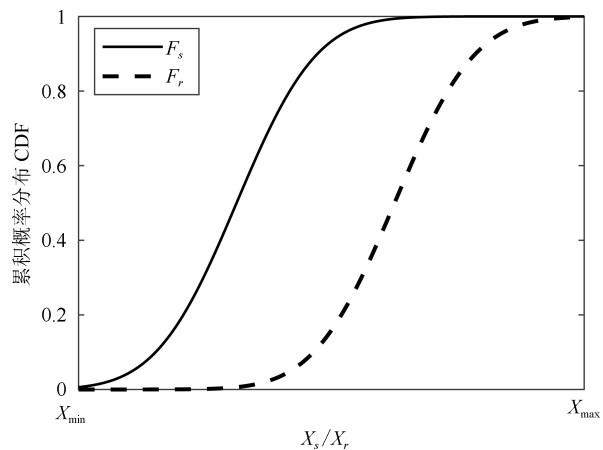


图 1 参考与仿真输出的 CDF 对比

Fig. 1 Comparing CDF curves of reference and simulation output

### 3.3 多元输出仿真结果验证方法流程

基于前文所述方法, 给出考虑相关性的多元输出仿真结果验证流程如图 2 所示.

- 1) 考虑不确定因素的影响, 分别进行  $n$  次仿真运行和实际试验, 获取多元仿真和参考输出  $\mathbf{Y}_s = \{\mathbf{Y}_{s1}, \mathbf{Y}_{s2}, \dots, \mathbf{Y}_{sm}\}$ ,  $\mathbf{Y}_r = \{\mathbf{Y}_{r1}, \mathbf{Y}_{r2}, \dots, \mathbf{Y}_{rm}\}$ ;
- 2) 利用多元输出变量选择方法提取  $\mathbf{Y}_s$ 、 $\mathbf{Y}_r$

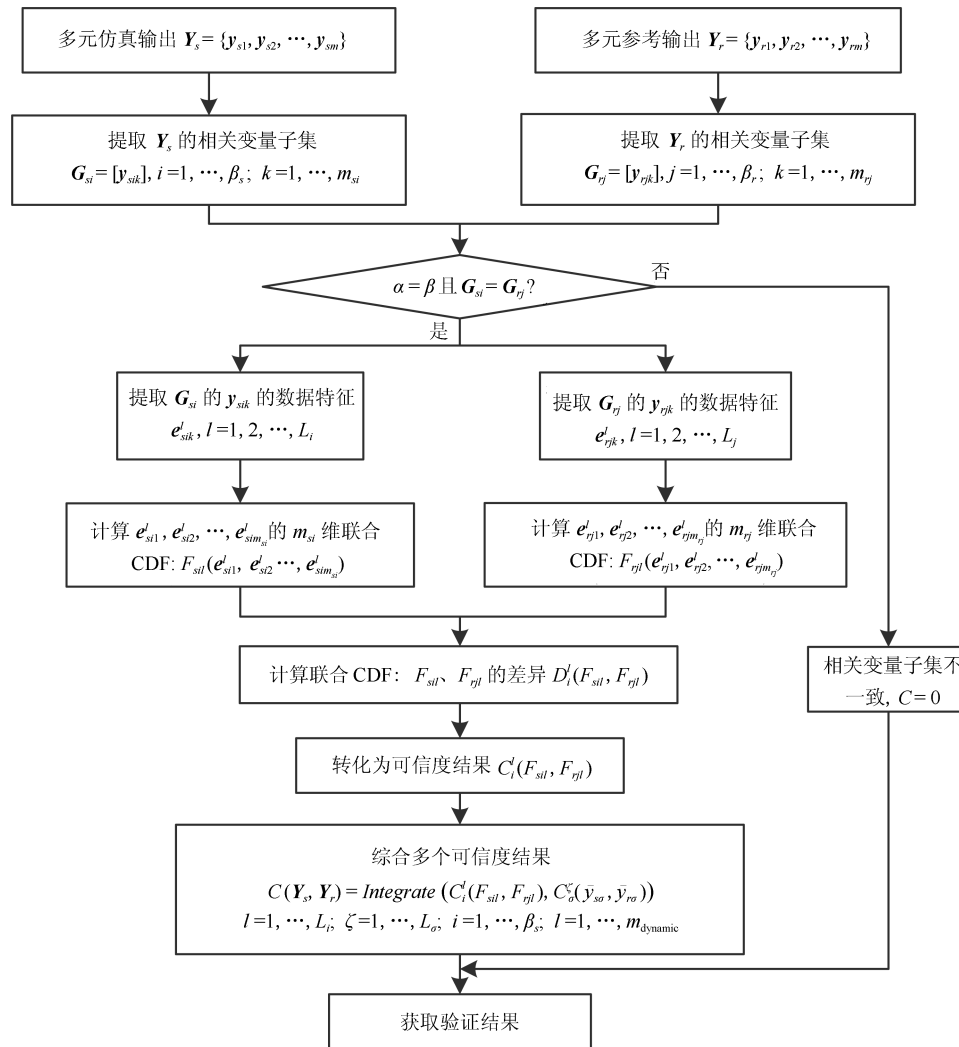


图 2 考虑相关性的多元输出仿真结果验证方法流程

Fig. 2 Procedures of multivariate simulation result validation under correlation

的相关变量子集  $\mathbf{G}_{si}, i = 1, \dots, \beta_s, \mathbf{G}_{rj}, j = 1, \dots, \beta_r$ ;

3) 若  $\beta_s = \beta_r$  且  $\mathbf{G}_{si} = \mathbf{G}_{rj}$ , 则依据式 (8) ~ (13) 提取  $\mathbf{G}_{si}, \mathbf{G}_{rj}$  中各变量的数据特征  $e_{sik}^l, e_{rik}^l$ ; 反之, 若  $\beta_s \neq \beta_r$  或  $\mathbf{G}_{si} \neq \mathbf{G}_{rj}$  的相关变量子集, 则认为该仿真模型不可信, 即  $C = 0$ , 算法结束;

4) 依据式 (14) 分别计算数据特征变量集  $e_{si1}^l, e_{si2}^l, \dots, e_{sim_{si}}^l$  和  $e_{rj1}^l, e_{rj2}^l, \dots, e_{rjm_{rj}}^l$  的联合 CDF:  $F_{sil}, F_{rjl}$ ;

5) 依据式 (15) ~ (19) 计算特征变量集的联合 CDF:  $F_{sil}, F_{rjl}$  的差异  $D_i^l(F_{sil}, F_{rjl})$ ;

6) 依据式 (20) 将  $D_i^l(F_{sil}, F_{rjl})$  转化为可信度结果  $C_i^l(F_{sil}, F_{rjl})$ ;

7) 通过 2) 可知,  $\alpha$  个相关变量子集之间是相互独立的, 且用户关注的多个数据特征 (包括位置、形状、频谱) 间也可认为是独立的, 进而可

采用加权方法综合多个可信度结果  $C_i^l(F_{sil}, F_{rjl})$ ,  $l = 1, \dots, L_i; i = 1, \dots, \beta_s; j = 1, \dots, \beta_r$ . 图 2 中 “Integrate( $\cdot$ )” 表示加权综合算子. 同时第  $\sigma$  个动态输出的均值曲线  $\bar{y}_{s\sigma}, \bar{y}_{r\sigma}$  可认为是对系统输出的一次抽样, 不考虑不确定性影响时的多元输出数据是近似独立的, 进而综合得到最终验证结果如下所示.

$$C(\mathbf{Y}_s, \mathbf{Y}_r) = w_1^- \cdot \sum_{i=1}^{\beta_s} w_i \cdot \left( \sum_{l=1}^{L_i} w_l C_i^l(F_{sil}, F_{rjl}) \right) + w_2^- \cdot \sum_{\zeta=1}^{L_{\sigma}} w_{\zeta} C_{\sigma}^{\zeta}(\bar{y}_{s\sigma}, \bar{y}_{r\sigma}) \quad (21)$$

其中,  $\sigma = 1, \dots, m_{dynamic}$  表示第  $\sigma$  个动态输出变量,  $\zeta = 1, \dots, L_{\sigma}$  表示动态输出均值曲线的第  $\zeta$  个特征,  $w_l, w_{\zeta}$  代表第  $l, \zeta$  个数据特征的可信度结果权重,  $w_i$  代表第  $i$  个相关变量子集的一致性分析结

果权重,  $w_1$ 、 $w_2$  代表相关变量子集和动态输出均值曲线一致性的权重.

#### 4 应用实例与分析

为验证本文方法的有效性, 针对文献 [2] 中给出的某飞行器纵向平面内末制导阶段的仿真模型进行结果验证. 该模型包括飞行器制导模型和目标运动模型. 图 3 给出纵向平面内弹目相对运动几何关系. 目标以恒定速度  $v_T$  沿  $x$  轴向右行驶. 假设飞行器无动力飞行且航向已对准目标, 忽略地球自转, 给出以时间为自变量的飞行器纵向质心运动方程

$$\begin{cases} \dot{v} = -\frac{D}{M} - g \sin \theta \\ \dot{\theta} = \frac{L}{Mv} - \frac{g \cos \theta}{v} \\ \dot{h} = v \sin \theta \\ \dot{d} = v \cos \theta \end{cases} \quad (22)$$

式中,  $v$  为速度,  $\theta$  为弹道倾角,  $h$  为高度,  $d$  为水平距离. 阻力  $D = 0.5\rho v^2 S C_D (Ma, \alpha)$ , 升力  $L = 0.5\rho v^2 S C_L (Ma, \alpha)$ .  $C_D$  与  $C_L$  分别为阻力系数与升力系数.  $\alpha$  为攻角, 马赫数  $Ma = v/v_s$ ,  $S$  为参考面积,  $M$  为质量,  $\alpha_M$  为法向加速度,  $\lambda$  为视线角. 声速  $v_s$  与大气密度  $\rho$  根据 1976 年美国标准大气计算. 相应的制导律设计可见文献 [30]. 根据上述信息建立该飞行器纵向平面内末制导仿真模型.

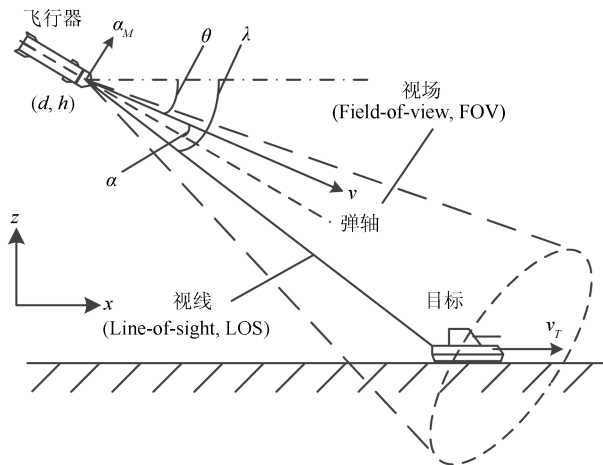


图 3 纵向平面内弹目相对运动几何关系

Fig. 3 Geometrical relationship of relative missile-target movement in longitudinal plane

利用此仿真模型精确研究此飞行器在末制导阶段的特性, 需要考虑其受到的不确定性因素. 飞行器升力与阻力均存在不确定性, 引入升力系数扰动  $C_{LC}$ , 与阻力系数扰动  $C_{DC}$  模拟阻力系数与升力系数的不确定性, 因此有

升力  $L = 0.5\rho v^2 S C_{LC} C_L (Ma, \alpha)$ , 阻力  $D = 0.5\rho v^2 S C_{DC} C_D (Ma, \alpha)$ , 分别采用不同分布对  $C_{LC}$  和  $C_{DC}$  进行描述. 同时, 大气密度会影响升力和阻力, 因每次飞行环境不同需考虑其不确定性的影响, 采用大气密度系数  $C_\rho$  表示. 此外, 飞行器进入末制导阶段时的初始视线角  $\lambda_0$  与弹道倾角  $\theta_0$  亦具有不确定性. 选取仿真模型和参考系统的不确定参数如表 2 所示.

表 2 飞行器末制导过程的不确定参数取值

Table 2 Uncertainty parameters values in the terminal guidance process of flight vehicle

| 变量名                    | 仿真模型参数分布        | 参考系统参数分布         |
|------------------------|-----------------|------------------|
| 大气密度系数 $C_\rho$        | $N(0, 0.033)$   | $N(0, 0.033)$    |
| 升力系数 $C_D$             | $N(0, 0.05)$    | $N(0.02, 0.07)$  |
| 阻力系数 $C_L$             | $N(0, 0.033)$   | $N(0.02, 0.033)$ |
| 初始弹道倾角 $\theta_0$ /rad | $N(0.17, 0.09)$ | $N(0.26, 0.07)$  |
| 初始视线角 $\lambda_0$ /rad | $N(0.17, 0.09)$ | $N(0.17, 0.09)$  |

选取用户关注的多元输出变量如表 3 所示. 选取静态输出变量有飞行器的最终落点位置坐标  $(x_f, z_f)$  和目标终点位置坐标  $(x_{Tf}, z_{Tf})$ , 同时选取待验证的动态输出变量有弹道倾角  $\theta$ 、攻角  $\alpha$ 、视线角  $\lambda$ 、弹目相对距离  $D_{MT}$ 、目标速度  $v_T$ . 利用拉丁超立方抽样法, 对模型不确定性参数进行抽样, 给定初始样本数为 1000, 运行仿真模型共得到 1000 组输出. 改变飞行器模型参数 (见表 2), 采用拉丁超立方抽样获得的 1000 组数据作为参考输出. 系统输出数据的包络线如图 4~10 所示. 其中目标速度  $v_T$ 、目标终点位置  $z_{Tf}$  为恒定值, 未在图中标出.

表 3 待验证的模型输出

Table 3 Model outputs to be validated

| 变量类型 | 变量名                      |
|------|--------------------------|
| 动态   | 弹道倾角 $\theta$ (rad)      |
| 动态   | 攻角 $\alpha$ (rad)        |
| 动态   | 视线角 $\lambda$ (rad)      |
| 动态   | 弹目相对距离 $D_{MT}$ (m)      |
| 动态   | 目标速度 $v_T$ (m/s)         |
| 静态   | 飞行器落点 X 坐标 $x_f$ (m)     |
| 静态   | 飞行器落点 Z 坐标 $z_f$ (m)     |
| 静态   | 目标终点位置 X 坐标 $x_{Tf}$ (m) |
| 静态   | 目标终点位置 Z 坐标 $z_{Tf}$ (m) |

利用本文方法对该飞行器末制导仿真输出进行验证. 首先利用多元输出变量选择方法分别对仿真和参考的动静态输出变量进行相关性分析及变量选择, 得到相关性分析结果如表 4 所示. 通过分析可知, 仿真和参考输出变量具有相同的变量子集, 动态



输出变量  $\theta$ 、 $\alpha$ 、 $\lambda$ 、 $D_{MT}$  具有相关性, 故将其归为一类.  $v_T$  为定值 (不随时间改变) 并与变量子集 I 相互独立; 静态输出  $x_f$ 、 $x_{Tf}$  具有相关性, 通过验证可知两者满足线性关系 (如图 11 所示), 同时  $z_f$  与  $x_f$  相互独立 (如图 12 所示), 进而可得  $z_f$  与  $x_f$  相互独立,  $z_{Tf}$  为定值 0 形成了变量子集 III. 由上述分析结果验证了变量选择方法的有效性.

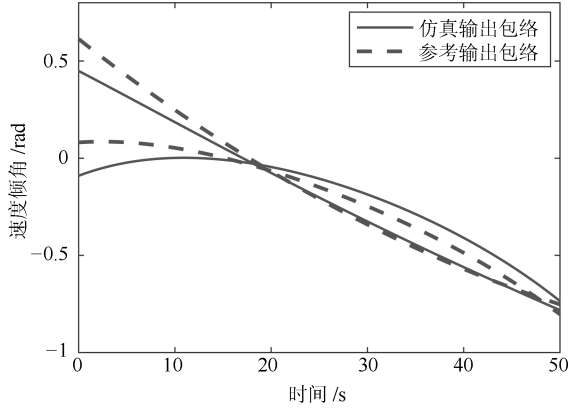


图 4 弹道倾角输出包络线

Fig. 4 Envelope lines of launch angle

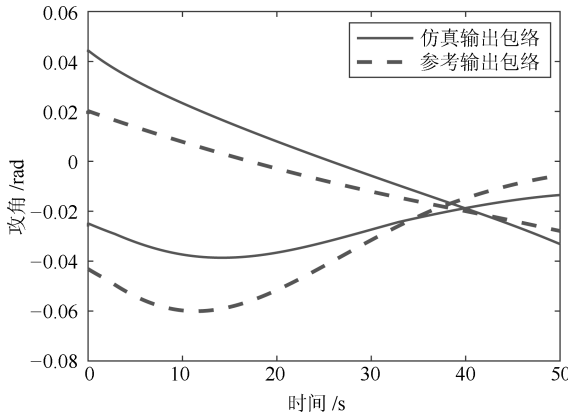


图 5 攻角输出包络线

Fig. 5 Envelope lines of angle of attack

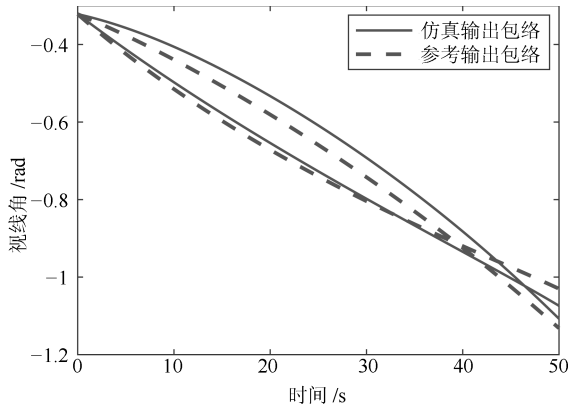


图 6 视线角输出包络线

Fig. 6 Line-of-sight angle envelopes

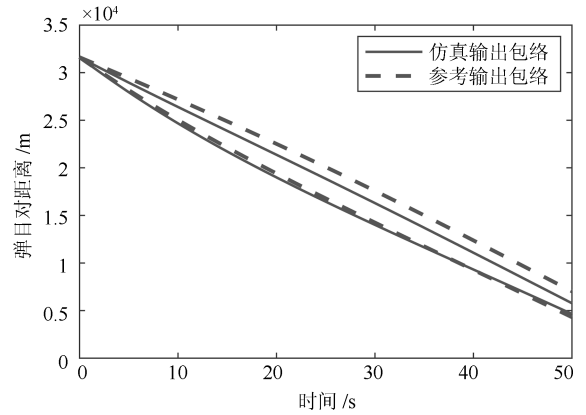


图 7 弹目相对距离输出包络线

Fig. 7 Envelope lines of the missile-target relative distance

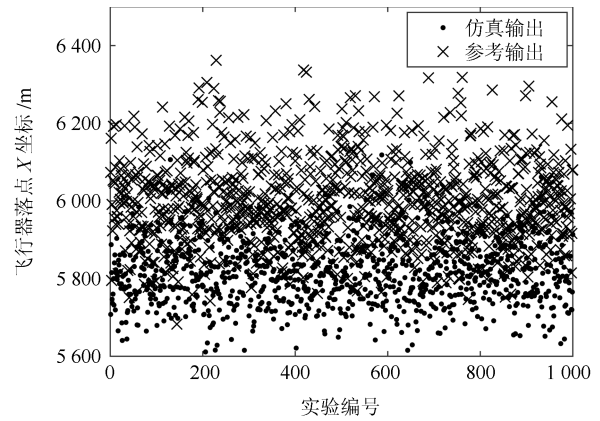


图 8 飞行器落点 X 坐标输出散点图

Fig. 8 Scatter diagram of X-direction drop point coordinates of the flight vehicle

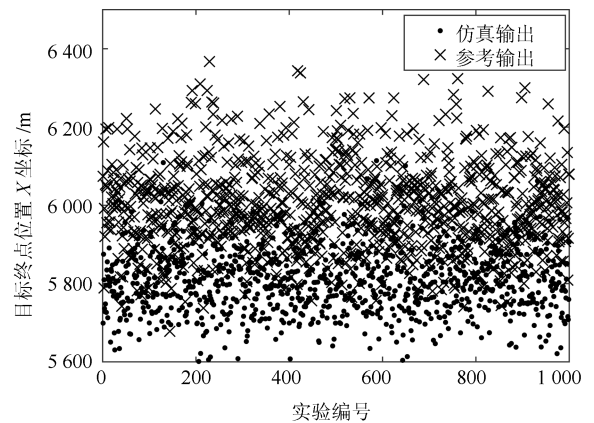


图 9 目标终点位置 X 坐标输出散点图

Fig. 9 Scatter diagram of X-direction terminal point coordinates of the target vehicle

根据表 4 的变量选择结果求取各变量子集关于某特征的联合 CDF, 选取动态输出的位置和形状特征, 分别求取变量子集 I 的联合 CDF, 变量子集 II 为恒定值在验证过程中直接采用相对误差方法进行

一致性分析即可; 对于静态输出变量子集 I 关于数据本身的联合 CDF 如图 13 所示, 变量子集 II 的 CDF 曲线如图 14 所示. 进而得到动态输出均值曲线的一致性结果 (见表 5) 以及多个变量组关于多个特征的 CDF 差异和可信度结果 (见表 6). 依据式 (21) 综合多个可信度结果得到最终验证结果为 0.82, 由于仿真和参考输出变量  $v_T$ 、 $z_{Tf}$  均相等且恒为 0, 故在计算模型可信度时不予考虑, 为方便计算采用均权的方式进行综合.

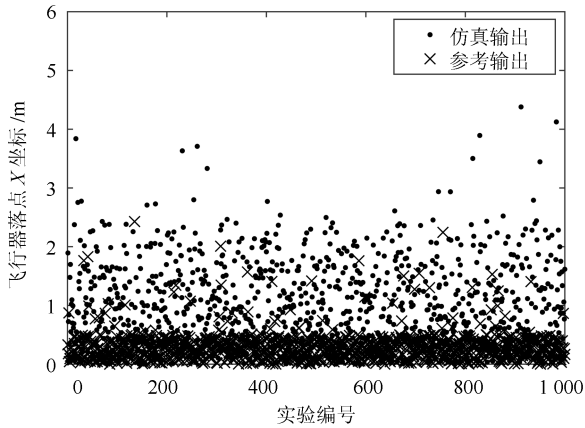


图 10 飞行器落点 Z 坐标输出散点图

Fig. 10 Scatter diagram of the terminal point of the target in the Z direction

表 4 多元输出变量选择结果

Table 4 Variables selection results of multiple outputs

| 输出类型 | 变量子集 I                            | 变量子集 II | 变量子集 III |
|------|-----------------------------------|---------|----------|
| 动态   | $\theta, \alpha, \lambda, D_{MT}$ | $v_T$   | —        |
| 静态   | $x_f, x_{Tf}$                     | $z_f$   | $z_{Tf}$ |

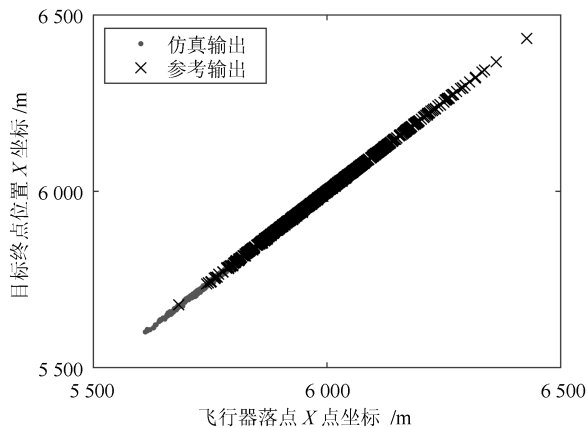


图 11 飞行器落点 X 坐标与目标终点位置 X 坐标间的关系

Fig. 11 Relationship of X-direction coordinates between drop point of flight vehicle and terminal point of target

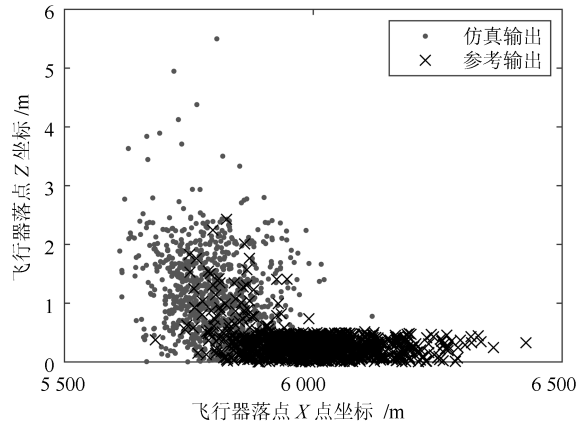


图 12 飞行器落点 X 方向坐标与 Z 方向坐标间的关系

Fig. 12 Relationship between X-direction and Z-direction coordinates of the drop point of flight vehicle

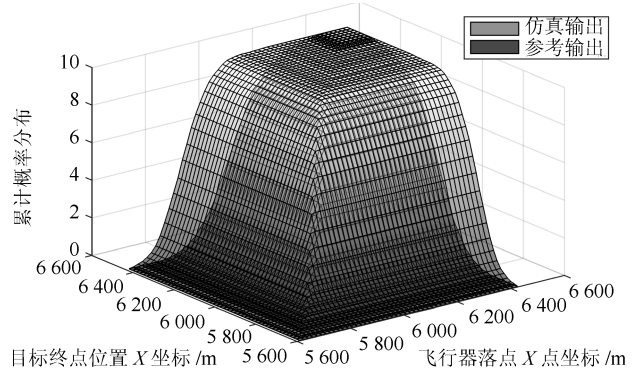


图 13 仿真和参考静态输出变量子集 I 的联合 CDF 对比

Fig. 13 JCDF comparison of variable subset I between static simulation and reference output

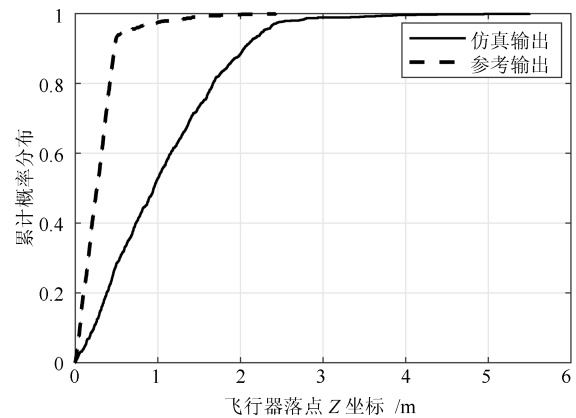


图 14 仿真和参考静态输出变量子集 II 的 CDF 对比

Fig. 14 Comparison of variable subset II between static simulation and reference output

此外, 为进一步验证本文方法对参数不确定性度量的有效性, 针对上述应用实例分别设计两组验证实验 (不确定性参数取值见表 7). 固定仿真模型和参考系统的不确定性参数大气密度系数  $C_p$ 、升力

系数  $C_D$ 、阻力系数  $C_L$  和初始视线角  $\lambda_0$  的取值. 分别调节仿真模型初始弹道倾角  $\theta_0$  的均值和方差, 得到最终验证结果如图 15~16 所示. 通过实验可得, 该方法能够度量仿真模型不确定参数取值的离散程度对验证结果的影响, 证明过大或过小的参数不确定度均会降低模型的可信度; 同时该方法能够度量不确定性参数的均值差异对验证结果的影响. 综上所述, 所提方法能够用于解决带有相关性的多元输出仿真结果验证问题.

表 5 动态输出均值曲线的一致性分析结果

Table 5 Consistency analysis results of the mean curves of dynamic outputs

| 变量名       | 位置特征一致性 | 形状特征一致性 |
|-----------|---------|---------|
| $\theta$  | 0.92    | 0.74    |
| $\alpha$  | 0.63    | 0.60    |
| $\lambda$ | 0.98    | 0.74    |
| $D_{MT}$  | 0.97    | 0.61    |

表 6 仿真和参考输出变量子集的一致性分析结果

Table 6 Consistency analysis results of the variables subset of the simulation and reference outputs

| 输出变量类型 | 变量子集标号   | 累积概率分布差异                    | 可信度结果      |
|--------|----------|-----------------------------|------------|
| 动态     | 变量子集 I   | 位置差异: $8.92 \times 10^{-8}$ | 位置特征: 0.99 |
| 动态     | 变量子集 I   | 形状差异: $1.1 \times 10^{-3}$  | 形状特征: 0.94 |
| 动态     | 变量子集 II  | 0                           | 1          |
| 静态     | 变量子集 I   | $1.6 \times 10^5$           | 0.84       |
| 静态     | 变量子集 II  | 0.5                         | 0.9        |
| 静态     | 变量子集 III | 0                           | 1          |

表 7 验证实验的不确定参数取值

Table 7 Uncertainty parameters values for validation experiments

| 试验编号 | 参考系统 $\theta_0$ 取值 | 实验组 I $\theta_0$ 取值 | 实验组 II $\theta_0$ 取值 |
|------|--------------------|---------------------|----------------------|
| 1    | $N(0.26, 0.07)$    | 0.26                | $N(0.15, 0.07)$      |
| 2    | $N(0.26, 0.07)$    | $N(0.26, 0.04)$     | $N(0.21, 0.07)$      |
| 3    | $N(0.26, 0.07)$    | $N(0.26, 0.07)$     | $N(0.26, 0.07)$      |
| 4    | $N(0.26, 0.07)$    | $N(0.26, 0.1)$      | $N(0.31, 0.07)$      |
| 5    | $N(0.26, 0.07)$    | $N(0.26, 0.13)$     | $N(0.37, 0.07)$      |

### 5 结论

针对带有相关性的多元输出仿真模型验证问题, 提出了考虑不确定性的联合验证方法. 首先对多变量输出提取相关变量子集, 并对各输出变量提取数据特征, 利用联合 CDF 差异法度量各相关变量子集的一致性程度, 进而综合得到模型可信度. 利用单变量验证方法进行多变量验证时需要满足输出变量相

互独立的条件. 本文方法考虑了多变量间的相关关系, 基于相关变量子集进行联合验证, 较单变量验证方法应用更合理; 同时在验证前引入了变量相关性分析, 使其能够适应输出变量之间关系未知的情况, 使验证结果更准确, 但也增加了计算开销. 此外, 该方法能够度量不确定性因素对模型可信度的影响.

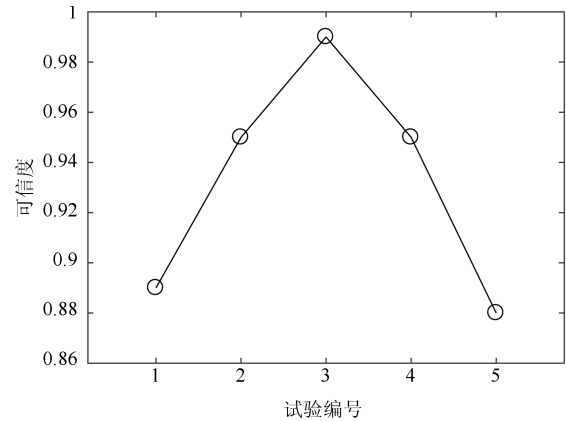


图 15 实验组 I 验证结果

Fig. 15 Validation result of experiment I

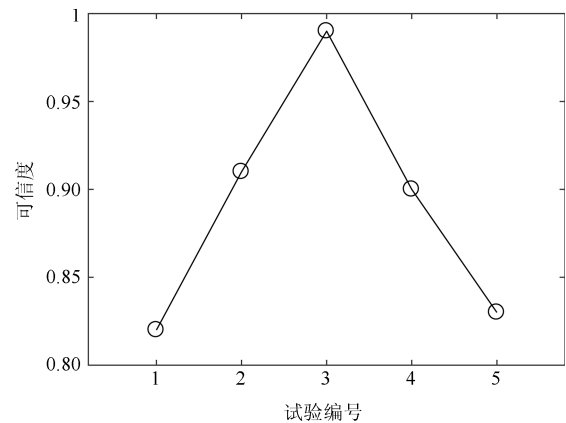


图 16 实验组 II 验证结果

Fig. 16 Validation result of experiment II

需要说明的是, 本文仅考虑同一类型输出 (动态或静态) 存在相关性的情况, 涉及的变量选择方法本质上属于数据挖掘方法, 为确保方法的准确性要求具备足够的样本容量, 对于参考数据缺乏的情况, 可采用专家给出参考数据的大致分布, 或可利用已有的历史数据、可信度较高且类似的半实物/纯数字仿真模型所产生的数据代替. 此外, 刻画动态输出的数据特征不限于距离、形状及频谱, 可依据具体应用需求而定 (例如, 超调量、相位误差等). 后续将对动态、静态输出间的相关性分析及变量选择方法进行研究; 同时针对参考数据缺乏以及存在认知和固有混合不确定性时的多元输出仿真结果验证问题展开进一步研究.

## References

- 1 Li Wei, Jiao Song, Lu Ling-Yun, Yang Ming. Validation and selection of simulation model based on the feature differences. *Acta Automatica Sinica*, 2014, **40**(10): 2134–2144 (李伟, 焦松, 陆凌云, 杨明. 基于特征差异的仿真模型验证及选择方法. *自动化学报*, 2014, **40**(10): 2134–2144)
- 2 Yang Ming, Qian Xiao-Chao, Li Wei. Simulation dynamic output validation method based on the data feature. *Systems Engineering and Electronics*, 2016, **38**(2): 457–463 (杨明, 钱晓超, 李伟. 基于数据特征的仿真动态输出验证方法. *系统工程与电子技术*, 2016, **38**(2): 457–463)
- 3 Oberkampf W L, Trucano T G, Hirsch C. Verification, validation, and predictive capability in computational engineering and physics. *Applied Mechanics Reviews*, 2004, **57**(5): 345–384
- 4 Oberkampf W L, Barone M F. Measures of agreement between computation and experiment: Validation metrics. *Journal of Computational Physics*, 2006, **217**(1): 5–36
- 5 Oberkampf W L, Trucano T G. Verification and validation in computational fluid dynamics. *Progress in Aerospace Sciences*, 2002, **38**(1): 209–272
- 6 Liu Y, Chen W, Arendt P, Huang H Z. Towards a better understanding of model validation metrics. *Journal of Mechanical Design*, 2011, **133**(7): 071005
- 7 Ling Y, Mahadevan S. Quantitative model validation techniques: New insights. *Reliability Engineering and System Safety*, 2013, **111**: 217–231
- 8 Rebba R, Huang S P, Liu Y M, Mahadevan S. Statistical validation of simulation models. *International Journal of Materials and Product Technology*, 2006, **25**(1-3): 164–181
- 9 Jiang X M, Yuan Y, Mahadevan S, Liu X. An investigation of Bayesian inference approach to model validation with non-normal data. *Journal of Statistical Computation and Simulation*, 2013, **83**(10): 1829–1851
- 10 Sankararaman S, Mahadevan S. Integration of model verification, validation, and calibration for uncertainty quantification in engineering systems. *Reliability Engineering & System Safety*, 2015, **138**: 194–209
- 11 Wang H Y, Li W, Qian X C. An improved Jousselme evidence distance. In: *Proceedings of the 16th Asia Simulation Conference*. Beijing, China: Springer Verlag, 2016. 08–11
- 12 Ferson S, Oberkampf W L. Validation of imprecise probability models. *International Journal of Reliability and Safety*, 2009, **3**(1-3): 3–22
- 13 Ferson S, Oberkampf W L, Ginzburg L. Model validation and predictive capability for the thermal challenge problem. *Computer methods in Applied Mechanics and Engineering*, 2008, **197**(29-32): 2408–2430
- 14 Jiang X M, Mahadevan S, Urbina A. Bayesian nonlinear structural equation modeling for hierarchical validation of dynamical systems. *Mechanical Systems and Signal Processing*, 2010, **24**(4): 957–975
- 15 Rebba R, Mahadevan S. Validation of models with multivariate output. *Reliability Engineering & System Safety*, 2006, **91**(8): 861–871
- 16 Jiang X M, Mahadevan S. Bayesian inference method for model validation and confidence extrapolation. *Journal of Applied Statistics*, 2009, **36**(6): 659–677
- 17 Zhan Z F, Fu Y, Yang R J, Peng Y H. Bayesian based multivariate model validation method under uncertainty for dynamic systems. *Journal of Mechanical Design*, 2012, **134**(3): 034502
- 18 Li W, Chen W, Jiang Z, Lu Z Z, Liu Y. New validation metrics for models with multiple correlated responses. *Reliability Engineering & System Safety*, 2014, **127**: 1–11
- 19 Zhao L F, Lu Z Z, Yun W Y, Wang W J. Validation metric based on Mahalanobis distance for models with multiple correlated responses. *Reliability Engineering & System Safety*, 2017, **159**: 80–89
- 20 Sousa E P, Traina A J, Wu L, Faloutsos C. A fast and effective method to find correlations among attributes in databases. *Data Mining and Knowledge Discovery*, 2007, **14**(3): 367–407
- 21 He Kuan, Chen Sen-fa. Comprehensive evaluation method based on one of causal diagram and its application. *Control and Decision*, 2010, **25**(10): 1513–1518 (何宽, 陈森发. 基于一类因果图的综合评价方法及应用. *控制与决策*. 2010, **25**(10): 1513–1518)
- 22 Yamamoto H, Yamaji H, Fukusaki E, Ohno H, Fukuda H. Canonical correlation analysis for multivariate regression and its application to metabolic fingerprinting. *Biochemical Engineering Journal*, 2008, **40**(2): 199–204
- 23 Nelsen R B. *An Introduction to Copulas. 2nd Edition*. New York: Springer-Verlag, 2006. 8–25
- 24 Tseng F M, Yu H C, Tzeng G H. Applied hybrid grey model to forecast seasonal time series. *Technological Forecasting and Social Change*, 2001, **67**(2-3): 291–302
- 25 Kwak N, Choi C H. Input feature selection by mutual information based on Parzen window. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(12): 1667–1671
- 26 Wang Jin-Jia, Chen Chun. Multi-channel eeg feature extraction using hierarchical vector autoregression. *Acta Automatica Sinica*, 2016, **42**(8): 1215–1226 (王金甲, 陈春. 分层向量自回归的多通道脑电信号的特征提取研究. *自动化学报*, 2016, **42**(8): 1215–1226)
- 27 Mao Wen-Tao, Jiang Meng-Xue, Li Yuan, Zhang Shi-Guang. Structural prediction of multivariate time series based on outlier elimination. *Acta Automatica Sinica*, 2017, **44**(4): 619–634 (毛文涛, 蒋梦雪, 李源, 张仕光. 基于异常序列剔除的多变量时间序列结构化预测. *自动化学报*, 2017, **44**(4): 619–634)
- 28 Han M, Liu X X. Feature selection techniques with class separability for multivariate time series. *Neurocomputing*, 2013, **110**: 29–34
- 29 Lin Sheng-Lin, Li Wei, Ma Ping, Yang Ming. Ranking evaluation of simulation models based on Hilbert-Huang transform. *Systems Engineering and Electronics*, 2017, **39**(9): 2137–2142 (林圣琳, 李伟, 马萍, 杨明. 基于 Hilbert-Huang 变换的仿真模型排序评估方法. *系统工程与电子技术*, 2017, **39**(9): 2137–2142)

30 Zhou Di, Sun Sheng, Teo K L. Guidance laws with finite time convergence. *Journal of Guidance, Control, and Dynamics*, 2009, **32**(6): 061401

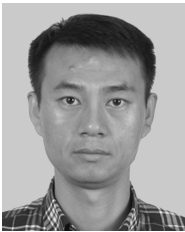


**林圣琳** 哈尔滨工业大学控制与仿真中心博士研究生. 主要研究方向为复杂系统仿真可信度评估与 VV&A, 基于仿真的作战效能评估.

E-mail: lin\_44627079@yeah.net

(**LIN Sheng-Lin** Ph.D. candidate at the Control and Simulation Center, Harbin Institute of Technology (HIT).

His research interest covers credibility assessment and VV&A for complex system simulation, simulation-based operational effectiveness assessment.)



**李 伟** 哈尔滨工业大学副教授. 2009 年获得哈尔滨工业大学航天学院博士学位. 主要研究方向为仿真分析与评估, 分布式仿真, 仿真实验设计.

E-mail: frank@hit.edu.cn

(**LI Wei** Associate professor at Harbin Institute of Technology (HIT).

He received his Ph.D. degree from School of Astronautics, HIT, Harbin, in 2009. His research interest covers simulation analysis and evaluation,

distributed simulation, simulation experiment design.)



**杨 明** 哈尔滨工业大学教授. 1997 年获得哈尔滨工业大学航天学院博士学位. 主要研究方向为飞行器制导与控制, 复杂系统仿真实理论与方法.

E-mail: myang@hit.edu.cn

(**YANG Ming** Professor at Harbin Institute of Technology (HIT). He received his Ph.D. degree from School of

Astronautics, HIT, Harbin, in 1997. His research interest covers vehicle guidance and control, complex system simulation theory and method.)



**马 萍** 哈尔滨工业大学教授. 2003 年获得哈尔滨工业大学航天学院博士学位. 主要研究方向为复杂系统建模与仿真, 分布式仿真与 VV&A. 本文通信作者.

E-mail: pingma@hit.edu.cn

(**MA Ping** Professor at Harbin Institute of Technology (HIT). She received her Ph.D. degree from School of

Astronautics, HIT, Harbin, in 2003. Her research interest covers complex system modeling and simulation, distributed simulation and VV&A. Corresponding author of this paper.)